



UNIVERSITE SULTAN MOULAY SLIMANE  
Faculté des Sciences et Techniques  
Béni-Mellal



*Centre d'Études Doctorales : Sciences et Techniques*  
*Formation Doctorale : Mathématiques et Physique Appliquées*

## THÈSE

Présentée par

**Rachid AIT DAOUD**

Pour l'obtention du grade de

**DOCTEUR**

*Spécialité : Informatique*

---

---

# **Réalisation d'un Système d'Aide à la Décision pour la Prédiction d'Attrition dans les Contextes Non-contractuels E-Commerce**

---

---

Soutenue le 17/06/2019 devant la commission d'examen composée de :

|                        |                                       |              |
|------------------------|---------------------------------------|--------------|
| Pr. Cherki DAOUI       | Professeur à la FST-USMS, Béni Mellal | Président    |
| Pr. Mohamed BASLAM     | Professeur à la FST-USMS, Béni Mellal | Rapporteur   |
| Pr. Hicham MOUNCIF     | Professeur à la FP-USMS, Béni Mellal  | Rapporteur   |
| Pr. Benayad NSIRI      | Professeur à la ENSET-UM5, Rabat      | Rapporteur   |
| Pr. Mohamed FAKIR      | Professeur à la FST-USMS, Béni Mellal | Examineur    |
| Pr. Rachid LBIBB       | Professeur à la FST-USMS, Béni Mellal | Encadrant    |
| Pr. Belaid BOUIKHALENE | Professeur à la FP-USMS, Béni Mellal  | Co-Encadrant |

# Liste des publications

## Publication dans des journaux

1. Rachid AIT DAOUD, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb, “Clustering Prediction Techniques in Defining and Predicting Customer Defection: The Case of E-commerce Context”, International Journal of Electrical and Computer Engineering (IJECE), Vol: 8, Num 4, 2018.
2. Rachid Ait daoud, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb, “Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco”, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol: 9, No: 8, c World Academy of Science, Engineering and Technology, pp1795-1805, 2015
3. Rachid Ait daoud, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb, “Combining RFM model and clustering techniques for customer value analysis of a company selling online”, In IEEE explorer, International Conference of Computer Systems and Applications (AICCSA), pp 1–6, 2015
4. Abdellah Amine, Rachid Ait daoud, Belaid Bouikhalene “Development of a Decision-Making System for Sultan Moulay Slimane University in Beni Mellal, Morocco”, Indonesian Journal of Electrical Engineering and Computer Science, 2, 469, pp469-477, 2016
5. Abdellah Amine, Rachid Ait daoud, Belaid Bouikhalene, “Efficiency Comparaison and Evaluation between Two ETL Extraction Tools”, Indonesian Journal of Electrical Engineering and Computer Science. 3. 174-181, pp174-181, 2016

## Conférences internationales

1. Rachid Ait daoud, Rachid Lbibb, Belaid Bouikhalene, “Approche data mining pour la classification des clients d’un site e-commerce”, Congrès Méditerranéen des Télécommunications (MTC, Faculté des Sciences et Techniques de Mohammedia, 22 et 23 Mai 2014
2. Rachid Ait daoud, Rachid Lbibb, Belaid Bouikhalene, “Data mining Techniques for Customer Relationship Management”, First International Conference on Business Intelligence (CBI’14), Beni Mellal, Morocco, April 29-30, 2014

## Résumé

Le thème de recherche abordé dans ce travail s'inscrit dans le cadre applicatif de la Gestion de la Relation Client (Customer Relationship Management ou CRM). Avec l'avènement des nouvelles Technologies de l'Information et de la Communication (TIC) et le développement du commerce électronique en particulier, une nouvelle philosophie de Gestion de la Relation Client est apparue : le CRM électronique (eCRM). L'eCRM se repose sur les concepts développés par le marketing relationnel et par les canaux électroniques afin qu'il automatise les processus de satisfaction, de fidélisation et le plus important ceux de rétention. Par la croissance ininterrompue du nombre de site e-commerce, les clients auront plus de choix qu'auparavant. Ce qui les encourage à diviser leurs achats sur plusieurs sites et à comparer les produits et les services de concurrents, cela augmente finalement le risque d'attrition (perte de clientèle). Les entreprises de vente en ligne perçoivent actuellement la nécessité d'une stratégie de rétention dans laquelle elles peuvent prédire les clients fidèles en voie de quitter partiellement ou totalement le site marchand afin que ces entreprises puissent anticiper un tel comportement défectueux. Le secteur de commerce électronique s'opère dans le cadre non contractuel où la perte des clients est difficile à définir et à tracer. Ce qui rend la prédiction d'attrition n'est pas facile à réaliser. L'objectif applicatif de ce travail de recherche est de concevoir une nouvelle approche de définition et de prédiction d'attrition adaptée à ce type d'environnement. L'approche conçue dans la thèse permet, en effet, d'élaborer une définition raisonnable d'attrition partielle et totale sur la base de laquelle des modèles prédictifs seront construits.

Cette approche est basée sur deux techniques de data mining : la clustering et la classification. Nous effectuons comme première étape, une segmentation de clients basée sur l'algorithme k-means et sur le modèle LRFM (Longueur, Récence, Fréquence et Montant) afin d'identifier les différents types de clients et d'analyser la valeur de chacun de ces types. Cette segmentation sert à arriver par la suite, à la définition du concept d'attrition partielle et d'attrition totale dans un cadre non contractuel. Sur la base des résultats obtenus au court de la première étape, nous utilisons dans la deuxième étape trois modèles de classification pour prédire l'attrition, à savoir, les réseaux de neurones, les arbres de décision et les méthodes d'ensemble dans lesquelles la variable cible (prédite) classe un client particulier soit en tant que client partiellement

défectueux, totalement défectueux, soit en tant que client poursuivant son comportement d'achat fidèle.

Pour tester nos modèles sur une base de données réelle d'un site e-commerce, nous appliquons une validation croisée à 10 fold. En termes de problèmes de prédiction multi-classes, nous évaluons l'efficacité de ces modèles par la mesure Macro-averaging comprenant l'efficacité moyenne, PrecisionM, RecallM et F-scoreM.

Nous effectuons une analyse comparative des différents modèles construits. Par conséquent, la performance du modèle basé sur les méthodes d'ensemble (Ensemble d'arbres de décision) est plus robuste et efficace que les autres modèles.

Nous présentons ainsi, une procédure qui permet de classer, selon l'importance, les variables prédictives utilisées dans la phase d'apprentissage de ces trois modèles.

En comparant les résultats de notre approche avec d'autres recherches récentes qui traitent la prédiction d'attrition dans les cadres de commerce électronique, nous montrons que les variables qui décrivent les taux d'abondance de session à chaque étape de processus de paiement, sont plus importantes dans la phase d'apprentissage que les autres variables prédictives classiques du modèle RFM, à savoir, la récurrence, la fréquence et le montant.

Mots clés : Aide à la décision, Data warehouse, Data mining, E-commerce, CRM, Prédiction d'attrition.

## **Abstract**

The research theme addressed in this work is in the context of Customer Relationship Management (CRM). With the advent of Information and Telecommunications Technology (ICT) and the development of electronic commerce in particular, a new CRM philosophy has emerged: the electronic CRM (eCRM). The CRM relies on the concepts developed by both relational marketing and the electronic channels to automate the processes of satisfaction, loyalty and the, more importantly, those of retention. With the growth of the e-commerce sector, customers have more choices, a fact which encourages them to divide their purchases amongst several e-commerce sites and compare their competitors' products, yet this increases high risks of churning. Online selling companies currently perceive the need for a retention strategy in which they can predict loyal customers likely to partially or totally leaves the e-commerce website in the future so that these companies can anticipate such defect behavior. The e-commerce sector operates in the non-contractual setting where the customer churn is difficult to define and trace. What makes the prediction of churn is not easy to achieve. The goal of this thesis is to develop new data mining approach adapted to this type of environment to defining and predicting churn. Consequently, the approach developed in this thesis allows to developing a reasonable definition of partial and total churn on the basis of which predictive models will be constructed.

This approach is based on two data mining techniques: clustering and classification techniques. We perform as a first step, a customer segmentation based on the k-means algorithm and on the LRFM model (Length, Recency, Frequency and Monetary) in order to identify the different types of customers and to analyze the value of each of these types. This segmentation is used next, to arrive at the definition of the concept of partial attrition and total attrition in a non-contractual setting. Based on the results obtained in the first step, we use in the second step three classification models to predict churn, including artificial neural networks, decision trees, and ensemble methods in which the target variable (dependent) classifies a particular customer either as a total defector, partial defector customer, or as a customer continuing his loyal buying behavior.

To test our models on a real database of an e-commerce site, we apply a 10-fold cross validation. In terms of multi-class prediction problems, we evaluate the accuracy of these models by the Macro-averaging measure including average accuracy,  $Precision_M$ ,  $Recall_M$ , and  $F-score_M$ .

We perform a comparative analysis of the different models built. As a result, the performance of the model based on the ensemble methods (Decision Tree Ensemble) is more robust and efficient than other models.

We present as well, a procedure which enables to rank the different predictors used in the learning phase of the models according to their importance.

By comparing the results of our approach with other recent research that deals with customer prediction churn in the e-commerce context, we show that variables which describe the dropout rates in the buying process steps, are more important than traditional predictors of the RFM model, namely, recency, frequency and monetary.

Keywords: Decision support, Data warehouse, Data mining, E-commerce, CRM, Churn prediction.

## Dédicaces

À ma mère, à mon épouse

À mes chères frères et sœurs

À la mémoire de mon père **Youssef AIT DAOUD**, qu'Allah ait pitié de son âme

À la famille Ait DAOUD

À tous mes amis(es)

## Remerciements

Je tiens tout d'abord à remercier Monsieur Rachid LBIBB professeur à la Faculté des Sciences et Techniques de Béni Mellal et Monsieur Belaid BOUIKHALENE professeur à la Faculté Polydisciplinaire de Béni Mellal, de m'avoir dirigé durant ces années de thèse. Je les remercie pour leur disponibilité, leurs précieux conseils concernant la méthodologie de recherche et pour la relecture du manuscrit de thèse. Je les salue pour leur compétence et pour leur ouverture d'esprit.

Je souhaiterais aussi adresser ma profonde reconnaissance aux membres du jury : Messieurs Cherki DAOUI, Mohamed BASLAM, Hicham MOUNCIF, Benayad NSIRI et Mohamed FAKIR qui m'ont fait l'honneur de participer à ma soutenance et d'avoir accepté de relire cette thèse et d'en être rapporteurs.

Un remerciement particulier à Monsieur Ahmed Boumezzough pour nos échanges constructifs durant cette recherche et pour les nombreuses discussions qui m'ont permis d'enrichir mes connaissances scientifiques et ma culture générale.

Je remercie infiniment ma mère, mon épouse ainsi que mes frères et mes sœurs, pour le soutien qu'ils m'ont apporté tout au long de cette thèse.

La réalisation de ce travail a été rendue possible grâce à la compréhension, le conseil, les commentaires et le soutien d'autres personnes qui ne sont pas mentionnés ici mais dont l'aide a tout autant été appréciée



# Table des matières

|  |      |
|--|------|
| Liste des publications .....                                   | i    |
| Publication dans des journaux .....                            | i    |
| Conférences internationales.....                               | i    |
| Résumé.....  | ii   |
| Abstract .....   | iv   |
| Dédicaces .....  | vi   |
| Remerciements .....  | vii  |
| Table des figures.....   | xii  |
| Liste des tableaux .....                                       | xiii |
| Abréviations .....   | xiv  |
| Contexte et objectifs.....                                     | 1    |
| Introduction.....  | 1    |
| 1. Motivation.....   | 2    |
| 2. Approches proposées .....                                   | 3    |
| 3. Principales contributions .....                             | 4    |
| 4. Organisation du manuscrit .....                             | 5    |
| Chapitre 1 :Le commerce électronique.....                      | 7    |
| Introduction.....  | 7    |
| 1. Généralité sur le e-commerce .....                          | 7    |
| 1.1. Définition .....  | 7    |
| 1.2. Evolution du commerce électronique .....                  | 11   |
| 1.3. Déroulement d'une opération du commerce électronique..... | 13   |

|  |    |
|--|----|
| 2. Types d'échanges .....  | 16 |
| 2.1. Le commerce électronique B to B (Business to Business).....                 | 16 |
| 2.2. Le commerce électronique B to C (Business to Consumer).....                 | 17 |
| 2.3. Le commerce électronique C to C (Consumer to Consumer) .....                | 17 |
| 2.4. Le commerce électronique B to G (Business to Government) .....              | 17 |
| 2.5. Le commerce électronique G to C (Government to Consumer) .....              | 18 |
| 3. Avantages et inconvénients du commerce électronique .....                     | 18 |
| 3.1. Les avantages.....  | 19 |
| 3.1.1. Les avantages pour l'entreprise.....                                      | 19 |
| 3.1.2. Les avantages pour les clients .....                                      | 20 |
| 3.2. Les inconvénients .....   | 21 |
| 3.2.1. Les inconvénients pour l'entreprise .....                                 | 21 |
| 3.2.2. Les inconvénients pour le client.....                                     | 22 |
| 4. Les méthodes de paiement.....   | 23 |
| 4.1. Les méthodes de paiement en ligne .....                                     | 25 |
| 4.2. La sécurité des paiements en ligne:.....                                    | 28 |
| 5. Le commerce électronique au Maroc : Quelques chiffres .....                   | 29 |
| 5.1. Les acteurs du commerce électronique au Maroc.....                          | 30 |
| 5.2. Les chiffres clés du e-commerce au Maroc (au 30 septembre 2018).....        | 31 |
| 5.3. Cinématique de paiement en ligne avec CMI.....                              | 33 |
| 5.4. Sécurisation des transactions et gestion de la fraude par CMI .....         | 35 |
| Conclusion .....   | 36 |
| Chapitre 2 : Gestion de la relation client à l'ère du commerce électronique..... | 37 |
| Introduction .....   | 37 |

|  |    |
|--|----|
| 1. Gestion de la relation client.....  | 38 |
| 1.1. Définition et objectifs de la CRM .....   | 38 |
| 1.2. Les trois dimensions de la CRM.....   | 39 |
| 1.2.1. Le CRM opérationnel.....  | 39 |
| 1.2.2. Le CRM collaboratif .....   | 39 |
| 1.2.3. Le CRM analytique .....   | 40 |
| 1.3. Cycle du CRM .....  | 41 |
| 1.4. Gestion de la relation client dans le commerce électronique .....                         | 44 |
| 1.4.1. L'e-CRM: une extension du CRM .....   | 44 |
| 1.4.2. Les stratégies CRM dans le contexte e-commerce. ....                                    | 45 |
| Conclusion .....   | 53 |
| Chapitre 3 : La découverte de connaissances et les techniques de data mining pour le CRM ..... | 54 |
| Introduction .....   | 54 |
| 1. Extraction de connaissances à partir de données.....  | 55 |
| 1.1. Enjeux .....  | 55 |
| 1.2. Processus de découverte de connaissances .....  | 56 |
| 1.2.1. Etapes du processus d'ECD.....  | 57 |
| 1.2.2. Le standard CRISP-DM comme modèle de processus d'ECD .....                              | 58 |
| 1.3. La fouille de données .....   | 61 |
| 1.4. Applications.....   | 65 |
| 2. Les techniques de data mining et la prédiction d'attrition (churn) : Etat de l'art ....     | 68 |
| 2.1. Prédiction d'attrition / churn (Revue de littérature) .....                               | 68 |
| Conclusion .....   | 73 |

|   |     |
|---|-----|
| Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision.....    | 74  |
| Introduction .....  | 74  |
| 1. Méthode de recherche .....   | 77  |
| 1.1. Phase de compréhension du problème.....  | 77  |
| 1.2. Phase de compréhension des données .....   | 79  |
| 1.3. Phase de préparation des données (conception et alimentation du data warehouse)..... | 80  |
| 1.4. Phase de modélisation.....   | 85  |
| 1.4.1. Modèle de segmentation : .....   | 85  |
| 1.4.2. Modèle de prédiction : .....   | 106 |
| 1.5. Phase d'évaluation.....  | 112 |
| 1.5.1. Les mesures de performances .....  | 112 |
| 2. Résultats et discussion .....  | 116 |
| Conclusion .....  | 126 |
| Conclusion et perspectives .....  | 128 |
| Bibliographie .....   | 132 |

## Table des figures

|   |     |
|---|-----|
| Figure 1.1 – Les cinq perspectives du e-Commerce (source [7]) .....   | 8   |
| Figure 1.2 – Le déroulement d’une opération commerciale en ligne : Cas de paiement par carte bancaire. ....   | 15  |
| Figure 1.3 – Les différents types de commerce électronique .....  | 18  |
| Figure 1.4 – Les chiffres clés de paiement via Internet 2014-2017.....  | 33  |
| Figure 1.5 - Schéma d’une transaction e-commerce par carte de paiement (source CMI [20])... ..  | 34  |
| Figure 2.1 - Les interactions entre les trois dimensions du CRM.....  | 41  |
| Figure 2.2 - Schéma du cycle de vie d'un client .....   | 42  |
| Figure 3.1 - Processus d’Extraction de Connaissances à partir des données [92]. ....  | 57  |
| Figure 3.2 - Structure séquentielle et feedback dans un modèle de processus d’ECD. ....   | 59  |
| Figure 3.3 – Cycle de vie d’un projet data mining avec le modèle CRISP-DM.....  | 60  |
| Figure 3.4 - Les différentes méthodes de data mining.....   | 62  |
| Figure 4.1- Schéma relationnel des tables. ....   | 81  |
| Figure 4.2- Le workflow utilisé pour la création et l’alimentation des dimensions de notre data mart (Knome Analytics Platform). ....                               | 83  |
| Figure 4.3- Modélisation des données en utilisant le schéma en étoile. ....   | 84  |
| Figure 4.4- Période d'observation. ....   | 91  |
| Figure 4.5 - Changement du modèle LRFM des clients au fil du temps pour définir le churn (perte de clientèle).....  | 92  |
| Figure 4.6 - La méthodologie proposée basée sur le modèle LRFM et la technique K-means pour définir et identifier le churn dans les contextes non-contractuels..... | 93  |
| Figure 4.7 - Le résultat des méthodes « Elbow » et « Silhouette » pour déterminer le nombre optimal de clusters .....   | 94  |
| Figure 4.8 – Présentation des 7 clusters trouvés par la méthode K-means (k=7).....  | 95  |
| Figure 4.9 - Graphique de silhouette pour k allant de 2 à 10.....   | 98  |
| Figure 4.10 : Vue d’ensemble sur les sept clusters. ....  | 101 |
| Figure 4.11 : Exemple de filtrage, vue d’ensemble sur les clients définis comme des clients partiellement défectueux.....   | 104 |
| Figure 4.12 : Vue d’ensemble sur le cluster LRFM.....   | 105 |
| Figure 4.13 - Exemple d'architecture d'un MLP à 5 variables d'entrée, 3 neurones pour la couche cachée et un neurone pour la couche de sorties. ....                | 107 |
| Figure 4.14- Vue d’ensemble sur les trois classes de clients en termes de taux d’abandon et de probabilité d’attrition. ....  | 124 |

Figure 4.15- Exemple de résultat du rapport lorsque nous choisissons un clients précis. .... 125

## Liste des tableaux

|  |     |
|--|-----|
| Table 3.2 Les modèles de prédiction de churn. ....   | 70  |
| Table 4.1 – Les données utilisées et leurs attributions.....   | 80  |
| Table 4.2 Tables de la base de données source.....   | 82  |
| Table 4.3 Dimensions et table du fait du data mart .....   | 82  |
| Table 4.4 Les actions utilisées dans la création et l'alimentation du data mart.....   | 83  |
| Table 4.5 - Définition du modèle LRFM .....  | 88  |
| Table 4.6 - Statistiques descriptives de la longueur, la récence, la fréquence et le montant .....   | 88  |
| Table 4.7 - Statistiques descriptives de sept clusters obtenus par la méthode k-means en T1 .....  | 98  |
| Table 4.8 - Les variables prédictives utilisées dans notre étude .....   | 110 |
| Table 4.9 - Les prédicteurs comportementaux de la défection et le type de variable cible utilisées dans les recherches antérieures .....     | 111 |
| Table 4.10 - Matrice de confusion .....  | 113 |
| Table 4.11 - Mesures de performance.....   | 113 |
| Table 4.12 - Matrice de confusion : cas de problème à trois classes .....  | 114 |
| Table 4.13 - Exactitude de prédiction des trois modèles avec une cross-validation à 10 folds..   | 116 |
| Table 4.14 - La matrice de confusion (par Artificial Neural Networks). ....  | 117 |
| Table 4.15 - La matrice de confusion (par Simple Decision Tree).....   | 117 |
| Table 4.16 - La matrice de confusion (par Decision Tree Ensemble). ....  | 118 |
| Table 4.17 - L'exactitude globale, les mesures de Précision, Rappel et F-score pour les trois classes (Réseaux de neurones). ....            | 118 |
| Table 4.18 - L'exactitude globale, les mesures de Précision, Rappel et F-score pour les trois classes (Arbre de décisions).....              | 118 |
| Table 4.19 - L'exactitude globale, les mesures de Précision, Rappel et F-score pour les trois classes (Ensemble d'arbres de décisions). .... | 119 |
| Table 4.20 – Les valeurs de la mesure Macro-averaging pour les trois modèles. ....   | 119 |
| Table 4.21 - Importance des différentes variables prédictives utilisées pour la prédiction de churn. ....                                    | 121 |

## **Abréviations**

CRM : Customer Relationship Management

eCRM : Electronic Customer Relationship Management

TIC: Technologies de l'information et de la communication

CMI : Centre Monétique Interbancaire

MTC : Maroc Télécommerce

LRFM : Longueur, Récence, Fréquence, Montant

T1 : Période de calibration

T2 : Période de prédiction

VP : Vrais Positifs

VN : Vrais Négatifs

FP : Faux Positifs

FN : Faux Négatifs

WCSS : La variance générale au sein du cluster

K : Nombre de clusters

N: Nombre de classes

DT: Décision Tree (Arbre de décision)

ANN : Artificiel Neural Networks (Réseaux de neurones)

DTE : Decision Three Ensemble (Ensemble d'arbres de décisions)

# Contexte et objectifs

## Introduction

Aujourd'hui, le marché du commerce électronique partout dans le monde est devenu très saturé en raison de nombre plus élevé des sites e-commerce compétitifs. Au Maroc par exemple, le développement de ce marché se fait à un rythme accéléré. Selon les chiffres publiés par le centre monétique interbancaire du Maroc (CMI), les sites marchands affiliés à ce dernier ont réalisé 6,3 millions d'opérations de paiement en ligne via cartes bancaires, marocaines et étrangères, pour un montant global de 2,4 milliard de DH durant les 9 premiers mois 2018, en progression de +30,4% en nombre et +21,1% en montant par rapport à la même période en 2017, ce qui signifie que l'internaute marocain est devenu plus familiarisé avec les systèmes de paiement en ligne. Par conséquent, afin de servir et faire face à la concurrence liée à la croissance explosive du commerce électronique, il est important pour une entreprise de commerce électronique de prendre les mesures nécessaires pour développer des activités de marketing innovantes afin d'identifier leurs différents clients et faire de son mieux pour les conserver ou, au moins, prendre soin des plus fidèles et obtenir leur satisfaction.

Par rapport aux méthodes de transaction traditionnelles, le plus grand inconvénient du marché de commerce électronique est que le taux d'attrition est très élevé, parce qu'avec le nombre croissant de sites marchands en ligne, les clients auront beaucoup plus de choix qu'auparavant, ce qui les incite à diviser leurs achats sur plusieurs sites de e-commerce et à faire des comparaisons de produits et services de concurrents ce qui augmente finalement le risque d'attrition (perte de clientèle), et si l'attrition des clients continue à se produire pour n'importe quelle entreprise, cela entraînerait une perte considérable de ses revenus. Dans cette situation, le seul moyen pour remédier à ces risques commerciaux relatif à la perte des clients, les entreprises sont obligées de trouver des solutions basées sur les techniques de fouille de données (data mining) et les outils de statistiques pour prédire les clients qui risquent de quitter le site marchand dans un proche avenir et prendre des mesures immédiates afin de minimiser le risque d'attrition. Ceci est possible si l'historique de comportement d'achat des clients est analysé systématiquement. Heureusement, parmi les points forts du commerce électronique est qu'il permet de sauvegarder toutes les actions et le flux de clics effectués par les clients sur le site



marchand, sous forme des bases de données. Cette quantité volumineuse de données garantit la possibilité d'appliquer les techniques de fouille de données qui sont essentiel pour découvrir des connaissances utiles et cachées dans ces données.

## 1. Motivation

L'un des principaux avantages qui caractérisent le marché du e-commerce par rapport aux marchés traditionnels est que le premier offre aux entreprises une très grande quantité d'informations sur les clients et sur leurs comportements d'achats. En d'autre point de vue, plus les enregistrements de transactions d'une entreprise deviennent beaucoup plus importants plus que l'analyse de la valeur des clients devient une tâche plus complexe [1,2]. Cependant, vu que les clients ne contribuent pas de façon identique aux profits, il est primordial pour une entreprise de segmenter d'abord tous les clients en nombre approprié de clusters homogènes et mutuellement hétérogènes en fonction de certaines similitudes, puis de déployer des ressources auprès des clients en fonctions de leurs valeurs. Allenby et al. (1998) ont décrit qu'un ensemble exact de variables de segmentation pour une segmentation complète du marché n'existe pas [3].

Un autre défi à relever par les gestionnaires de site marchand est se reflète dans la capacité de prédire l'attrition ou la perte de clients « lorsque le client met fin à ses relations commerciales avec une entreprise ». L'attrition (Customer churn en anglais) représente une préoccupation importante pour toute entreprise et une question centrale d'intérêt dans l'analyse de base de clientèle, parce que la perte des clients non seulement implique des coûts d'opportunité en raison de la réduction des ventes, mais conduit également à un besoin accru d'attirer de nouveaux clients, soit 5-6 fois plus chère que la vente aux clients existants [4]-[5]. En outre, le marché de commerce électronique souffre de la nature de la relation entre le client et le site marchand ; il n'y a pas de contrat entre les deux parties, c'est un cadre non contractuel dans lequel les clients peuvent facilement changer leur comportement d'achat vers les concurrents sans être confrontés à des coûts de commutations et sans informer l'entreprise, ce qui rend l'attrition n'est pas facile à tracer et à prédire comme dans le cas des secteurs qui travaillent dans le cadre contractuel dans lequel il est facile de déterminer le moment précis où les clients interrompent leur relation : clôture de compte (Secteur bancaire) ou changement d'opérateur téléphonique (Télécommunication) ou fin de contrat (Assurance).

Une complication supplémentaire se produit dans les contextes non contractuels réside sur la difficulté de définir l'attrition, car les caractéristiques qui doivent être observées pour dire qu'un client a totalement ou partiellement défectueux ne sont pas clairement définies.

## 2. Approches proposées

La segmentation de la base de données client en nombre approprié de clusters homogènes à des fins de marketing ciblé et personnalisé, et la rétention des clients en évitant leur défection, sont considérées comme les défis les plus critiques à relever dans le domaine de gestion de la relation client. Cependant, une grande partie des études existantes liées à la gestion de la relation client traite la segmentation des clients et la prédiction d'attrition d'une façon indépendante.

Dans le cadre de notre travail, nous développons un modèle à deux étapes dans lequel un modèle de clustering (segmentation) et un modèle de prédiction sont utilisés pour définir et prédire l'attrition des clients dans les contextes non contractuels et plus précisément dans le secteur du e-commerce. Pour résoudre les problématiques citées dans la section précédente, nous divisons le processus de construction de notre modèle en deux phases :

Phase de segmentation, qui a deux objectifs principaux, le premier sert à identifier et analyser la valeur de chaque segment, tandis que le deuxième objectif sert à arriver à la définition du concept d'attrition partielle et totale dans le cadre non contractuel. Pour atteindre ces objectifs, nous avons proposé une nouvelle approche qui combine l'algorithme de clustering de data mining (K-means) et le modèle LRFM (Length, Recency, Frequency, Monetary). L'idée générale de cette approche est basée sur l'utilisation de deux intervalles de temps T1 et T2. T1 est utilisé pour identifier les différents groupes de clients on se basant sur l'extraction des variables du modèle LRFM qui caractérisent le comportement d'achat pour chaque client ; L : "Longueur de la relation client/entreprise qui est égale au nombre de jours entre la première et la dernière transaction", R : "La récence qui fait référence au nombre de jours entre le premier jour de la période d'étude et le jour de la dernière transaction", "F : Fréquence des achats sur la période considérée", "M : Le montant, qui indique le montant total dépensé par un client au cours de la période analysée", ensuite ces variables seront utilisées comme des paramètres d'entrées pour la technique de clustering K-means qui va générer comme résultat (K) groupes et chaque

client sera affecté à son groupe approprié selon ses valeurs de L, R, F et M. Dans la deuxième période T2 le même processus d'extraction des variables LRFM est répété, puis nous introduisons un nœud d'attribution de groupe qui va attribuer les clients existants en T2 aux groupes déjà obtenus par la technique k-means en T1. Finalement, nous définissons la migration du client d'un groupe avec une valeur importante en T1 vers un autre groupe de moins valeur en T2 comme un signal de défection partielle ou totale.

Nous proposons de combiner les méthodes somme des erreurs aux carrés et la silhouette pour déterminer à la fois le nombre optimal de segments et aussi pour mesurer la qualité de segmentation de notre modèle de clustering.

Phase de prédiction : En se basant sur les résultats de la première phase, trois techniques de classification : les arbres de décisions, les méthodes d'ensembles et les réseaux de neurones sont utilisées pour prédire l'attrition.

Pour évaluer la performance prédictive des modèles construits, la métrique d'évaluation Macro-moyenne (Macro-averaging) incluant les mesures suivantes *Accuracy*, *Precision*, *Recall* et *F-mesure* et la méthode validation croisée « K-fold » sont utilisées. Par conséquent, une analyse comparative des différents modèles est présentée pour sélectionner celui qui offre le meilleur pouvoir prédictif. Enfin nous proposons un workflow qui permet d'extraire les variables qui peuvent contribuer davantage à la prédiction de l'attrition (partielle/totale) dans le secteur de e-commerce.

### **3. Principales contributions**

Basé sur une gamme d'études sur les modèles de prédiction de la perte de clients, nous révélons que : (1) bien que plusieurs modèles prédictifs ont été développés pour modéliser l'attrition dans les secteurs de Télécommunication, Banque, Assurance, et la vente au détail, le secteur de commerce électronique en général et les paramètres non contractuels en particulier ont reçu moins d'attention à cet égard. En outre, aucune recherche antérieure n'a traité l'attrition partielle et totale simultanément dans un environnement en ligne non contractuel, elles sont plutôt concentrées seulement sur une parmi les deux, et dans la plupart des cas, un tel comportement de défection est défini comme une attrition totale.

Par conséquent, notre travail de recherche contribue à la littérature existante de deux façons importantes. Tout d'abord, nous proposons une nouvelle approche qui permet de définir et d'identifier à la fois les deux types d'attrition (partielle et totale) dans les contextes non contractuels, et particulièrement dans le e-commerce.

Deuxièmement, contrairement aux études antérieures qui se concentrent essentiellement sur les modèles de classification binaire où le modèle prédit l'état d'un client en tant que client perdu ou non, le présent travail aborde le cas des problèmes de classification multi-classes où la variable dépendante (cible) classe un client particulier, soit en tant que client partiellement défectueux, totalement défectueux, ou bien en tant que client poursuivant son comportement d'achat fidèle. Cette contribution est importante pour plusieurs raisons. Premièrement, puisque nous considérons les deux types de défection (partielle et totale), le degré de risque lié à la défection partielle est différente de celui de la défection totale, et en raison des coûts associés aux stratégies de rétention, il est conseillé de ne pas concentrer les efforts et les ressources dédiés à la gestion d'attrition de la même manière pour les deux niveaux de défection, mais il faut optimiser l'affectation des ressources en fonction de degré d'attrition [6]. En d'autres termes, un client prédit par le modèle comme un client en voie de défection partielle ne devrait pas être ciblé par le même programme d'incitation dédié à ceux qui sont susceptibles de quitter définitivement l'entreprise dans l'avenir, et vice versa. Cela aidera les décideurs et les responsables de marketing à faire de bonnes interactions au bon moment pour retenir ces clients sans gaspiller de ressources. Deuxièmement, les managers auront la possibilité de vérifier si l'attrition totale est toujours précédée par une attrition partielle, ou qu'il y'a des cas dans lesquels le client part définitivement sans laisser des signes d'insatisfaction, ce qui permettra aux managers de réfléchir à des solutions à de telles situations.

## 4. Organisation du manuscrit

Cette thèse est organisée comme suit :

Dans le **premier chapitre** nous donnons un aperçu sur le commerce électronique comme nouvelle forme du commerce classique. Nous montrons que l'émergence de ce dernier est directement liée à l'apparition des technologies de l'information et de la communication (TIC) et

surtout l'internet. Nous décrivons ensuite les différents types de cette nouvelle forme de commerce, ses principaux acteurs, les étapes d'une activité commerciale en ligne et les mécanismes principaux nécessaires pour la sécurité des transactions en ligne. Enfin nous donnons un état des lieux du commerce électronique au Maroc pour mettre en lumière la valeur de ce secteur dans l'économie marocaine, ainsi pour savoir comment les internautes marocains interagissent avec le paiement en ligne.

Le **deuxième chapitre** est dédié à la présentation du concept de système de gestion de la relation client (Customer Relationship Management CRM). Nous décrivons ce concept dans son sens général, ainsi nous essayons de montrer eCRM comme une extension du CRM, qui est intéressante pour la création des relations étroites entre les entreprises et les clients dans le contexte de commerce électronique. Nous montrons également comment adopter les stratégies de marketing relationnel traditionnelle pour la gestion de la relation client, à l'ère du commerce électronique.

Le **troisième chapitre** : La première partie de ce chapitre est consacré aux processus d'extraction de connaissances et plus précisément à la fouille de données (data mining). Nous présentons les techniques de data mining pour la relation client. Les techniques de data mining sont très variées, ici nous présentons quelques techniques populaires et leurs applications en CRM. La deuxième partie fait l'état de l'art sur les approches qui ont été développées pour la prédiction d'attrition (perte de clientèle).

Le **chapitre quatre** : Finalement, le dernier chapitre présente une nouvelle approche de définition et de prédiction d'attrition dans les cadres non contractuels, et en particulier dans le secteur de commerce électronique - la contribution principale de la thèse. Nous appliquons par la suite cette approche sur des données réelles d'un site e-commerce marocain pour prédire les clients qui sont en voie de défection partielle ou totale, ainsi ceux qui sont susceptibles de rester fidèles.

Nous concluons cette thèse en exposant nos contributions et quelques perspectives de recherche dans ce domaine.

# Chapitre 1 :Le commerce électronique

## Introduction

Avec l'évolution des technologies de l'information et de la communication (TIC) et surtout l'internet, le domaine de commerce a connu la naissance d'une nouvelle forme de commerce : Le commerce électronique, aussi appelé e-commerce qui s'oriente de plus en plus vers un échange axé principalement sur les valeurs immatérielles à travers des réseaux numériques qui ignorent les frontières et qui se caractérise par une forte croissance, touchant l'ensemble des secteurs économiques tels que le secteur touristique et hôtelier, secteur bancaire, distributions, secteur de l'habillement et de la mode.

Dans ce chapitre, nous présentons en revue la notion de commerce électronique, son évolution, ses différents types d'échange. Nous présentons également les instruments de paiement adaptés au commerce électronique. Ensuite, nous citons les avantages et les inconvénients de ce type de commerce. Finalement, nous présentons quelques chiffres montrant l'état du e-commerce au Maroc.

## 1. Généralité sur le e-commerce

### 1.1.Définition

On pourrait aisément être tenté de résumer le commerce électronique comme étant l'utilisation d'un média électronique pour la vente et l'achat de produits ou services. Mais il s'avère que cette convergence n'est pas suffisante pour expliquer ce qu'est la notion aujourd'hui. Il existe aujourd'hui plusieurs définitions concurrentes du commerce électronique qui présentent une image déroutante et incohérente du secteur. [7], ont collecté et analysé une grande variété de définitions de e-commerce pour arriver à une taxinomie de celles-ci suivant cinq points de vue : la vue commercial, la vue d'échange d'informations, les activités Business, les effets, et finalement la chaîne de valeurs (figure 1.1).



Figure 1.1 – Les cinq perspectives du e-Commerce (source [7])

- **La perspective “*Vue d’échange*”** : Ce type de commerce électronique est celui qui note l’ensemble des échanges liés à des activités commerciales en utilisant exclusivement un moyen électronique, c’est celui qui se pose probablement à l’esprit lorsqu’on utilise le terme de « commerce électronique ». [7] citent huit définitions de divers auteurs se reportant à cette vision du e-commerce. Nous prendrons comme illustration la définition de [8]: “*Le commerce est l’échange de produits et des marchandises, généralement à grande échelle. La mise en œuvre du commerce électronique consiste alors à échanger des produits et des marchandises en tout ou en partie par voie électronique.*”.

- **La perspective “ *Vue d’Echange d’Informations* ”** : Contrairement à la perspective Trading View qui se concentre seulement sur l’échange de produits et de services via des moyens électroniques, la perspective Information Exchange View note que le flux d’informations transféré lors d’une transaction est plus important que le flux matériel, parce que en réalité, le commerce électronique implique non seulement l’échange de produits et services,

mais également un flux très important d'informations qui le permet, le facilite et le guide[9]. Ces informations peuvent être utilisées par la suite pour améliorer les activités telles que le service à la clientèle et la conception de nouveaux produits. Une examination de l'ensemble de définitions de divers auteurs données par [7] et qui sont regroupées dans la vue d'échange d'informations montre que chaque une entre elle fait l'accent sur l'importance du transfert d'informations dans le commerce électronique et que les termes transactions, vente-achat, peuvent être contournés dans ces définitions. Parmi ces huit définitions nous reprendrons celle de [10] pour illustration : “ *Le commerce électronique vise à fournir des solutions électroniques à toutes les activités de communication d'informations entre deux parties.* ”.

- “ **Vue d'activité** ” : est la troisième perspective du commerce électronique qui considère le commerce électronique plus comme une simple activité de commerce mais qu'elle touche d'autres activités dites non-commerciale telles que l'exécution de transactions, les efforts de la poste et de pré-vente, et d'autres activités auxiliaires toujours en utilisant la technologie[11]. En d'autres termes, cette vision du commerce électronique est plus expansive que celle de la vue commerciale et complète la vue de l'échange d'informations. Nous prenons celle de [12].

“*Le commerce électronique peut être défini comme l'ensemble des actions soutenant les activités commerciales sur un réseau*”.

- **La perspective “ Vue des effets”** : L'ensemble de définitions données par les trois classes précédentes se concentrent dans leurs définitions du commerce électroniques seulement sur la nature des transactions (transactions commerciales, autres activités) et sur la façon dont elles sont effectuées (via des échanges d'informations basés sur la technologie), tandis que la perspective dite «*La vue des effets*» met l'accent sur deux aspects qui n'ont pas été adressé par les trois classes précédentes qui sont les raisons et les effets résultant de l'utilisation du commerce électronique. Les effets de celui-ci apparaissent dans tous les domaines d'activité, du service client à la conception de nouveaux produits[13]. Il facilite la mise en œuvre de nouveaux types de processus d'affaire et augmente le degré d'interaction avec les consommateurs. Cela peut influencer positivement sur la gestion des commandes en réduisant les coûts associés à cette dernière, et facilite également l'élargissement de la base des fournisseurs et des partenaires commerciaux, augmente la rapidité de réponse des consommateurs vers l'organisation et inversement, et améliore la qualité des services. La définition suivante du commerce électronique



est symptomatique de ceux qui concentrent sur les effets du commerce électronique donnée par [13] *“Au sens large, le commerce électronique est une méthodologie commerciale moderne qui répond aux besoins des organisations, des commerçants et des consommateurs pour réduire les coûts tout en améliorant la qualité des biens et des services et en accélérant la fourniture de services.”*.

En effet, une activité commerciale passe de plus en plus par un processus complexe qui fait appel à des entités intermédiaires, ce qui permet de créer des interdépendances (liens) entre la chaîne de valeur de l'entreprise et celles de ses fournisseurs et ses clients. Dans cette direction, la dernière perspective du commerce électronique nommée *«Vue de la chaîne de valeur»*, note que l'application des technologies de l'information et de la communication aux différents maillons de la chaîne de valeur se traduisait par l'amélioration de la compétitivité de la chaîne en général, ainsi l'utilisation des TIC dans différentes parties de la chaîne de valeur mettent en évidence plusieurs avantages concrets permettant de contribuer à la réduction des coûts de transaction des intermédiaires, d'apporter de la valeur au client grâce à un meilleur service et de réduire les coûts tangibles et intangibles de ses affaires. En guise d'illustration, nous prendrons la définition de [14] *“Le commerce électronique est un commerce rendu possible par les technologies de l'ère WWW, permettant l'intégration transparente des technologies de l'information, de la communication et la technologie logistique tout au long de la chaîne de valeur des processus métier, des fournisseurs de produits bruts et services aux clients finaux”*.

Cette taxinomie n'est pas exhaustive et pourrait faire l'objet de changements en fonction de l'évolution technologiques, sociales ou encore économiques qui vont sans cesse évoluer le commerce électronique et donc, sa propre définition. A partir de cette taxinomie du commerce électronique, [7] *donnent une définition plus intégrée « Le commerce électronique est une approche permettant d'atteindre des objectifs commerciaux dans lesquels la technologie d'échange d'informations permet ou facilite l'exécution d'activités dans et entre les chaînes de valeur, ainsi que le soutien à la prise de décision sous-jacente à ces activités»*. On doit noter que très beaucoup d'emphase est mise sur la prise de décision qui est omniprésente tout au long des activités au sein de la chaîne de valeur.

## 1.2.Evolution du commerce électronique

Le phénomène commercial que nous appelons maintenant le commerce électronique a eu une histoire intéressante. En général, la croissance du commerce électronique est étroitement liée au développement et à la diffusion de nouvelles technologies de l'information et surtout internet. Pour illustrer les phases d'évolution du commerce électronique dans le monde nous considérons quatre phases grâce auxquelles nous allons pouvoir observer l'évolution qu'a vécue le commerce électronique depuis sa création : phase avant 1990, début des années 90, milieu des années 90 à 2000 et finalement la phase de début du 21ème siècle jusqu'à présent.

- **Avant 1990** : Avant l'avènement de l'internet, les échanges électroniques de données étaient réalisables grâce à la technologie EDI (Electronic Data Inter change). Cette technologie a été inventé en 1948, à cette époque (les années cinquante) elle consistait à l'échange de données informatisées par un réseau de partenaires et de fournisseurs, a pour but principal l'automatisation de l'échange de documents comme la prise de commande, le suivi des colis envoyés et la facturation. Cependant, l'utilisation de cette technique par les entreprises reste relativement faible en raison des lourds investissements qu'exige l'EDI lors de son implantation, et à la nature de communication entre les deux parties ce qui a ramené à une migration vers un nouveau moyen de commerce électronique, à savoir l'Internet.

- **Début des années 90** : Au début des années 90, les activités commerciales sur Internet ont évolué pour fonctionner dans un environnement qui était ouvert et facilement accessible. Et avec l'apparition du service principal de l'internet qui est le WWW, l'avènement de l'identifiant unifié URL et la disponibilité et la gratuité des navigateurs, le partage et l'affichage de l'information sont devenus beaucoup plus faciles.

Peu à peu, les sites web primitifs commencent à remplir le web. Parallèlement à ces progrès techniques, des moteurs de recherche puissants basés sur le web ont été développés ce qui permet de rendre le WWW comme un canal d'atteindre efficacement les clients potentiels par les entreprises. A cette époque, les sites web étaient généralement une source d'information, ainsi les portails généralistes comprennent les fonctions de base telles que le listage, le catalogage, la publication et le regroupement sont devenus populaires. Toutefois, malgré le progrès remarquable du Web, la nature de la communication entre les entreprises et leurs clients reste unidirectionnelle, ce qui limite les entreprises à interagir qu'aux demandes des clients, tandis que

la possibilité d'extraire des informations sur les clients potentiels pour améliorer les relations existantes ce n'était pas disponible à l'époque. Un autre obstacle au développement du commerce électronique à cette époque était l'absence d'un système qui garantit la transmission sécurisée de données confidentielles.

- **Milieu des années 90 à 2000 :** Alors que la nature de la communication était unidirectionnelle dans les débuts des années 90, celle-ci a été remplacé par une communication bidirectionnelle entre l'entreprise et le client, assurant une interactivité entre les deux parties. Cette interactivité a été obtenue principalement en raison de l'émergence d'une technologie de base appelée "cookies" qui permet de recueillir et stocker des informations à partir du terminal utilisé par le visiteur lors de sa visite afin de mieux servir les clients lors des prochaines visites. Ainsi grâce à l'interactivité et à l'utilisation de d'autres langages de programmation web, les sites web e-commerce arrivent à fournir des services personnalisés, en offrant des contenus adaptés aux intérêts des clients. En 1995, le commerce électronique concerne peu à peu tous les secteurs d'activités et toutes les fonctions de l'entreprise : la distribution des produits culturels, livres ou musique, l'industrie automobile, le tourisme, les intermédiaires financiers, la publicité et également la recherche d'emploi, Amazon.com et eBay étaient les plus remarquables sites de e-commerce qui ont été publiés dans cette période. Compte tenu du risque d'utilisation de données confidentielles telles que le numéro de carte de crédit dans des environnements publics comme Internet, il été nécessaire de trouver des solutions à ce problème avant que les activités en ligne puissent s'épanouir. Par conséquent, en 1995 des systèmes de cryptographie ont été inventés pour développer le protocole SSL (Secure Socket Layer) sur le web afin d'établir une liaison chiffrée entre les deux parties (entre le serveur web et le visiteur), et pour garantir la confidentialité et l'intégrité des données transmises. La technologie de base a continué d'évoluer pour que sa fonctionnalité s'améliore tout comme la capacité de gérer le trafic de site Web lourd, fournir des interfaces visiteurs sophistiquées et également l'amélioration des interfaces de langues avec des bases de données telles que ODBC. L'année 1998 a connu la naissance de deux grands organismes, Google et PayPal, qui ont joué un rôle très important dans l'évolution et la création de nouvelles activités de commerce électronique.

- **Du 2001 jusqu'à présent :** La croissance de l'internet autour du monde a incité de nombreuses entreprises à établir leurs présences sur le web en vendant directement leurs produits par internet, et avec le nombre croissant de personnes qui ont commencé à faire des affaires en

ligne, un besoin de communication et de transactions sécurisées est devenu évident. En 2004, le Conseil des normes de sécurité PCI a été créé pour s'assurer que les entreprises respectent les différentes exigences de sécurité, et afin d'augmenter le niveau de contrôle des informations du titulaire de la carte dans le but de réduire l'utilisation frauduleuse des instruments de paiement d'une part, et d'autre part pour favoriser la confiance d'utiliser ces moyens de paiements.

La croissance mondiale de l'utilisation des médias sociaux comme Facebook, Twitter et Youtube a influé directement et positivement sur le secteur de commerce électronique, la conversation entre les entreprises et les consommateurs est devenue plus engageante, ce qui facilite l'accès des échanges transactionnels en ligne. Par exemple la nouvelle innovation du « Bouton acheter » de Facebook et Twitter permet aux utilisateurs d'acheter un produit en direct sur le réseau, sans passer par le site e-commerce de l'entreprise.

Finalement, avec l'utilisation croissante de l'Internet, des tablettes PC, des smartphones et avec la confiance des consommateurs envers les systèmes de paiement électronique, le commerce électronique continuera à évoluer et offrira des services de plus en plus intéressants et sophistiqués dans l'avenir.

### **1.3.Déroulement d'une opération du commerce électronique**

Avant d'illustrer le déroulement d'une opération commerciale en ligne, il est nécessaire de citer les principaux types d'acteurs impliqués dans le commerce électronique. On distingue trois types d'acteurs :

**Les clients :** Sont des particuliers ou des entreprises qui achètent un produit ou un service pour satisfaire un besoin quelconque. Ces clients peuvent effectuer des achats en ligne en utilisant plusieurs modes de paiement offerts par le vendeur à savoir les cartes bancaires (Visa, MasterCard, Carte blue), porte-monnaie électronique (Paypal, Paybox, Olkypay), virement, chèque et cash à la livraison.

**Le vendeur :** Une personne ou une entreprise qui vend un bien ou un service sur l'internet en utilisant un site web marchand pour gérer la commercialisation de l'ensemble des produits et services mis en vente. Le vendeur doit gérer un certain nombre des tâches qui sont essentiels pour répondre aux besoins de ses clients à savoir la gestion du site web et sa sécurité, gestion des demandes et de paniers, mise en place des moyens de paiement sécurisés et finalement la livraison des objets de transactions aux clients. La gestion de toutes ces données par le vendeur

peut faire appel à des intermédiaires qui établissent ou facilitent le processus de transaction entre le vendeur et le client par le biais d'un support informatique.

**Les intermédiaires :** Ils représentent des acteurs principaux dans le commerce électronique, on distingue deux types :

- Les intermédiaires techniques : Les fournisseurs d'accès Internet qui sont chargés d'assurer la disponibilité de la liaison Internet, les administrateurs des systèmes et serveurs qui sont responsables de la sécurité au niveau physique et la traçabilité des interactions et le web master qui contrôle la bonne mise en ligne des contenus de site marchand et sa maintenance.
- Les intermédiaires financières : Ils sont notamment chargés d'émettre les cartes de crédits, d'effectuer les transferts d'argent du compte client à celui du vendeur et également ils assurent la sécurité de paiement. Par exemple, on trouve les organismes de paiement qui se charge d'assurer l'honnêteté et la protection des informations nécessaires au paiement transmises par les clients telles que le type de carte bancaire, son numéro, sa date d'expiration, et leurs codes PIN. On trouve aussi d'autres médiateurs comme Paypal, Neteller, Skrill qui offrent un service de paiement en ligne pour payer des achats ou de recevoir de paiement sans communiquer les coordonnées bancaires.

Les différentes étapes d'une activité commerciale en ligne sont détaillées dans la figure 1.2

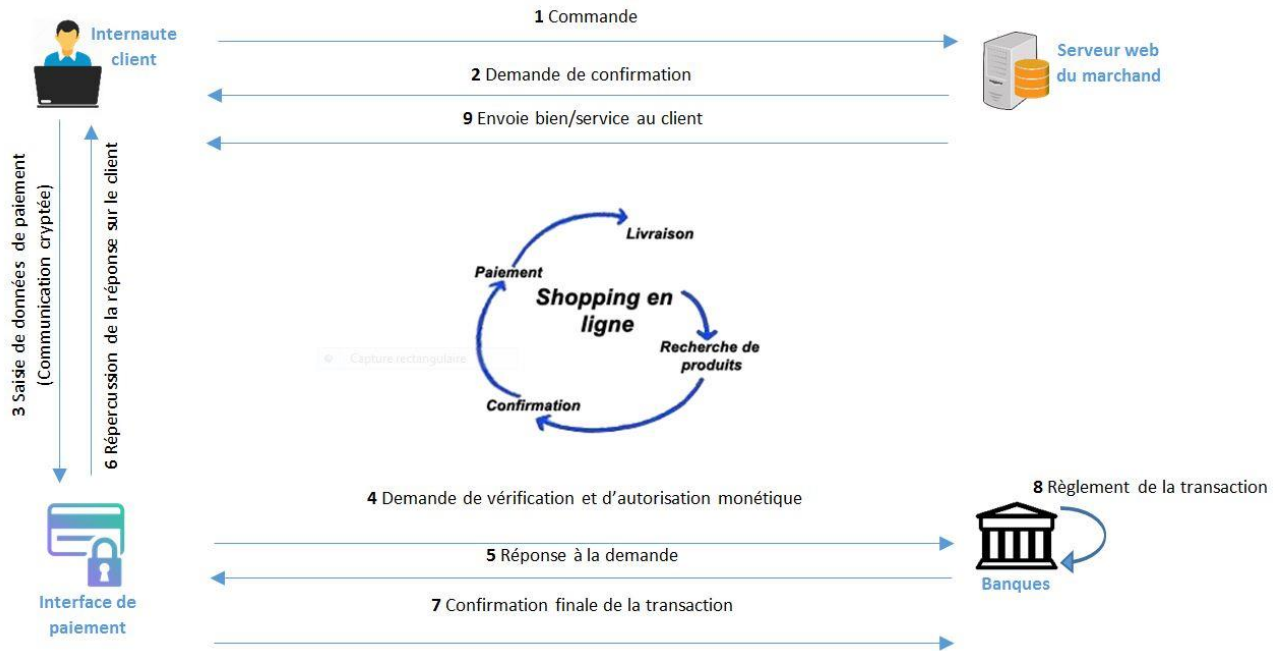


Figure 1.2 – Le déroulement d’une opération commerciale en ligne : Cas de paiement par carte bancaire.

- L’étape 1 : Le client se connecte au site marchand et procède à ajouter des produits à son panier ou aux créances à régler.
- L’étape 2 : Une fois le client valide son choix, une demande de confirmation lui sera envoyée en cliquant généralement sur le bouton « Payer ».
- L’étape 3 : Une fois sa confirmation validée, il sera redirigé automatiquement vers l’interface de paiement d’un opérateur de paiement en ligne qui se charge d’assurer une connectivité sécurisée en temps réel depuis le site marchand jusqu’à la plate-forme de passerelle de paiement. Le client sera ensuite invité à communiquer quelques informations relatives à la carte bancaire telles que sa type (Visa, MasterCard, Maestro, Carte Blue, etc.), son numéro, sa date d’expiration et son code de vérification.
- L’étape 4-5 : En temps réel, la plate-forme de passerelle de paiement transite en toute sécurité les informations de paiement de client vers les banques appropriées par le biais du réseau financier. Ces dernières vérifient ces informations, et donc soient elles acceptent ou refusent la demande d’autorisation.

- L'étape 6 : Lorsque la réponse d'autorisation est reçue du réseau financier, elle sera vérifiée et enregistrée par la plate-forme de passerelle de paiement, ensuite elle sera retournée par la même session au client sur le site marchand sous forme d'un reçu de paiement si l'autorisation a été accordée et acceptée, ou bien sous forme d'un message de refus dans le cas échéant. Le marchand reçoit à son tour la réponse d'autorisation (paiement autorisée) par courrier électronique ou fax.
- L'étape 7 : A la réception d'une confirmation de prise en charge de la part du marchand, la plate-forme de passerelle de paiement envoie une confirmation finale de compensation aux banques appropriées.
- L'étape 8 : Une fois la confirmation finale de compensation est arrivée, les banques procèdent au règlement en débitant le compte du client et en créditant le compte du marchand.
- L'étape 9 : Le marchand procède finalement à la livraison selon l'adresse fournie par le client.

## 2. Types d'échanges

En général, lorsque nous pensons au commerce électronique, nous pensons à une transaction commerciale en ligne entre un vendeur et un acheteur. Cependant, et bien que cette idée soit juste, nous pouvons être plus précis si nous basons sur la nature des parties. En effet, les trois acteurs les plus courants du commerce électronique sont les entreprises, les consommateurs et les gouvernements. Sur cette base, nous pouvons proposer quatre types principaux de commerce électronique, tous avec des caractéristiques différentes.

### 2.1.Le commerce électronique B to B (Business to Business)

Cette forme de commerce est la plus ancienne, elle est apparue dans les années 50 avec l'invention de la technologie EDI, cette dernière consistait en l'échange électronique de certaines informations traditionnellement communiquées sur papier d'une entreprise à une autre par le biais d'un réseau de télécommunication pour faciliter les transactions entre eux. Dans ce type de commerce électronique, les deux participants sont des entreprises. En conséquence, le volume et la valeur du commerce électronique B2B peuvent être énormes. Les fabricants et les grossistes

de commerce traditionnel opèrent généralement avec ce type de commerce électronique. En d'autres termes, Le commerce électronique B to B est le support électronique des transactions commerciales entre entreprises et il couvre un large éventail d'applications qui permettent à une entreprise de former des relations électroniques avec leurs distributeurs, revendeurs, fournisseurs et autres partenaires.

## **2.2.Le commerce électronique B to C (Business to Consumer)**

Le commerce électronique B to C est probablement le type que la plupart des gens connaissent, et comme son nom l'indique, c'est le type de commerce électronique qui prend en compte l'interaction entre les entreprises et les consommateurs. Il permet aux consommateurs de parcourir le catalogue, de choisir le produit et de le commander en ligne. Cette forme de commerce électronique simplifie l'expérience d'achat pour les consommateurs et les entreprises, mais la complexité et le coût de la logistique peuvent être un obstacle à la croissance de ce type de commerce électronique B2C.

## **2.3.Le commerce électronique C to C (Consumer to Consumer)**

Le commerce électronique de consommateur à consommateur ou C2C est simplement un commerce entre particuliers ou consommateurs. Ce type d'e-commerce se caractérise par la croissance des marchés électroniques et des ventes aux enchères en ligne. Par exemple : eBay est une plate-forme où vous pouvez publier sur internet une annonce pour la vente d'un produit ou d'un service et une autre personne achète simplement les produits directement auprès de vous. Ce type de commerce électronique est connu pour son anonymat et pour la facilité d'enregistrement. Cependant, il y a aussi le risque qu'un produit ne soit pas envoyé ou que la qualité de l'article ne corresponde pas aux attentes de l'acheteur. Les sites marchands dans ce genre de commerce électronique ne sont que des intermédiaires assurent que l'interaction entre les consommateurs et ne s'occupent pas de vérifier la qualité des produits offerts.

## **2.4.Le commerce électronique B to G (Business to Government)**

Business to Government (B2G) est une forme de commerce électronique qui se réfère aux entreprises vendant des produits, des services ou des informations aux gouvernements ou aux organismes gouvernementaux via Internet.



## 2.5.Le commerce électronique G to C (Government to Consumer)

Bien que nous ne puissions pas définir strictement les échanges en ligne entre les administrations et les individus comme étant des activités commerciales, nous pouvons voir plusieurs applications G2C dans le cadre des transactions qui sont faites et traitées plus efficacement avec les systèmes et technologies du commerce électronique y compris le paiement des taxes, l'immatriculation des véhicules et la fourniture d'informations et de services.

Le schéma ci-dessous récapitule les différents types du commerce électronique.

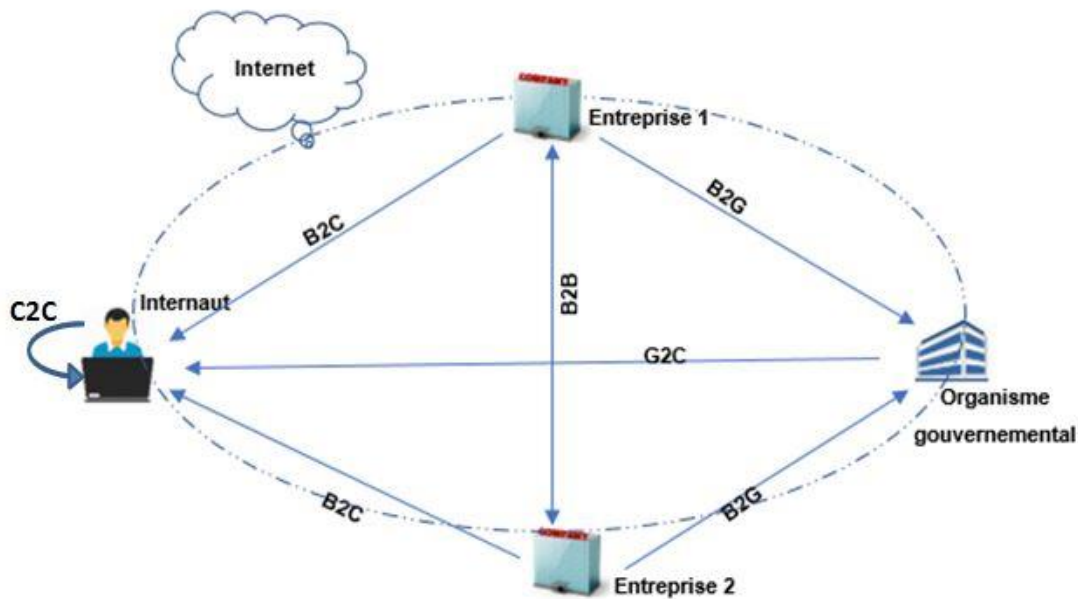


Figure 1.3 – Les différents types de commerce électronique

## 3. Avantages et inconvénients du commerce électronique

Le commerce électronique a changé la pratique, le calendrier et la technologie de toutes les formes du commerce (B2B, B2C, C2C, B2G et G2C). Il a influé positivement sur les prix, la disponibilité des produits, les modes de paiement et également sur le comportement des consommateurs. Grâce à cette nouvelle forme de commerce qui est basée sur l'internet, l'internaute pourra acheter n'importe quel article dans le monde entier. En général, le commerce électronique a été conçu dans le but d'offrir de nombreux avantages aux entreprises et aux acheteurs, mais comme toute technologie elle présente également des inconvénients.

### **3.1.Les avantages**

#### **3.1.1. Les avantages pour l'entreprise**

- Surmonter les limitations géographiques :

Le principal avantage du commerce électronique est qu'il permet à l'entreprise d'atteindre le marché mondial. Le commerce électronique répond simultanément aux demandes des entreprises nationales et internationales et permet de faire sortir leurs activités de leurs frontières géographiques. Avec le commerce électronique, même les petites entreprises obtiennent un accès au marché mondial.

- Plus besoin d'avoir un local professionnel

Comme il n'y a pas besoin d'un magasin physique, les entreprises de commerce électronique économisent sur l'un des frais généraux les plus élevés que les détaillants doivent supporter.

- Contact permanent avec les clients pendant 24 heures et 7 jours

Les sites Web de commerce électronique peuvent fonctionner tout le temps, cela rend les horaires d'ouverture de magasin 24/7/365. Du point de vue du commerçant, cela augmente le nombre de commandes reçues. Du point de vue du client, un magasin « toujours ouvert » est plus pratique.

- Nouveau canal de distribution

Le commerce électronique ouvre un nouveau canal de distribution, complémentaire, pour certains produits et services de l'entreprise.

- Réduction des coûts salariaux

Puisque les processus de commerce électronique sont largement automatisés, moins d'employés sont nécessaires pour les emplois de bas de gamme. L'automatisation supprime les employés traditionnels qui sont affectés souvent aux tâches de traitement des commandes et de logistique.

- Fournir des informations abondantes

La quantité d'informations pouvant être affichées dans un magasin physique est limitée. Cependant, il est difficile d'équiper les employés pour répondre aux clients qui ont besoin

d'informations à travers les gammes de produits par exemple. Les sites Web de commerce électronique peuvent facilement fournir des informations supplémentaires accessibles aux clients. La plupart de ces informations sont fournies par les fournisseurs et ne coûtent rien à créer ou à maintenir.

- Large gamme de produits et services :

En raison de la contrainte d'espace, un fournisseur ne peut stocker qu'une quantité minimale de marchandises dans le cas d'un magasin physique. Mais avec les magasins virtuels où il n'y a pas de limite d'espace ou de taille de magasin, les entreprises de commerce électronique peuvent répertorier de nombreux articles différents et sans tenir compte du coût de l'inventaire.

- Personnalisation supérieure :

En utilisant des cookies et d'autres méthodes de surveillance du comportement d'un client, un site Web de commerce électronique peut personnaliser de nombreux aspects afin d'offrir le bon contenu, au bon moment et à la bonne personne.

### **3.1.2. Les avantages pour les clients**

- Fournir des comparaisons des prix et des produits

Le commerce électronique facilite les comparaisons des prix et des produits. Il existe plusieurs services en ligne qui permettent aux clients de parcourir plusieurs marchands de commerce électronique et de trouver les meilleurs prix afin de pouvoir finalement prendre les décisions d'achat les plus appropriées.

- Abondance de choix

Le commerce électronique permet aux clients de choisir un produit ou un service de leur choix auprès de n'importe quel fournisseur du monde entier. Cela accroît la concurrence entre les fournisseurs et, par conséquent, ces derniers offrent des rabais substantiels aux clients.

- Commodité et disponibilité 24/7 :

Les clients peuvent acheter par Internet n'importe quel produit à n'importe quelle heure (Il n'y a pas d'heure d'ouverture et de fermeture pour effectuer les achats) dans n'importe quel endroit dans le monde sans quitter leur lieu de travail ou leur domicile. En raison de certaines circonstances telles que les mauvaises conditions climatiques et la distance à parcourir, les gens

peuvent restreindre leurs achats même si la nécessité se fait sentir. Le commerce électronique offre la commodité d'acheter des biens ou des services sans causer de contraintes physiques aux consommateurs.

- L'interactivité avec le site web marchand

La plupart des plates-formes de commerce électronique donnent le client la possibilité d'ajouter des commentaires sur un produit et voir ce que les autres achètent, ou voir les commentaires des autres clients avant de faire un achat final.

- La baisse des prix

Le commerce électronique réduit le besoin d'intermédiaire de marché et leurs coûts associés, une partie de ces coûts réduits pourrait être répercutée sur les clients sous la forme de prix réduits.

## **3.2.Les inconvénients**

### **3.2.1. Les inconvénients pour l'entreprise**

- Une très grande dépendance sur le site

Pour une entreprise de commerce électronique, son site Web représente la base de leur activité. Même quelques minutes d'indisponibilité peuvent entraîner une perte substantielle d'argent, sans parler de l'insatisfaction du client.

- La nécessité d'une stratégie de logistique inversée

Elle s'appelle aussi logistique des retours, dans le cas de la vente physique, les clients sont généralement disposés à se rendre au point de vente pour retourner ou remplacer des marchandises si nécessaire. Mais recevoir des marchandises est un plus grand cauchemar pour une entreprise de commerce électronique. Même que cette stratégie peut apparaître coûteuse, elle occupe une importance croissante dans les stratégies du e-commerce.

- Résistance des consommateurs

Le client peut ne pas faire confiance au vendeur inconnu du site. Une telle méfiance rend difficile le passage des consommateurs des magasins physiques aux magasins en ligne.

- Problèmes techniques

L'industrie du développement de logiciels continue d'évoluer et évolue rapidement, et parfois, il devient difficile d'intégrer un logiciel de e-commerce avec les applications ou les bases de données existantes.

### **3.2.2. Les inconvénients pour le client**

- Incapacité à découvrir le produit avant l'achat :

Il y a beaucoup de produits que les consommateurs veulent toucher, sentir, entendre, goûter et sentir avant de les acheter. Le commerce électronique enlève ce luxe.

Par exemple : Nous ne pouvons pas toucher le tissu du vêtement que nous voulons acheter, nous ne pouvons pas vérifier comment la chaussure se sent sur nos pieds et également nous ne pouvons pas tester le parfum avant de l'acheter.

- Nécessité d'avoir un dispositif et une connexion Internet :

Le commerce électronique ne peut être traité qu'à l'aide de deux facteurs : Avoir un dispositif d'accès Internet tel qu'un ordinateur, tablette ou un smartphone et une connectivité Internet.

- Problème de retard de marchandises :

Dans le cas des biens numériques, par exemple un ebook, fichier de musique ou bien le règlement d'une facture, le commerce électronique pourrait être plus rapide que l'achat de bien ou de service dans un magasin physique, mais dans le cas de marchandises physiques le commerce électronique offre beaucoup plus de temps pour obtenir ses marchandises entre nos mains. Même avec la livraison express, le plus tôt que nous obtenons les marchandises est généralement "demain". Par exemple, si nous voulons acheter une lampe pour remplacer instantanément celle défectueuse, ou bien si nous voulons acheter un stylo parce que nous avons besoin d'écrire quelque chose en ce moment, nous ne pouvons pas les acheter sur un site de commerce électronique.

- Incapacité d'identifier les arnaques :

Il y a des sites web de commerce électronique qui offrent des produits qu'ils ne possèdent pas, ou bien ceux qui publient des petites annonces immobilières, ou encore pour vendre des voitures à des prix imbattables, mais en réalité ces biens ne sont pas disponibles. Ces escrocs acceptent souvent les commandes et disparaissent.

- La sécurité et la confidentialité des données numériques :

Lorsque nous effectuons un achat en ligne, nous devons fournir au moins les informations de notre carte bancaire et notre adresse de livraison. Dans de nombreux cas, les sites Web de commerce électronique peuvent recueillir d'autres informations sur notre comportement et nos préférences en ligne via l'utilisation des cookies. Cela pourrait nuire à la confiance du client envers le système de paiement en ligne ou, pire encore, au vol d'identité, parce que les sites de commerce électronique et les passerelles de paiement sont susceptibles d'être attaqués par des pirates et, par conséquent, l'identité du client est menacée.

- La qualité :

Dans le cas d'un achat en ligne nous ne pouvons pas toucher ou tenir physiquement le produit que nous achetons. Nous devons faire confiance à ce que l'entreprise nous a montré. Malheureusement, Il y a des sites web de commerce électronique qui offrent des produits avec des descriptions incorrectes, des images ne correspondent pas ou sont de mauvaise qualité. C'est la raison pour laquelle la qualité n'est pas assurée.

## **4. Les méthodes de paiement**

Les méthodes de paiement facilitent l'échange de biens et de services et répondent à des besoins spécifiques. Chaque méthode a sa propre histoire sociale et technologique qui oriente son utilisation dans des domaines spécifiques. Aujourd'hui, les banques offrent un grand nombre de moyens liés au traitement automatique des transactions et à la dématérialisation progressive des supports monétaires. Les moyens utilisés varient d'un pays à l'autre. En général, les moyens de paiement prennent les formes suivantes :

- Espèces (sous forme de pièces métalliques ou de billets)
- Chèques
- Virements
- Débits directs
- Transferts interbancaires

- Cartes de paiement (débit ou crédit)

Avec l'essor des nouvelles technologies en général et le commerce électronique en particulier, des nouveaux moyens de paiement sont apparus et qui sont basés sur la dématérialisation de l'argent. La monnaie dématérialisée est souvent stockée dans des cartes intelligentes (puces) ou dans des porte-monnaie électroniques ou virtuels.

Quel que soit le moyen de paiement choisi pour régler les achats sur Internet, plusieurs paramètres de sécurité doivent être pris en considération pour assurer le lien entre consommateur, marchand et acquéreurs de paiements (passerelle de paiement, banques) afin de garantir un niveau plus élevé de la sûreté des transactions.

La sécurité des transactions en ligne comprend en générale les services suivants [15] :

- **La confidentialité** : se définit comme « le fait de s'assurer que les messages échangés entre deux parties ne sont pas divulgués à un tiers non autorisé ». Dans certaines applications, la confidentialité des adresses peut également être nécessaire pour empêcher l'analyse des modèles de trafic et la dérivation des informations secondaires qui pourraient être utilisées.

Pour assurer la confidentialité, le client utilise une clé de session secrète qui sera cryptée avec la clé publique de la banque. La réponse de la banque, qui comprend les pièces signées à l'aveuglette, est signée avec la clé privée de la banque puis cryptée avec la clé de session.

- **L'intégrité des données** : Ce service garantit que les données reçues par l'acheteur, le vendeur et les acquéreurs de paiements sont exactement celles qui ont été transmises par l'expéditeur et qu'elles n'ont pas été corrompues, intentionnellement ou par erreur, pendant leur transfert sur le réseau.

- **La non-répudiation** : est le service qui offre une preuve irréfutable de l'intégrité des données et de leur origine d'une manière qui peut être vérifiée par un tiers, par exemple la non-répudiation que l'expéditeur a envoyé le message ou qu'un récepteur a reçu le message. Ce service peut aussi être appelé authentification de l'origine des données.

- **La fiabilité** : Du point de vue de consommateur final, la fiabilité du système dans son ensemble dépend de l'atomicité des transactions ; c'est-à-dire qu'une transaction doit avoir lieu dans son intégralité ou pas du tout. Aucun acheteur ne devrait être contraint de payer pour une

transaction interrompue. Pour cela le système de paiement doit être pourvu de mesures nécessaires en cas de mauvais fonctionnement.

#### **4.1.Les méthodes de paiement en ligne**

Les méthodes de paiement en ligne se réfèrent à la manière dont les acheteurs peuvent payer leurs achats sur internet. Le choix de la méthode de paiement en ligne est présenté à la phase de caisse ou sur la page de paiement de la passerelle de paiement en ligne, ces méthodes de paiement doivent être clairement visibles par l'acheteur au moyen des logos bien connus (par exemple : *Verified by Visa ou MasterCard SecureCode, PayPal, iDEAL*), ou bien par des mots universels communs comme « cartes de crédits », « virement bancaire » ou « paiement à la livraison ». Les méthodes de paiement en ligne reposent sur différents instruments de paiement, et bien qu'il existe une large gamme sur le marché (Les espèces, les cartes de paiement, les chèques, les virements, les débits directs et également les transferts interbancaires), il y'a ceux qui sont utilisables uniquement sur internet à savoir les cartes virtuelles, les comptes virtuels et la monnaie numérique. En règle générale, les méthodes de paiement en ligne font référence aux marques de méthodes de paiement en ligne (Visa, MasterCard, Maestro, American Express, CMI), aux marques de solutions de paiement en ligne (par exemple *Paypal, Klarna, MasterPass*) ou directement à l'un des instruments de paiement. Les méthodes de paiement les plus utilisées dans le e-commerce sont les suivantes :

**Paiement par cartes de paiement :** Regroupe les cartes de débit, les cartes de crédit, les cartes à puce. Elles sont largement utilisées à l'échelle internationale et bénéficient d'un statut d'être largement accepté comme mode de paiement commun.

Le traitement d'une transaction par carte de paiement en ligne implique deux processus généraux, l'acceptation du paiement et la compensation de la transaction. L'acceptation de paiement comprend les étapes nécessaires pour déterminer que la carte est valide et que la transaction ne dépassera pas la limite de crédit qui pourrait exister pour la carte. La compensation de la transaction comprend toutes les étapes nécessaires pour transférer les fonds du compte bancaire du titulaire de la carte vers le compte bancaire du commerçant.

**La monnaie virtuelle :** Bien que les cartes de paiement (crédit ou débit) dominent aujourd'hui les paiements en ligne, la monnaie virtuelle est prometteuse pour l'avenir. La monnaie virtuelle appelée aussi « crypto monnaie » est un terme général qui décrit à la fois un



système de paiement décentralisé (de pair à pair) et tout système de stockage et d'échange de valeurs créé par une entité privée (non gouvernementale) qui n'utilise pas de documents papier ou de pièces de monnaie et qui peut servir de substitut à la monnaie physique émise par le gouvernement.

Ce type de paiement est apparu en 2009 et une dizaine de million de personnes l'utilisent déjà comme moyen de paiement pour échanger des biens et des services. Le bitcoin est la toute première monnaie virtuelle, créée en 2009, il n'appartient à aucun état, et n'est régulé par aucune banque centrale, il est géré par ces propres utilisateurs en réseau et il est beaucoup moins vulnérable face aux attaques informatiques.

Jusqu'à ce jour, le paiement en ligne par la monnaie virtuelle n'est pas encore devenu universel en raison que chaque émetteur de ce type de monnaie crée ses propres normes. En d'autres termes, un système de trésorerie électronique efficace aura besoin de normes communes pour que la monnaie virtuelle d'un émetteur puisse être acceptée par un autre émetteur. Un classement des crypto-monnaies en fonction de leur valorisation boursière réalisé par Coinmarketcap.com en août 2017 a classé le Bitcoin, L'Ethereum et Ripple parmi les trois premières monnaies cryptographiques décentralisées les plus importantes (respectivement 54.117, 26.920 et 6.977 milliards de dollars).

**Les portefeuilles électroniques :** Ce sont des services de paiement en ligne permettent de payer des achats en ligne, d'effectuer et de recevoir des virements. Cette méthode de paiement est liée à un ou plusieurs cartes de paiement ou directement à un compte bancaire qui représentent la source de paiement. Les portefeuilles électroniques rendent la phase de paiement plus simple, plus fluide, et surtout plus sécurisé, car l'utilisation de ces derniers ne nécessitent pas de communiquer les informations financières (numéro de carte, sa date de validité, ainsi le cryptogramme) à chaque opération de paiement au site marchand. Ces données ne sont en effet demandées que lors de la création du portefeuille électronique. Par la suite, l'utilisateur doit uniquement saisir ses identifiants (nom d'utilisateur et mot de passe) pour réaliser des transactions en ligne.

Le service de paiement en ligne PayPal représente le portefeuille électronique le plus répandu, il est accepté par des millions de marchands dans le monde et constitue le mode de paiement le plus utilisé sur les géants du commerce électronique tels qu'eBay, Amazon et Alibaba.

### **Paiement en espèces à la livraison / Paiement en espèces dans un magasin :**

Selon [16], 2.5 milliards de personnes adultes dans le monde n'ont pas accès aux produits bancaires formels. La grande majorité de ces personnes (près de 2,2 milliards) non desservis sont dans les pays en voie de développement (Asie, Afrique, Amérique latine et au Moyen-Orient). Afin de vendre des biens et des services en ligne aux consommateurs de ces régions, il est essentiel d'offrir des paiements en espèces.

Nous distinguons deux méthodes couramment utilisées pour les paiements en espèces, les paiements à la livraison et les paiements dans les points de paiement physiques.

**Paiement en espèces à la livraison :** Est un type de transaction dans laquelle l'acheteur effectue le paiement d'un bien au moment de la livraison. Si l'acheteur n'effectue pas de paiement lors de la livraison du bien, ce dernier sera retourné au vendeur. L'acheteur peut effectuer le paiement en espèces ou par chèque certifié, selon ce que stipule le contrat d'expédition. Dans ce type de transaction, la livraison des biens est habituellement effectuée par l'intermédiaire des compagnies de transport (exemple : FedEx, UPS, DHL,DPD, ...) l'acheteur remet donc le paiement à la compagnie de transport, qui relaie ensuite le paiement au vendeur. Selon le rapport mondial de paiement de WorldPay de novembre 2015 [17], 6,5% du chiffre d'affaires mondial du commerce électronique est payé en espèces à la livraison.

Le paiement à la livraison présente un faible risque pour l'acheteur puisqu'il ne paie pas jusqu'à ce qu'il reçoive le produit dans un état satisfaisant, il protège également l'acheteur contre les escroqueries et les fraudeurs en ligne qui se présentent souvent comme des entreprises légitimes pour collecter des informations de carte de crédit à des fins illégales. Du côté commerçant, offrir une option de paiement à la livraison peut améliorer la confiance des acheteurs surtout dans le cas d'une nouvelle entreprise qui n'a pas encore une forte reconnaissance dans le marché. Cela contribue à augmenter le nombre de clients et à booster les ventes.

**Paiement en espèces dans un magasin:** Le principe est très simple, lorsque l'acheteur utilise cette méthode de paiement, il obtient un code unique (souvent un code barre) qui lui permet de régler son achat dans un des points de paiement physiques proposés par le site marchand (ex :

YesByCash, SisalPay,HiPay, Binga...) et dès que l'enregistrement de paiement est effectué dans ces points de paiement, le vendeur sera informé des paiements et par la suite la commande sera préparée et livrée selon le mode de livraison choisi par l'acheteur.

#### **4.2.La sécurité des paiements en ligne:**

Il existe deux mécanismes principaux pour la sécurité des transactions en ligne, le premier basé sur l'utilisation de deux approches de chiffrement qui servent à établir des connexions sécurisées entre les serveurs Web et les clients, en chiffrant les informations sensibles fournies par les clients sur un site web telles que le mot de passe et le numéro de la carte bancaire. Ces deux approches reposent sur le système SSL (Secure Sockets Layer) développé par Netscape Communications et le protocole S-HTTP (Secure Hypertext Transfer Protocol) développé par Commerce Net. SSL et S-HTTP permettent aux dispositifs utilisés par les clients (Ordinateurs, Tablettes, Smartphone...) et aux serveurs de gérer les activités de chiffrement et de déchiffrement entre eux au cours d'une session Web sécurisée ; Cependant, SSL et S-HTTP ont des objectifs différents. SSL sécurise les connexions client/serveur, et S-HTTP établit la sécurité pour chaque message individuel.

Même si qu'au départ le système SSL n'était pas conçu pour être un protocole de sécurisation de paiement, il est devenu jusqu'à aujourd'hui le fondement de la plupart des transactions de commerce électronique sécurisées. Du cryptage des informations de paiement sensibles à la vérification de l'authenticité d'un site Web, SSL fournit un canal de communication sécurisé pour le commerce électronique via l'utilisation des certificats numériques.

Un autre protocole de sécurisation de transaction électronique SET (Secure Electronic Transaction) qui s'appuie sur le standard SSL a été développé par Visa et MasterCard en 1996. Il est destiné spécialement à sécuriser les paiements électroniques par carte bancaire et le principal motif derrière le développement de ce protocole était la prévention de la fraude. En plus de chiffrer les transactions, SET fournit des mesures d'authentification rigoureuses. Dans le cas de SSL, le certificat d'authentification est fourni uniquement par le serveur avant de sécuriser le canal de communication. Cependant, après que le canal de communication est sécurisé, seulement le commerçant peut vérifier l'authenticité du consommateur grâce à ses informations de carte bancaire, ce qui signifie que les cartes bancaires volées peuvent être utilisées. SET répond à cette lacune de SSL car, en plus des certificats de serveur des commerçants, les

consommateurs sont également tenus d'obtenir des certificats pour empêcher le vol d'identité par l'utilisation d'un système complémentaire de type 3D-Secure qui amené l'acheteur à saisir un code secret lors de chaque opération de paiement en ligne avant de pouvoir procéder à la demande de l'autorisation de paiement pour s'assurer que la carte est utilisée par son véritable porteur. Malgré ses atouts, SET n'a pas réussi à obtenir le même niveau d'acceptation que SSL à cause de la complexité et du coût supporté par le commerçant.

Le deuxième mécanisme fait appel à un intermédiaire « tiers de confiance » qui met à disposition des sites marchands un système de paiement sécurisé. Lors de la phase de paiement, le navigateur de client est redirigé vers la page de paiement de serveur du tiers de confiance sur laquelle le client sera invité à saisir le détail de ses informations financières. Généralement après vérification auprès de la banque ayant délivré la carte bancaire de l'acheteur, via le réseau interbancaire, la passerelle de paiement renvoie un code de retour positif au site marchand, lui indiquant que le paiement s'est effectué correctement. Le passage par une passerelle de paiement « tiers de confiance » offre un plus haut niveau de confidentialité aux consommateurs que d'effectuer des paiements directement sur le site marchand, car le système de la passerelle de paiement crypte les informations de paiement afin que seule la banque ait accès aux données financières des clients et non au marchand. Le marchand ne voit que les détails de la commande sous une forme non cryptée et transmet les instructions de paiement dans une demande d'autorisation à la passerelle de paiement pour authentification. Dans le cas classique, les données des clients telles que les informations de compte de carte bancaire sont traitées et stockées dans les bases de données des commerçants et donc ces derniers ont accès aux données financières des clients.

Du côté marchand, le système de tiers de confiance libère le vendeur de toutes complexités techniques et les conformités par rapport aux normes et standard de paiement et de sécurité et offre aux marchands un système de paiement sécurisé, certifié et évolutif.

## **5. Le commerce électronique au Maroc : Quelques chiffres**

Une étude détaillée menée par [18] en 2015 sur le commerce électronique en Afrique, a montré que parmi les pays étudiés (Maroc, Tunisie, Sénégal, Cote d'ivoire), le Maroc est le plus avancé dans le domaines de commerce électronique. Ceci s'explique par plusieurs facteurs :

- La disposition d'une infrastructure considérable qui permet au Maroc d'aller loin dans le domaine de l'économie numérique
- La naissance de la société Maroc Telecommerce (MTC) en 2001, 1ère plateforme marocaine de paiement via Internet et le premier opérateur de paiement multicanal.
- Le lancement du Centre Monétique Interbancaire (CMI) en 2004, acquéreur national de transactions monétiques.
- Le Maroc compte selon les derniers rapports de l'Agence nationale de réglementation des télécommunications (ANRT) : Un parc de plus de 18 millions de smartphones (à fin juin 2017) (près de deux marocains sur trois disposent d'un smartphone) [19].
- La baisse des prix des abonnements haut débit et le lancement du service 4G.
- Une croissance annuelle de 29% des abonnés internet (19,2 millions d'abonnés internet à fin juin 2017).
- Le taux de pénétration de l'Internet a atteint 55% de la population
- Près de 12% des individus ont utilisé internet pour effectuer des achats en ligne.

Tous ces facteurs soutiennent de manière cohérente la croissance du commerce électronique au Maroc.

### **5.1.Les acteurs du commerce électronique au Maroc**

Comme dans la plupart des pays, le marché du commerce électronique au Maroc est régi essentiellement par plusieurs organismes et dépend de plusieurs acteurs, privés et publiques.

**Les sites en ligne :** Comprend les sites marchands dans lesquels Jumia, Hmizate, Hmall, microchoix et superdeal représentent les principaux acteurs locaux et les sites des facturiers qui permettent aux particuliers/entreprises de régler les factures périodiques (Electricité, eau, téléphone, taxes, impôts, etc...).

**Le Prestataire de Service de Paiement (PSP) :** C'est un acteur privé, qui joue un rôle très important dans le processus d'une transaction en ligne. Au Maroc, la société Maroc Telecommerce (MTC) représente le 1er PSP qui a démarré ses activités sur le marché en 2001 à l'initiative des banques marocaines et d'investisseurs dans les TIC. Le développement rapide du commerce électronique a favorisé l'émergence de nouveaux concurrents à savoir AmanPay, PayZone et F-Pay. La mission principale de ces PSP est de sécuriser le processus de paiement, en mettant à la disposition des entreprises une infrastructure technologique sécurisée et évolutive, comprenant les dernières avancées technologiques pour assurer le lien entre commerçant, consommateur et acquéreurs de paiements.

**Le Centre Monétique Interbancaire (CMI) :** C'est un établissement privé qui a été fondé en 2004 par neuf banques marocaines qui en sont les actionnaires, et ayant pour vocation d'améliorer la monétique au Maroc. Il est connecté au réseau interbancaire national, aux banques internationales et aux PSP du marché (MTC). Parmi ces principales missions : la mise en place des solutions de paiement électroniques par carte bancaire auprès des commerçants et des sites marchands, le traitement des données de transactions électroniques, la gestion du flux financiers résultant de ces transactions en demandant à la banque du client de débiter le compte de ce dernier et crédite le compte bancaire du marchand et la lutte contre les risques de fraude liés à l'utilisation des cartes bancaires. En effet, depuis octobre 2016, le CMI est devenu l'actionnaire majoritaire de la société MTC.

**Les banques :** Comprend la banque du client et celle du marchand, elles sont connectées entre eux et avec le CM via le réseau interbancaire. Lors d'une demande de paiement, la banque du client vérifie et contrôle l'éligibilité de la carte du client (solde suffisant, carte non bloquée) et décide d'autoriser ou non le paiement. L'opération prend quelques instants et si elle est approuvée, la banque effectue la transaction et transfère le paiement sur le compte du marchand. En cas de refus, le consommateur est informé par le CMI que la transaction ne peut pas être terminée et que sa demande est annulée.

## **5.2.Les chiffres clés du e-commerce au Maroc (au 30 septembre 2018)**

Ces dernières années, le commerce électronique au Maroc est en forte progression. Les dernières statistiques du Centre monétique interbancaire (CMI) confirment cette donnée.

Suite à la publication de son rapport sur l'activité monétique au 30 septembre 2018, le Centre Monétique Interbancaire (CMI) poursuit ses actions de sensibilisation à l'usage des cartes bancaires, notamment pour le e-commerce. Selon le rapport en question les sites marchands et sites des facturiers affiliés au CMI ont réalisé 6,3 millions d'opérations de paiement en ligne via cartes bancaires, marocaines et étrangères, pour un montant global de 2,4 milliard de DH durant la période des 9 premiers mois, en progression de +30,4% en nombre et +21,1% en montant par rapport à la même période en 2017.

L'activité des cartes marocaines a évolué de +30,1% en nombre d'opérations, en passant de 4,7 millions de transactions durant la période des 9 premiers mois 2017 à 6,7 millions de transactions durant la période des 9 premiers mois 2018, et de +19,9% en montant, en passant de 1,8 milliard de DH durant la période des 9 premiers mois 2017 à 2,2 milliards de DH durant la période des 9 premiers mois 2018.

L'activité des cartes étrangères a évolué de +47,9% en nombre d'opérations, en passant de 84.711 transactions durant la période des 9 premiers mois 2017 à 125.300 transactions durant la période des 9 premiers mois 2018, et a progressé de +35,5% en montant, en passant de 147,5 millions de DH durant la période des 9 premiers mois 2017 à 199,9 millions de DH durant la période des 9 premiers mois 2018.

L'activité reste très fortement dominée par les cartes marocaines à hauteur de 98,0% en nombre de transactions et de 91,7% en montant [20].L'internaute marocain semble donc se prendre de plus en plus au jeu du paiement en ligne.

1.726 contrats ont été signés à ce jour, avec les organismes certificateurs que sont Maroc Telecommerce (MTC) et le Centre monétique interbancaire(CMI). Ils étaient 140 à fin 2010. Au final, ce sont 1.726 sites e-commerce qui sont actuellement actifs et référencés par CMI, dont 283 nouveaux en 2016.

La table 1.1 récapitule les chiffres clés de paiement via Internet entre 2014 et 2017 (source CMI [20]).

| Périodes            | Cartes Marocaines |               | Cartes Etrangères |              | Cartes Marocaines & Etrangères |               |
|---------------------|-------------------|---------------|-------------------|--------------|--------------------------------|---------------|
|                     | Nombre            | Montant       | Nombre            | Montant      | Nombre                         | Montant       |
| <b>9 MOIS 2015</b>  | 1 699 445         | 893 157 327   | 66 419            | 100 707 346  | 1 765 864                      | 993 864 673   |
| <b>9 MOIS2016</b>   | 2 512 476         | 1 190 235 685 | 83 440            | 118 194 352  | 2 595 916                      | 1 308 430 037 |
| <b>9 MOIS2017</b>   | 4 745 236         | 1 833 795 191 | 84 711            | 147 546 421  | 4 829 947                      | 1 981 341 612 |
| <b>9 MOIS 2018</b>  | 6 172 064         | 2 198 908 280 | 125 300           | 199 877 347  | 6 297 364                      | 2 398 785 627 |
| <b>Var. 18 / 17</b> | <b>30,1%</b>      | <b>19,9%</b>  | <b>47,9%</b>      | <b>35,5%</b> | <b>30,4%</b>                   | <b>21,1%</b>  |

Figure 1.4 – Les chiffres clés de paiement via Internet 2014-2017.

Et selon le bilan annuel du e-commerce publié par le moteur de recherche Vendo.ma nous trouvons :

#### **TOP 5 des catégories recherchées en 2016**

1. Mode & accessoires
2. High Tech
3. Deals et Offres
4. Hôtels et Voyages
5. Electroménagers

#### **TOP 10 des boutiques les plus cliquées en 2016**

6. Jumia
7. Hmall.ma
8. zara
9. boutika.ma
10. Bershka
11. booking.com
12. LinksSolutions
13. LaRedoute
14. PULL&BEAR
15. bestmark

### **5.3.Cinématique de paiement en ligne avec CMI**

La cinématique de paiement réalisée avec CMI correspond à la cinématique d'un paiement à redirection.

L'utilisateur au cours de la transaction sera redirigé vers une plate-forme technologique internationale « *Nestpay* » du CMI afin de choisir sa carte de paiement (Visa, Mastercard,



Maestro ou cmi) et de renseigner son numéro de carte bancaire, la date d'expiration et le code de vérification ensuite CMI s'occupe de la gestion des échanges monétaire électronique entre les consommateurs, les e-commerçants et les établissements bancaires. La figure 1.5 montre le schéma d'une transaction e-commerce avec CMI.

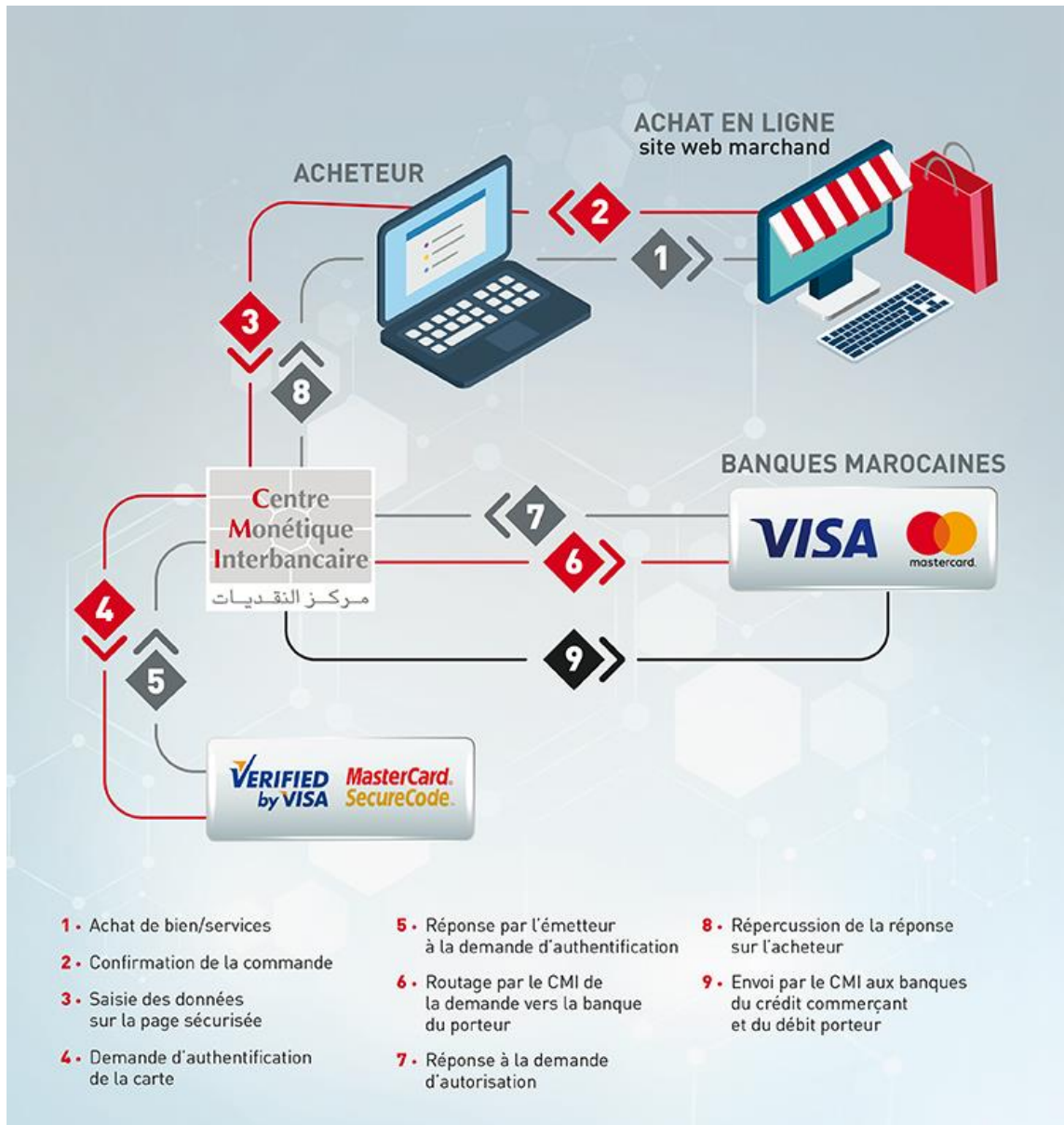


Figure 1.5 - Schéma d'une transaction e-commerce par carte de paiement (source CMI [20]).

## **5.4.Sécurisation des transactions et gestion de la fraude par CMI**

Le Centre Monétique Interbancaire (CMI) met en place des outils efficaces afin de détecter et prévenir les tentatives de fraude lors des transactions e-commerce, et de garantir un paiement en ligne sécurisé.

Pour se faire le CMI offre :

- Une surveillance en temps réel de l'ensemble des paiements électroniques qui transitent sur ses plateformes.
- Des règles anti-fraudes performantes aux banques et aux commerçants. Ainsi, des alertes qui sont mises en place selon des seuils précis, alignés sur les standards VISA et MASTERCARD.
- Une plateforme certifiée PCI DSS (Payment Card Industry Data Security Standard), permet de garantir le traitement, le stockage et la transmission des données des propriétaires de cartes. Le standard PCI-DSS a été créé en 2006 par les grandes marques de cartes, à savoir : Visa, MasterCard, American Express, Discover et JCB, regroupées ensuite en un consortium fondé en 2006 : le PCI SSC (Payment Card Industry Security Standards Council), il est basé sur les programmes de sécurité Visa AIS (Account Information Security) et MasterCard SDP (Site Data Protection). Ceux-ci comprennent des mesures de sécurité extrêmement complexes et coûteuses qui sont nécessaires pour protéger la sécurité des données des sociétés émettrices de cartes de paiement et qui doivent être vérifiées et auditées à intervalles réguliers.
- 3D Secure : C'est un système de paiement authentifié pour améliorer la sécurité des transactions en ligne et encourager la croissance des paiements de commerce électronique, et qui consiste à s'assurer, lors de chaque paiement en ligne, que la carte est utilisée par son véritable titulaire en demandant une authentification supplémentaire sous forme d'un code dynamique à usage unique. 3D Secure est appelé "Verified by Visa" chez Visa, et "SecureCode" chez Mastercard.

## Conclusion

Dans ce chapitre, nous avons dans un premier temps introduit le concept de commerce électronique, ses différentes définitions, son histoire, son évolution, ses différentes formes, ses avantages et ses inconvénients pour l'entreprise et le client. Nous avons aussi cité et décrit le rôle de chaque acteur impliqué dans les transactions électroniques et par la suite nous avons détaillé les étapes entrant dans le processus de transaction électronique. Nous avons constaté que les débuts et l'évolution du commerce électronique sont liés avec l'évolution et la généralisation d'Internet et qu'il peut prendre plusieurs formes selon la nature des acteurs mis en relation via l'internet.

Nous avons vu également que le commerce électronique offre plus d'avantages que d'inconvénients pour les marchands et les clients. Cette multitude d'avantages justifient l'évolution fulgurante qui a connu le secteur du commerce électronique partout dans le monde.

Nous avons ensuite présenté les principales méthodes de paiement en ligne, en soulignant que des nouveaux moyens de paiement basé sur la dématérialisation de l'argent sont apparus avec l'émergence du commerce électronique comme la monnaie virtuelle et les portefeuilles électroniques. Enfin, nous avons abordé la partie essentielle du commerce électronique qui est la sécurité des paiements en ligne, dans laquelle nous avons décrit les deux principaux mécanismes utilisés pour sécuriser l'acte d'achat sur internet et assurer le lien acheteur, commerçant et acquéreurs de paiements.

Dans un deuxième temps, nous avons jeté la lumière sur l'état des lieux du commerce électronique au Maroc ainsi sur les organismes professionnels qui le représente. En effet, les chiffres clés pour 2017 indiquent clairement que le commerce électronique est en pleine explosion au Maroc et que l'internaute marocain est devenu plus familiariser avec les systèmes de paiement en ligne.

Nous verrons dans le chapitre 2 une présentation des systèmes CRM (Customer Relationship Management) et plus particulièrement, ceux qui sont dédiés au secteur de commerce électronique, et qui sont indispensables à la réalisation de ses résultats souhaitables.

# **Chapitre 2 : Gestion de la relation client à l'ère du commerce électronique**

## **Introduction**

Dans ces dernières années, le commerce électronique a connu une forte croissance dans le monde entier. Et comme tout secteur en plein essor tels que, les secteurs de la banque, de l'assurance, de télécommunication et d'automobile, la croissance et la prospérité du commerce électronique l'amène à se confronter à des problématiques et des défis dont la plupart sont reflétés dans la capacité de l'entreprise à gérer au mieux ses relations avec ses clients existants et potentiels. Dans le monde d'aujourd'hui, la richesse d'une entreprise se trouve dans ses clients, la raison pour laquelle la compréhension des attentes des clients et l'anticipation de leurs besoins devient un objectif majeur de nombreuses entreprises qui souhaitent accroître la rentabilité et la fidélité de leurs clients. Par conséquent, un système de gestion de la relation client (Customer Relationship Management ou CRM) dont les objectifs sont de trouver, d'attirer, d'acquérir de nouveaux clients, de fidéliser les clients existants et de les retenir dans l'optique d'augmenter le chiffre d'affaire et les bénéfices, est devenu très urgent surtout dans un secteur qui connaît une forte concurrence comme le secteur du commerce électronique.

Avec l'explosion d'Internet et l'émergence de commerce électronique, une nouvelle philosophie de gestion de la relation client est apparue : la gestion électronique de la relation client e-CRM. La matière première de l'e-CRM est constituée de données sur les clients et sur leurs comportements d'achat, qui sont collectées à travers les sites e-commerce des marchands. En effet, le grand volume de données provenant des sites e-commerce exige l'utilisation des outils et des techniques afin de découvrir des connaissances cachées dans ces données, cela permet à l'entreprise d'exploiter par la suite ces connaissances dans la prise de décisions et également dans le pilotage de la fonction marketing. Parmi ces outils et techniques, les techniques fouille de données (data mining) représentent les techniques les plus efficaces pour atteindre ces objectifs [21].

## 1. Gestion de la relation client

### 1.1. Définition et objectifs de la CRM

Au cours de la dernière décennie, les universitaires et les professionnels ont manifesté un intérêt croissant pour la gestion de la relation client. Il reste, cependant, une certaine divergence sur ce qu'est vraiment le CRM en termes de définition, et sur la façon avec laquelle une stratégie CRM devrait être développée. Ceci dépend en partie des différentes solutions proposées par les éditeurs de logiciels ou les intégrateurs de systèmes et que chaque prestataire n'associe que des aspects particuliers du CRM à sa gamme de produits et aux technologies qu'il utilise. Bien qu'il n'existe pas de définition unanime de CRM, nous pouvons identifier un certain nombre d'attributs de base de CRM et les intégrer dans une définition précise afin de fournir un aperçu général des différentes sections du CRM et de ses objectifs. Dans un contexte plus large et stratégique, CRM peut être donc défini comme « une stratégie commerciale qui intègre les processus et les fonctions internes, ainsi que les réseaux externes en s'appuyant sur la technologie et sur des méthodes scientifiques pour identifier de nouveaux clients, construire des relations à long terme avec eux, les fidéliser en répondant à leurs besoins et enfin réduire le taux d'attrition tout en réduisant les coûts de marketing et de services à la clientèle ».

Les années 90 marquent le début de l'ère du client, dans laquelle les entreprises connaissent une profonde mutation dans leurs stratégies de marketing, elles sont passées d'une stratégie centrée produit (amélioration de leurs produits, de leur fonctionnement interne), à une stratégie centrée client. Le concept de marketing relationnel est devenu l'approche majeure de marketing, dans laquelle l'accent est mis sur l'attraction, la rétention et l'amélioration de la relation client [22]. La richesse d'une entreprise provient, donc, en grande partie de sa capacité à identifier, acquérir et à fidéliser ses meilleurs clients. Selon [23], ces trois principales motivations représentent les trois phases du cycle de vie du client.

Nous pouvons donc conclure que l'idée du CRM est qu'il aide les entreprises à améliorer l'efficacité de leurs stratégies de marketing relationnel en combinant la technologie de l'information et le marketing relationnel afin de bien comprendre les attentes des clients et d'anticiper de leurs besoins pour les fidéliser et d'augmenter leurs rentabilités [24]. C'est à cet objectif que souhaite répondre tous système de gestion de la relation client [25]. Un système de CRM efficace peut augmenter les revenus d'une entreprise en :

- Fournir des services et des produits qui correspondent exactement à ce que veulent ses clients
- Offrant un meilleur service à la clientèle
- Fidéliser les clients existants et en découvrir de nouveaux
- Traverser les produits plus efficacement
- Aider les vendeurs à conclure leurs affaires plus rapidement
- Diminuer les coûts globaux
- Simplifier les processus de marketing et de vente

## **1.2. Les trois dimensions de la CRM**

Compte tenu de la diversité des définitions du CRM, il n'est pas surprenant qu'il y ait eu plusieurs tentatives pour produire des taxonomies qui reconnaissent différentes dimensions de CRM. Un premier effort a identifié trois dimensions différentes de CRM : le CRM opérationnel, collaboratif et analytique [26]. La collaboration de ces trois parties a pour finalité de réaliser l'objectif ultime de la stratégie CRM de l'entreprise, c'est-à-dire le développement et l'optimisation de la relation client.

### **1.2.1. Le CRM opérationnel**

Le CRM opérationnel est centré sur la gestion quotidienne de la relation avec le client, il comprend les processus d'activités et les technologies qui peuvent aider à améliorer l'efficacité et l'exactitude des opérations de vente, de marketing et de service client. Toutes les interactions avec le client sont enregistrées dans l'entrepôt de données de l'entreprise à des fins d'analyse et de suivi.

### **1.2.2. Le CRM collaboratif**

Le CRM collaboratif se concentre sur l'interaction avec les clients (interaction personnelle, sites web, centre d'appel, courrier individuel et ventes directes, etc.) Le CRM collaboratif traite de la synchronisation et de l'intégration des interactions clients et des canaux de communication tels que téléphone, email, fax, web etc. dans le but de référencer les clients d'une manière cohérente et systématique. L'idée est non seulement d'améliorer les interactions, mais aussi d'augmenter et d'améliorer la fidélisation et la liberté des clients.

Aussi le CRM collaboratif a pour objectif d'intégrer les différentes parties internes de l'entreprise (département de vente, de marketing, de support technique), et les parties prenantes externes (fournisseurs, vendeurs et distributeurs) afin de partager les informations utiles collectées à partir des interactions avec les clients tout au long de la chaîne d'approvisionnement pour mettre en évidence une meilleure compréhension des clients. Par exemple, les informations sur les produits préférés pourraient être partagées avec le département marketing afin que l'analyse puisse être effectuée dans cet aspect pour fournir des produits préférés aux clients.

Certains éditeurs de technologie CRM ont développé des applications de gestion des relations avec les partenaires (PRM) qui permettent aux entreprises de gérer des écosystèmes de partenaires ou de canaux complexes et de réduire les coûts de gestion des partenaires ou des canaux.

### **1.2.3. Le CRM analytique**

Dans le CRM analytique, les données relatives aux clients générées en bonne partie par le CRM opérationnel et collaboratif sont analysées par les techniques d'exploration de données (data mining) qui englobent des techniques avancées d'apprentissage statistique et d'apprentissage automatique, de text mining et de techniques d'exploration de données Web. L'analyse de ces données est exploitée pour améliorer le service client en découvrant des informations commerciales cachées qui ne peuvent être obtenues autrement par des règles commerciales standard [27,28].

Cette très vaste quantité de données est stockée généralement dans l'entrepôt de données de l'entreprise (data warehouse), et comprend généralement des données de vente (historique des achats), données financières (historique des paiements, crédit), données marketing (réponse de la campagne, données du programme de fidélité) et des données de service. Autres données provenant de sources externes peuvent être ajoutées, et avec l'application d'outils de fouille de données (data mining) et des outils de business intelligence dans l'analyse de données, l'entreprise peut par exemple diviser les clients en groupes homogènes, construire des profils et créer des modèles de comportement en évaluant un certain nombre de paramètres tels que la loyauté et la rentabilité, modéliser la valeur à vie du client, adapter une commercialisation ciblée, recommander de produits et de services par analyse de panier de marché, détection de fraude, prévenir le désabonnement des clients (churn). En bref, le CRM analytique transmet le résultat

de l'analyse de données (c'est-à-dire les connaissances acquises) aux technologies CRM opérationnel et collaboratif, en d'autres termes, les décisions commerciales prises dans le contexte analytique sont ensuite réalisées dans la dimension opérationnelle, où les opérations ont lieu. La figure 2.1 illustre les interactions existantes entre les trois dimensions de CRM.

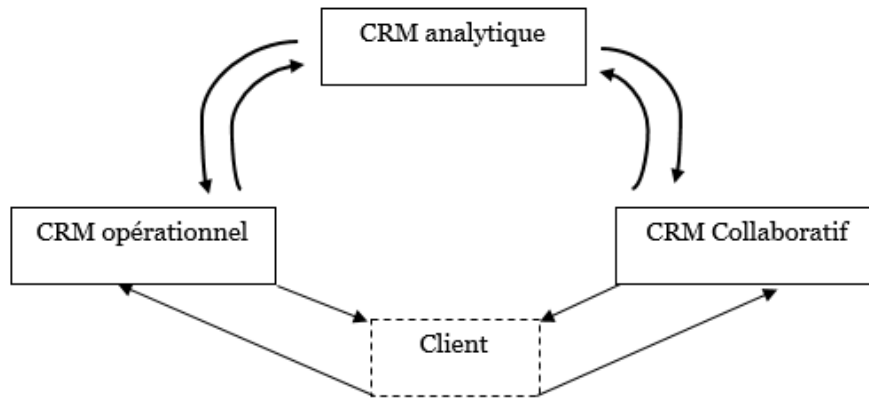


Figure 2.1 - Les interactions entre les trois dimensions du CRM

Notre travail se focalisera essentiellement sur l'analytique et spécifiquement à la fonction des techniques de data mining dans les environnements e-commerce.

### 1.3.Cycle du CRM

Comme il est mentionné ci-dessus l'objectif principal du CRM est d'exploiter le capital client en utilisant des stratégies, des processus et de la technologie pour répondre de manière efficace aux besoins des clients afin de les fidéliser, ce qui permet ensuite d'augmenter le chiffre d'affaire et les bénéfices. Pour atteindre ces objectifs, des activités comme la collecte, le stockage, l'analyse de données clients, la restitution des informations pertinentes et la construction des stratégies de marketing et de rétention sont indispensables. Ces activités sont réalisées au cours des trois phases principales qui constituent les principales étapes du cycle de vie du client : Phase d'acquisition, phase de fidélisation et la phase d'extension. La gestion de la relation client s'articule donc autour de la gestion du cycle de vie du client.



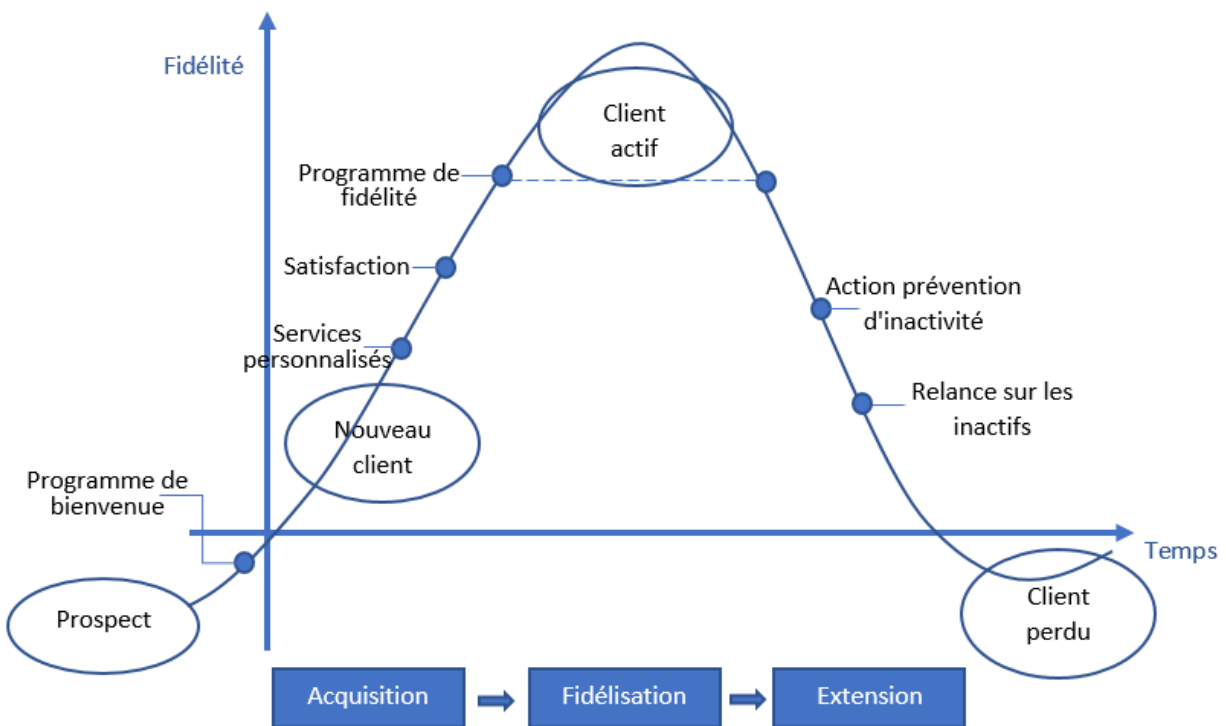


Figure 2.2 - Schéma du cycle de vie d'un client

Le cycle de vie du client (Figure 2.2) est une représentation des étapes que le client traverse dans leur relation avec une entreprise. A l'aide des technologies CRM les entreprises développent des stratégies et des processus pour déplacer leurs clients à travers ces trois phases. Ces stratégies et processus déterminent comment les entreprises identifient et acquièrent de nouveaux clients, augmentent leur valeur pour l'entreprise et les gardent à long terme.

**Phase d'acquisition :** L'acquisition des clients a toujours été la première étape importante dans l'établissement de relations d'affaires, elle commence soit par la publicité traditionnelle, soit par le biais des recommandations des clients. Ainsi acquérir les clients est toujours l'objectif le plus important lors du lancement de nouveaux produits et lors de la création de nouvelles entreprises, et même dans les grandes entreprises bien gérées, la perte des clients reste encore un problème crucial ce qui les obligent à acquérir autant de nouveaux clients que possible pour remplacer ceux qui ont été perdus.

Dans le contexte du CRM, il convient d'acquérir des clients qui ont de fortes chances d'être rentables dans le temps. Par conséquent, l'entreprise doit être en mesure de faire la différence

entre les chasseurs de bonnes affaires, les clients fidèles et les clients compétitifs, car la réussite du CRM commence par l'acquisition des bons clients [29].

La répartition des efforts d'acquisition des clients prend plusieurs formes. L'accent est mis sur la publicité, la promotion des ventes, le bouche-à-oreille et les médias sociaux. Généralement, au moment du premier contact avec le client, des outils de base de données sont utilisés pour capturer les données clés de client.

**Phase de fidélisation :** Le but réel de la collecte de données sur les clients acquis est d'améliorer le taux de rétention. En analysant ces données, les entreprises peuvent identifier les caractéristiques des prospects et des clients qui offrent le meilleur potentiel de gains à vie, ce qui permet de se concentrer davantage sur la fidélisation des clients principaux. [30] a déclaré que, « La clé de la fidélisation de la clientèle est la satisfaction du client ». En effet, la satisfaction des clients n'est pas suffisante pour les fidéliser, selon [31], une autre composante de qualité relationnelle qui comprend la qualité globale, l'engagement et la confiance est nécessaire.

Les clients ont des attentes concernant, par exemple, la qualité du produit, la gamme de produits, la réactivité du service, la stabilité des prix et les activités promotionnelles. La gestion de ces attentes pour améliorer la satisfaction des clients est l'une des meilleures stratégies de fidélisation de la clientèle que l'entreprise doit développer. Cette stratégie de fidélisation vise à conserver une forte proportion de clients importants en réduisant leur défection (churn), car plusieurs études [32,33,34] ont révélé que les clients ciblés par un programme de fidélisation démontrent une plus grande loyauté envers ses entreprises. En outre, dans le contexte de CRM, il convient de cibler que les clients dont la contribution future semble prometteuse.

**Phase d'extension :** La phase d'extension comprend des activités destinées à la croissance de la valeur des clients retenus. A cette phase, l'entreprise bénéficie de la disponibilité de grande quantité d'informations sur les clients et sur leurs comportements d'achat, et s'elle arrive à analyser ces données méticuleusement ainsi d'effectuer des ventes croisées et des ventes incitatives, les clients retenus deviendront plus rentables qu'auparavant.

Les techniques de la fouille de données (data mining) peuvent être utilisées comme des technologies CRM qui sont utiles pour l'extension des clients. A l'aide des données stockées dans les bases de données, et qui comprennent souvent des données sur les transactions effectuées par les clients, des techniques de data mining comme les techniques d'extraction de

règles d'associations peuvent être utilisées comme une méthodologie de recommandations de produits pour accroître les ventes croisées et donc augmenter les revenus, ainsi l'utilisation de ces techniques peut indiquer aux entreprises la probabilité qu'un client achètera un autre produit (propension à acheter), ceci est très utile lorsque l'entreprise souhaite lancer des campagnes de vente croisée ciblées.

En bref, le CRM peut être considéré comme une approche systématique de la gestion du cycle de vie du client, qui aligne les processus et la technologie.

## **1.4. Gestion de la relation client dans le commerce électronique**

### **1.4.1. L'e-CRM: une extension du CRM**

Avec la croissance d'internet et l'explosion des sites e-commerce dans le monde, le CRM a pris une nouvelle direction, donnant naissance au concept d'eCRM (Electronic Customer Relationship Management). En effet, cette nouvelle forme de gestion de la relation client englobe toutes les fonctions du CRM avec l'utilisation de technologies basées sur Internet telles que les courriels, les sites Web, les forums de discussion et d'autres canaux pour atteindre les objectifs du CRM. En d'autres termes, le eCRM se repose sur les concepts développés à la fois par le marketing relationnel et les canaux électroniques afin d'automatiser les processus de marketing, de vente et de service à la clientèle [35], ce qui nous amène à considérer que le concept eCRM est une implication de CRM.

Dans le contexte de e-commerce, une stratégie eCRM efficace consiste à collecter des informations sur les clients, l'historique des transactions et les informations sur les produits, ainsi que des informations sur le flux de clics. Cette quantité d'information provenant des sites e-commerce devra ensuite être analysée par un processus d'exploration de données web. Ce dernier suit les processus d'exploration de données traditionnels, exécutant des fonctions telles que la segmentation, la classification, l'association et la prédiction. Par conséquent, les sites web de commerce électronique seront plus capables de capturer de nouveaux clients, de suivre leurs préférences et leurs comportements en ligne, et de personnaliser les communications, les offres, les produits, les services et les prix. Cependant, le concept de personnalisation de masses, ou l'approche One to One (Une offre personnalisée pour chaque client), promue par des auteurs comme Peppers et Rogers (1993), est devenu le pivot principal de l'eCRM [36].

### **1.4.2. Les stratégies CRM dans le contexte e-commerce.**

Depuis l'adoption du marketing relationnel, beaucoup de recherches commencent à étudier l'impact de CRM sur l'entreprise, par exemple l'auteur Frederick Reichheld dans son livre "*The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*" [37] a nous démontré grâce aux quelques statistiques le rôle crucial de la satisfaction de la clientèle pour l'entreprise, parmi ces chiffres nous trouvons :

- Acquérir un nouveau client coûte quatre à cinq fois plus cher que de retenir un client déjà actif.
- Un client entièrement satisfait a six fois plus de chance de racheter un produit sur une période de 1 à 2 ans qu'un client insatisfait.
- Une augmentation de 5 % au niveau du taux de rétention de la clientèle peut impliquer une augmentation de 30 à 85% des profits.
- Une augmentation de 2% de la capacité d'une entreprise à retenir sa clientèle équivaut à une diminution des coûts d'opérations de 10%.

De ce fait, les entreprises commencent à investir de plus en plus de ressources et d'efforts dans ce domaine. Parallèlement, des chercheurs universitaires commencent également à étudier les différents facteurs qui améliorent la fidélisation des clients [38,39]. En outre, le secteur de commerce électronique essaie également d'attraper les mêmes tendances. En raison des caractères uniques de ce secteur, les entreprises de vente en ligne ne peuvent pas simplement adopter les théories classiques du marketing relationnel. Ces entreprises ont besoin de quelques théories dans le contexte du commerce électronique pour soutenir leur fonctionnement [40]. Dans ce contexte, un cadre analytique examine comment les sites de commerce électronique mettent en œuvre des stratégies de CRM en ligne a été élaboré par [41] sur la base du modèle de de Winer (2001) [36]. Ce cadre considère les processus eCRM sur trois phases d'interactions clients/site e-commerce : l'acquisition, la vente/service et la rétention après-vente. Les interactions dans chaque phase sont également examinées selon quatre facteurs qui sont essentiels aux solutions CRM basées sur internet : (1) la satisfaction des clients, (2) la fidélité des clients, (3) la personnalisation et (4) la confiance.

Comme le soulignent les auteurs dans leurs études : le modèle de relation client de Winer [36], le modèle d'indice de satisfaction client américain de [42] et la recherche de [43] qui porte sur la confiance dans le marketing relationnel, ces cinq facteurs représentent les caractéristiques principales du marketing relationnel.

#### **1.4.2.1. La satisfaction du client**

Le but ultime de toute entreprise est de s'assurer que le client qui achète son produit ou service est satisfait [34]. La satisfaction du client a suscité un réel intérêt pour les organisations et les chercheurs. Elle est définie comme « les sentiments de plaisir ou de déception d'une personne qui résultent de la comparaison de la performance perçue d'un produit ou de ses résultats avec ses attentes » [44]. [45] suggèrent que la satisfaction du client se produit lorsque l'expérience d'achat et le service après-vente répondent aux attentes du client. Ainsi [46] ont affirmé que la satisfaction du client est perçue comme le degré auquel les attentes des clients en termes de qualité des produits et des services sont atteintes ou dépassées par rapport à la performance perçue.

En général, dans toute entreprise, la mise en œuvre d'une stratégie de satisfaction du client implique les activités suivantes : identification des clients, identification de leurs besoins et attentes, et enfin, mesure de leurs perceptions [47]. En étendant cette conclusion au contexte du commerce électronique, nous pouvons suggérer l'utilisation de comportement de navigation Web des utilisateurs pour l'identification de clients et leurs attentes. Identifier les utilisateurs au cours de leurs activités commerciales en ligne est extrêmement important dans le commerce électronique, car ceci fournit aux entreprises de vente en ligne un large éventail d'applications qui peuvent affecter positivement la satisfaction des clients, telles que la personnalisation du contenu Web et les recommandations de produits/services basées sur le profil de client [48].

En outre, les acteurs de commerce électronique peuvent améliorer la satisfaction des clients en intégrant les caractéristiques suivantes dans leurs sites e-commerce :

- Large gamme de produits
- Un bon design et un processus d'achat en ligne simple et sécurisé
- Une interface de paiement électronique sécurisée et facile à utiliser

- Large choix de moyen de paiement (Cash à la livraison, Cartes bancaires, Virement, porte-monnaie électronique, etc.)
- Différentes options de livraison
- Des conditions générales de ventes clairement affichées.
- Une politique de retour bien expliquée pour chaque situation
- Plusieurs supports de communication pour simplifier la prise de contacts (téléphone, email, media sociaux, chat, FAQ interactive...)
- Outil de recommandation de produits et de services
- Service après-vente comme suivi de colis et la réponse aux demandes des clients
- Module commentaires, notes et avis après commande.

L'insertion des descriptions détaillées de produits sur un site de commerce électronique augmente la possibilité pour le client d'obtenir des produits de meilleure qualité. Ainsi, un site de commerce électronique facile à naviguer, rapide et bien contenu peut rendre l'expérience d'achat en ligne plus satisfaisante, et de plus, le client gagne également du temps et des efforts pour trouver les produits ou services attendus.

#### **1.4.2.2. La fidélisation du client**

La fidélisation de la clientèle peut être définie comme « un engagement profond exprimé par le consommateur d'acheter à nouveau le produit ou service qu'il préfère, de manière cohérente dans le futur, en dépit des influences situationnelles et des effets marketing susceptibles de provoquer un changement de comportement d'achat » [49]. Pour une entreprise, la fidélisation de la clientèle doit être améliorée parce que les clients fidèles augmenteront les revenus et créeront une efficacité pour le fonctionnement de l'entreprise [50]. Et plus encore, les clients fidèles donnent une appréciation positive envers la marque et partagent la valeur positive de l'entreprise avec les autres [51]. Selon certaines recherches fondatrices du marketing de services [37,46,52]), conserver un client fidèle coûte cinq à dix fois moins cher que d'en acquérir un nouveau. Par conséquent, la fidélisation devient alors un facteur important sur lequel une entreprise devrait se concentrer pour atteindre un avantage économique optimal.

En outre, la satisfaction du client semble avoir un effet positif sur la fidélité des clients [53,54,55]. De nombreux chercheurs ont également montré que la qualité du service et la satisfaction du client sont les principaux antécédents de la fidélisation [56,57,58]. Ces constatations vont dans le même sens que celles d'autres recherches récentes qui ont suggèrent une relation directe et positive entre la satisfaction du client et la fidélité du client [59,60], de sorte que lorsque la satisfaction atteint un certain niveau, la fidélité augmente considérablement ; en même temps, lorsque la satisfaction tombe à un certain point, la fidélité diminue également de façon spectaculaire. Des études antérieures [61,62] ont révélé que plus le niveau de satisfaction de la clientèle envers l'entreprise est élevé, plus le niveau de loyauté est élevé.

Les programmes de fidélisation sont parmi les différentes stratégies de marketing relationnel utilisées par les entreprises. La création ou le choix de ces programmes de fidélisation est défini selon l'activité de l'entreprise et les objectifs à atteindre. En général, un programme de fidélisation ciblé peut être basé sur la communication ou sur le concept de compensation. Dans le premier cas, il aura comme objectifs globaux d'informer, d'apporter des conseils et d'assurer une forme de suivi du client et il peut prendre plusieurs formes, à titre d'exemple : des magazines et des newsletters envoyés aux clients, à intervalles réguliers, ou bien encore sous forme d'un accès privé sous un blog ou un site internet dans lequel le client peut obtenir des conseils ou des informations spécifiques à son besoin. Dans le deuxième cas, le principe est de remercier les clients fidèles en se basant sur un système de compensation qui permet aux clients d'accumuler plus facilement des bons de réduction, des offres promotionnelles ou des points de fidélité qui peuvent ensuite les utiliser contre des remises ou les convertir en cadeaux.

Du point de vue du marketing en ligne, [63] ont étudié la fidélité des clients dans le contexte du commerce électronique et définissent la fidélité électronique comme l'attitude et l'engagement favorables du client envers un site de commerce électronique qui se traduisent par un comportement d'achat récurrent.

Les acteurs du commerce électronique peuvent fidéliser leurs clients en fournissant les fonctionnalités CRM suivantes :

- Des informations détaillées et claires sur le programme de fidélisation
- Des incitations ciblées pour adhérer et participer au programme de fidélité

- Un programme de récompense en fonction de centres d'intérêt et préférences d'achat des clients
- L'envoi automatique des newsletters contenant les dernières actualités et les promotions dont le client fidèle peut bénéficier grâce à ses points de fidélité
- Des services personnalisés pour les clients fidèles
- Un compte fidélité privé sur le site e-commerce pour chaque client afin de donner un état sur le statut de fidélisation, les options de mise à niveau, les procédures de remboursement, les remises et les offres promotionnelles
- Des services spéciaux pour les acheteurs fréquents : livraison rapide et sans frais
- Recueillir les avis clients et les afficher clairement sur le site e-commerce
- Des outils d'assistance en ligne (click to chat, click to call) pour répondre d'une façon immédiate aux réclamations des clients

#### **1.4.2.3. La personnalisation**

La personnalisation dans la gestion de la relation client fait référence à la mesure dans laquelle une technologie, un bien ou un service peut être créé, sélectionné ou modifié pour se conformer aux préférences de consommateur [64], cette stratégie commerciale visant à fournir au consommateur de façon rentable tout ce qu'il désire, en tout temps, en tous lieux et de toutes les façons [65]. Le Jupiter Research Group définit la personnalisation comme « l'utilisation d'analyses prédictives sur les données des clients pour conduire à une livraison ciblée d'informations ou de messages promotionnels » [66]. Cette définition introduit la stratégie de collecte et traitement des données relatives aux clients (données liées aux achats ou services, pages consultées, produits recherchés, temps passé en ligne, etc.). En adaptant les produits et les services aux besoins et aux préférences des clients individuels, une entreprise peut satisfaire et dépasser les attentes des clients et augmenter leur perception de la qualité des produits, de plus, cette tactique de personnalisation peut contribuer l'entreprise à maintenir une relation à long terme avec ses clients [67]. Cette démarche suppose que l'entreprise est capable de prévoir les demandes et les attentes des clients, d'optimiser l'offre par rapport à ces attentes prévues et de la délivrer [68]. En outre, les consommateurs tirent davantage parti de la personnalisation, car elle



leurs permet de bénéficier de la meilleure adéquation entre les services/produits et leurs besoins [69].

Dans les environnements de commerce électronique, la personnalisation est définie comme la capacité d'un site e-commerce à permettre aux clients de trouver et d'accéder aux produits/services adaptés le plus précisément possible à leurs besoins individuels [70], ce qui affecte positivement la e-satisfaction et la e-fidélisation des clients en économisant leurs temps et leurs efforts pour trouver ce qu'ils veulent sur le site e-commerce. Ceci a été également confirmé par [71] qui ont identifié la personnalisation parmi les principaux facteurs qui génèrent la fidélité électronique. Des travaux antérieurs [72], ont souligné également l'importance de la personnalisation dans le eCRM en abordant le rôle des systèmes de recommandation. Selon eux, les systèmes de recommandation représentent un moyen essentiel d'automatiser la personnalisation de masse pour les sites de commerce électronique et permettent de diminuer le taux de rebond et d'augmenter le taux de transformation. [73] postulent également qu'internet offre aux entreprises des opportunités inédites pour connaître en profondeur leurs clients en analysant la navigation et les éventuels achats passés de chaque client pour comprendre ses préférences et ses attentes, et lui recommander de manière très fine les produits qui lui correspondent (des recommandations totalement 1-to-1).

Un marchand en ligne peut réussir la personnalisation de son site e-commerce (contenus, produits et services) en intégrant les fonctionnalités suivantes :

- Un affichage de page personnel pour voir le statut de commande actuel, l'historique des achats du client, l'historique des livraisons et les paniers abandonnés.
- Proposition des produits complémentaires qui s'accordent bien avec les produits consultés, recherchés ou les produits mis au panier.
- Extraire de la masse des résultats, les produits les plus adaptés à chaque internaute en mettant en œuvre des recommandations personnalisées.
- Favoriser la recommandation de produits similaires aux produits complémentaires lors de la consultation du produit.
- Les résultats de moteurs de recommandation doivent donc se trouver de préférence au-dessus de la ligne de flottaison

- Relancer les paniers abandonnés

Chaque étape d'un eCRM efficace, du premier contact avec le site e-commerce, à l'achat jusqu'au service après-vente en passant par l'utilisation du service ou du produit, doit être adaptée aux attentes et besoins du client. C'est à cet objectif que souhaite répondre tous système de personnalisation de la relation client.

#### **1.4.2.4. La confiance**

De nombreuses études récentes ont été réalisées pour identifier l'importance de la confiance dans le développement des relations entre un acheteur et un vendeur [74,75,76,77]. La confiance en tant que concept est souvent liée à la volonté du client de s'appuyer à l'avenir sur les actions et le comportement d'une entreprise [78]. Des travaux antérieures [79,80,81], ont également décrit que la confiance est un aspect clé dans la construction de relations, ainsi [82] affirme que la confiance devrait être l'un des principaux éléments des activités de marketing relationnel pour établir des relations fructueuses avec les clients.

Dans les environnements en ligne, la confiance repose sur le niveau de risque, qui peut être déterminé par l'utilisation des outils et des services impliquant des échanges financière (les applications Web et mobiles, l'infrastructure réseau), les problèmes de confidentialité des clients, transferts sécurisés de ces données et l'authentification du système [83]. D'un côté, la qualité des services offerte par l'entreprise qui sont reflétés dans la fourniture des informations cohérentes et complètes, la personnalisation des services/produits, la simplicité et la facilité d'utiliser les interfaces système, la fiabilité de la connectivité, la disponibilité du support client et la sécurité de données confidentielle affectent directement et significativement la confiance des clients envers un vendeur en ligne. D'un autre côté, des recherches antérieures ont révélé que les risques perçu associés à l'envoi et la récupération d'informations confidentielles (Numéro de carte bancaire, date de validité et code cryptogramme) par voie électronique introduisent certaines vulnérabilités qui rendent les échanges électroniques plus risqués [84]. De plus, l'absence d'interaction face à face entre l'acheteur et le vendeur en ligne augmente la vulnérabilité de l'acheteur. Ces risques perçus affectent toutes les décisions d'achat et le comportement des clients en les dissuadant d'effectuer des transactions en ligne. Ainsi, de nombreux chercheurs ont étudié les liens de cause à effet entre la confiance en ligne, la satisfaction et la fidélité

électronique, et elles ont montré que la confiance affecte la satisfaction et la fidélité du client, et influence directement l'efficacité de la stratégie eCRM [85,86,87].

Selon [88], la stratégie clé des spécialistes du marketing en ligne permettant d'accroître la confiance des consommateurs en ligne consiste à afficher des sceaux de confidentialité et de sécurité sur leurs sites Web. En général, la sécurité consiste en quatre éléments : l'authentification, l'autorisation, le cryptage et l'audit. Du côté entreprise, les sites e-commerce devraient offrir des mécanismes garantissant la vie privée du client et le transfert sécurisé des fonds électroniques grâce à la coopération avec des systèmes de paiement tiers et la mise en place de sceaux Internet [89]. Du côté client, l'internaute devrait rechercher dans les sites de e-commerce les composants qui assurent et vérifient la crédibilité des normes de paiement telles que : SET (Secure Electronic Transaction) et SSL (Secure Socket Layer) [90]. Par conséquent, l'acquisition, le renforcement de la confiance et la réduction des risques perçus pour le commerce électronique peuvent être basées sur l'intégration des fonctionnalités suivantes :

- Des informations cohérentes sur les produits et services
- Politique de gestion des retours clairement définie.
- Donner aux clients l'opportunité d'évaluer la crédibilité de l'entreprise (Montrer ce que d'autres clients disent sur l'entreprise sous forme de commentaires, avis, notes, etc.)
- Des informations sur les prix et les options de paiement
- L'affichage de la politique en matière de confidentialité, de remboursement, d'expédition, etc.
- L'affichage des logos de méthodes de paiement bien connus qui garantissent la sécurité des transactions en ligne (par exemple : Verified by Visa ou MasterCard SecureCode, PayPal, iDEAL)
- Mécanisme d'authentification approprié
- L'intégration d'une plateforme de paiement certifiée PCI DSS (Payment Card Industry Data Security Standard), qui garantit le traitement, le stockage et la transmission des données confidentielles.
- Des règles anti-fraudes performantes aux banques et aux commerçants

Les fonctionnalités de renforcement de confiance sont importantes pour les trois phases de CRM. Des mesures de confidentialité et de sécurité doivent être clairement mises en évidence pour les nouveaux clients, ainsi des transactions en ligne sécurisées et une meilleure expérience d'achat favorisent l'efficacité, la confiance et l'image du site e-commerce, ce qui particulièrement important pour encourager les clients à revenir et faire des achats répétitifs.

## **Conclusion**

Dans cette partie, nous nous sommes attachés à décrire le concept CRM dans son sens général, ainsi nous avons essayé de montrer tout au long de ce chapitre le eCRM comme une extension du CRM qui permet de créer un lien étroit avec les clients dans le secteur de e-commerce. La nature et les caractères uniques de ce secteur conduisent plusieurs chercheurs à adopter les théories classiques du marketing relationnel dans le contexte de commerce électronique. En résumé, comme dans les contextes traditionnels, la conception des stratégies CRM efficaces nécessite une connaissance approfondie de la clientèle présente et prospectée (Padmanabhan et Tuzhilin 2003). Autrement dit, les données des clients jouent un rôle essentiel pour atteindre les objectifs principaux d'un système CRM qui sont l'acquisition, la satisfaction, la fidélisation et l'augmentation de taux de rétention. C'est particulièrement le cas dans le monde en ligne où les serveurs web collectent automatiquement de grandes quantités de données transactionnelles et comportementales sur les clients. Toutefois, l'incapacité à découvrir de précieuses informations cachées dans ces données rend leur stockage au niveau des entrepôts de données inutile, c'est la raison pour laquelle la mise en œuvre d'un eCRM analytique basé sur le processus de découverte de connaissance dont la matière première est la donnée, et les principaux composants sont les entrepôts de données (data warehouse), la fouille de données (data mining) et les outils de reporting, et qui permet d'extraire, de stocker, d'analyser et d'exploiter les données brutes de l'entreprise dans le but d'améliorer la connaissance et la compréhension du client est devenue une tâche essentielle dans tous projets CRM. Dans le chapitre suivant relatif aux techniques d'extraction des connaissances, nous verrons les étapes fondamentales du processus de découverte de connaissances et les techniques de data mining applicables aux stratégies de relation clients en ligne.

# Chapitre 3 : La découverte de connaissances et les techniques de data mining pour le CRM

## Introduction

Avec le développement et à la diffusion de nouvelles technologies de l'information et de la communication (TIC) et surtout internet, les sources de données sont devenues très nombreuses et très riches. Par exemple, dans les environnements en ligne, les sites de commerce électronique collectent et stockent des données sur les clients, les achats (transactions), les habitudes de navigation et les durées d'utilisation, ce qui conduit à générer des bases de données volumineuses, complexes et pas nécessairement structurées. Sans ces données il est impossible d'arriver à établir une véritable connaissance client. Mais le problème réside dans le fait que les pratiques commerciales traditionnelles ne parviennent souvent pas à découvrir des informations utiles cachées dans les grandes bases de données, ce qui empêche les entreprises de bénéficier de ces données en les transformant en valeur ajoutée, c'est-à-dire en connaissances actionnables [91]. Il est donc crucial de faire appel à d'autres techniques puissantes qui sont capables de donner un sens à ces données. Cependant, les modèles de processus d'Extraction de Connaissances à partir de Données (ECD) dont la matière première est la donnée, et les principales composantes sont le data warehouse (entrepôt de données) et les techniques de data mining, sont conçus pour remédier à cette lacune. Le présent chapitre a pour objectif de présenter, dans un premier temps, le processus d'ECD, et le standard CRISP-DM « Cross-Industry Standard Process for Data mining » comme un modèle de référence, puis, dans un second temps, d'identifier les techniques de data mining existantes applicables aux stratégies de relation clients, et en particulier celles de rétention qui sont fondées sur les modèles de prédiction d'attrition. L'idée est d'examiner les techniques et les approches les plus couramment utilisées, et d'évaluer leur potentiel dans le contexte de la thèse. Cette étude bibliographique permet de positionner nos problématiques dans un cadre académique.

## 1. Extraction de connaissances à partir de données

### 1.1. Enjeux

L'Extraction de Connaissances à partir de Données (ECD) ou en anglais, Knowledge Discovery in Databases (KDD) est un processus dont l'objet est d'identifier et d'extraire une connaissance ou un savoir, à partir de grande quantité de données, par des méthodes automatiques ou semi-automatiques, et l'utilisation industrielle ou opérationnelle de cette connaissance.

L'expression ECD a été inventée lors du premier atelier de « Knowledge Discovery in Databases » en 1989 pour souligner que la « connaissance » est le produit final d'une découverte axée sur les données. Il existe un large éventail de définitions de l'ECD dans la littérature, et toutes ces définitions s'accordent sur le fait que l'intention est de récolter des informations en reconnaissant des modèles dans des données brutes. Examinons la définition proposée par [92], « L'extraction de connaissances à partir des données (ECD) est l'acquisition de connaissances nouvelles, intelligibles et potentiellement utiles à partir de faits cachés au sein de grandes quantités de données ». Le but est de chercher surtout à isoler des traits structuraux (patterns) qui soient valides, non triviaux, nouveaux, utilisables et si possible compréhensibles ou explicables. L'expression ECD a été popularisée dans l'intelligence artificielle et l'apprentissage automatique. Elle a évolué et continue d'évoluer à partir de l'intersection de domaines de recherche tels que l'apprentissage automatique, la reconnaissance de formes, l'intelligence artificielle, les bases de données, les statistiques, l'acquisition de connaissances pour les systèmes experts, la visualisation de données et le calcul haute performance [92], ce qui mène à la création d'une grande variété d'applications dans le monde réel comme par exemple, dans la santé, la météorologie, le e-commerce, le marketing, la finance, la télécommunication, la fabrication, les jeux et bien d'autres. En général, les applications d'ECD ont le plus souvent l'un des objectifs suivants :

- Gestion de la relation client : marketing direct, acquisition des clients, détection de la fraude, etc.

- L'aide à la décision s'applique à presque tous les domaines, allant de la médecine au marketing en passant par la logistique. Les applications ECD visent à justifier les décisions en fonction des données en mettant en relation les actions et les résultats.

- Conception des systèmes de recommandations : Ces derniers suggèrent généralement des objets aux clients selon leur profil, tout en fournissant des informations qui les aident à décider quels objets acheter. Les objets peuvent être, par exemple, des produits comme dans le cas des sites de e-commerce, ou des documents comme dans l'apprentissage des moteurs de recherche.

- Développement des outils et des services de prédiction : la prédiction de faillite des entreprises, la gestion du risque de crédit, prévention des risques naturels, prédiction de l'attrition, prédiction de la consommation d'énergie, etc.

La découverte de connaissances devient alors un enjeu stratégique fondamental dans différents domaines [93].

## **1.2.Processus de découverte de connaissances**

Les termes "ECD" et "data mining" sont souvent utilisés comme synonymes. Comme il est mentionné ci-dessus, l'ECD désigne le processus global de transformation des données de bas niveau en connaissances de haut niveau. D'un autre côté, data mining fait référence à l'application d'algorithmes afin d'extraire des modèles ou des règles de comportement à partir des données et de corriger l'interprétation des résultats. Par conséquent, data mining n'est qu'une étape dans le processus global de ECD.

La Figure 3.1 montre que le processus de découverte de connaissance s'effectue sur plusieurs étapes interrompues continuellement par des prises de décision. La description de chaque étape est donnée dans la partie suivante :

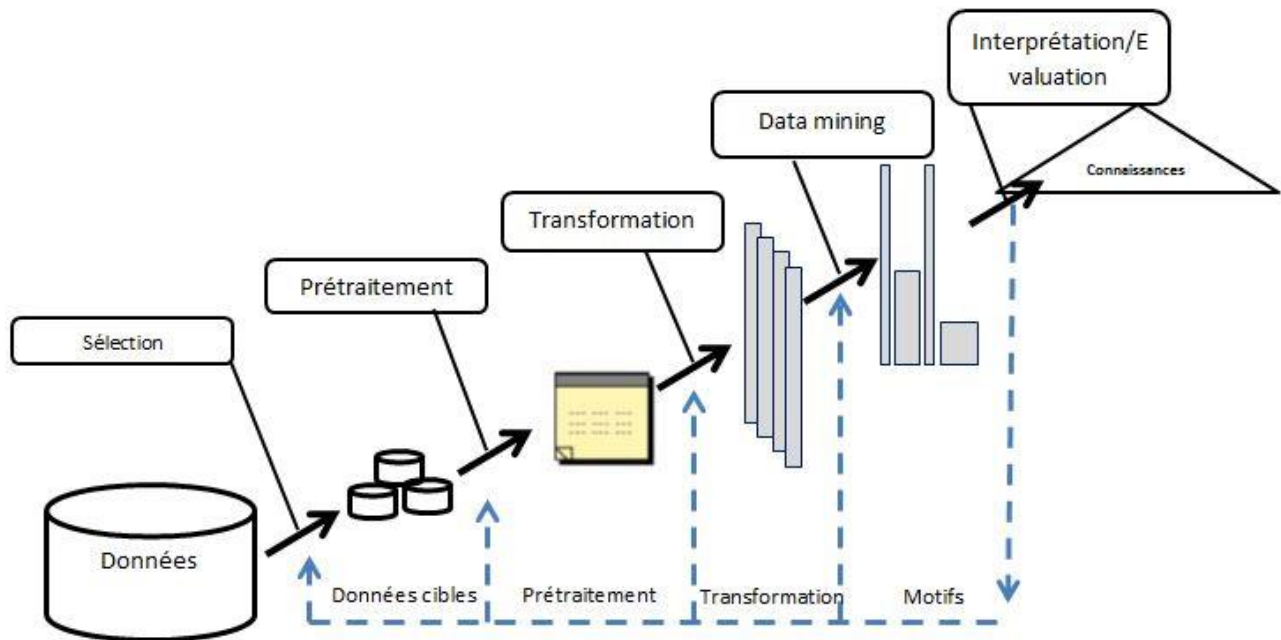


Figure 3.1 - Processus d'Extraction de Connaissances à partir des données [92].

### 1.2.1. Etapes du processus d'ECD

Les étapes fondamentales qui ont été proposées initialement pour le processus d'ECD (Fayyad et al., 1996) [92], et qui ont comme objectif principal d'atteindre l'étape de découverte de connaissances sont résumées comme suit :

- 1- **Sélection** : Le processus de sélection développe une compréhension claire du domaine d'application. L'identification de l'objectif du processus ECD du point de vue de l'utilisateur est également une tâche importante. Sur la base de cette compréhension et de ce plan global, un sous-ensemble de variables et de données ou échantillonnage de données sont sélectionnés pour les utiliser dans le processus de découverte de connaissances.
- 2- **Nettoyage et prétraitement des données** : Au cours de cette étape, les données sélectionnées dans l'étape précédente sont nettoyées et traitées. Ces opérations sont fondamentales pour réduire la présence de bruit dans les données ou des valeurs aberrantes. Cette étape comprend également la conception d'un moyen de traitement de champs de données manquants.
- 3- **Transformation** : Elle inclut la réduction et la projection de données, en définissant des modes de représentation des données en référence aux objectifs à atteindre. L'utilisation de méthodes de réduction ou de transformation a pour objectif de réduire le nombre



effectif de variables prises en considération ou les représentations invariantes des données.

- 4- Data mining :** Cette étape comprend la sélection de techniques d'exploration de données à utiliser pour rechercher des modèles dans les données selon l'objectif du processus ECD, c'est-à-dire s'il s'agit d'une régression, d'une classification, d'un clustering ou d'une prédiction. Au cours de cette étape, plusieurs modèles sont généralement élaborés, le plus performant étant choisi.
- 5- Interprétation et évaluation :** La dernière étape implique l'interprétation des modèles identifiés dans l'étape précédente. Des rétroactions aux points précédents sont possibles pour d'autres itérations jusqu'à l'obtention de résultats complètement satisfaisants pour les experts métier. Cette étape peut également impliquer des techniques de visualisation aidant les experts à interpréter et à valider le modèle choisi.

Le processus de découverte de connaissances peut nécessiter un nombre important d'itérations, comme le montrent les lignes pointillées de la Figure 3.1. Il pourrait même créer des boucles entre deux étapes ou plus.

### **1.2.2. Le standard CRISP-DM comme modèle de processus d'ECD**

Tous les modèles de processus d'ECD consistent en un ensemble d'étapes de traitement et utilisent le même flux de processus lors de la conception d'un système d'extraction de connaissances [94]. Le modèle comprend la description de toutes ses étapes et les tâches comprises dans chacune de ces dernières, et aussi une explication des relations entre ces tâches. Un modèle de processus complet est un modèle qui rend les gros projets d'exploration de données plus rapides, plus efficaces et moins coûteux.

Plusieurs modèles de processus ont été développés, les plus cités dans la littérature du domaine de l'ECD sont les suivants :

- Knowledge Discovery in Databases (KDD) Process by Fayyad et al. (1996).
- Information Flow in a Data Mining Life Cycle by Ganesh et al. (1996).
- SEMMA by SAS Institute (1997)
- Knowledge Discovery Life Cycle (KDLC) Model by Lee and Kerschberg (1998).
- CRoss-Industry-Standard Process for Data Mining (CRISP-DM) by CRISP-DM (2000).
- Generic Data Mining Life Cycle by (DMLC) by Hofmann (2003).
- Ontology Driven Knowledge Discovery Process (ODKD) by Gottgroy (2007).

La représentation du processus d'ECD sous forme d'un ensemble d'étapes qui sont exécutées en séquence continue est le point commun entre ces différents modèles (voir figure 3.2) [95]. Ainsi la plupart des modèles, utilisent le même flux de processus, y compris la plupart des étapes suivantes : compréhension de la problématique, compréhension des données, préparation des données, modélisation, évaluation des modèles et en fin le déploiement des résultats. Tandis que, la portée, le nombre et les tâches spécifiques à chaque modèle représentent les principales différences entre ces modèles.

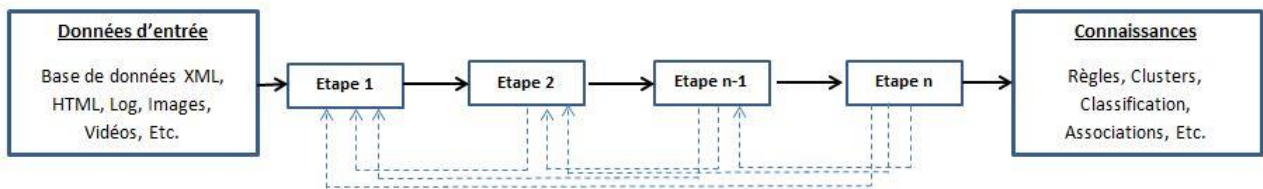


Figure 3.2 - Structure séquentielle et feedback dans un modèle de processus d'ECD.

Nous allons à présent présenter le modèle CRISP-DM (Shearer, 2000) [96], sur lequel notre approche sera développée, ce choix est motivé par le fait que ce dernier est devenu actuellement le modèle de référence pour le processus d'ECD.

CRISP-DM (CRoss-Industry Standard Process for Data Mining) est un modèle à six étapes qui a été au départ initiée dans la fin de l'année 1996 par un large consortium d'entreprises européennes : SPSS (fournisseur de solutions d'exploration de données commerciales), NCR (fournisseur d'entrepôt de données Teradata), Daimler Chrysler (fabricant d'automobiles), et OHRA (compagnie d'assurance néerlandaise) ; Les deux dernières sociétés ont servi comme sources de données et des études de cas [96]. Il reste aujourd'hui le seul modèle utilisable efficacement pour tous les projets *Data Science*, et s'applique à tous les domaines d'activité, indépendamment du logiciel de data mining utilisé.

Comme illustré dans la figure 3.3 le modèle CRISP-DM décompose le cycle de vie d'un projet de data mining en six étapes allant de la compréhension du problème métier au déploiement et la mise en production, et chacune de ses étapes est composée d'un ensemble des tâches spécifiques.

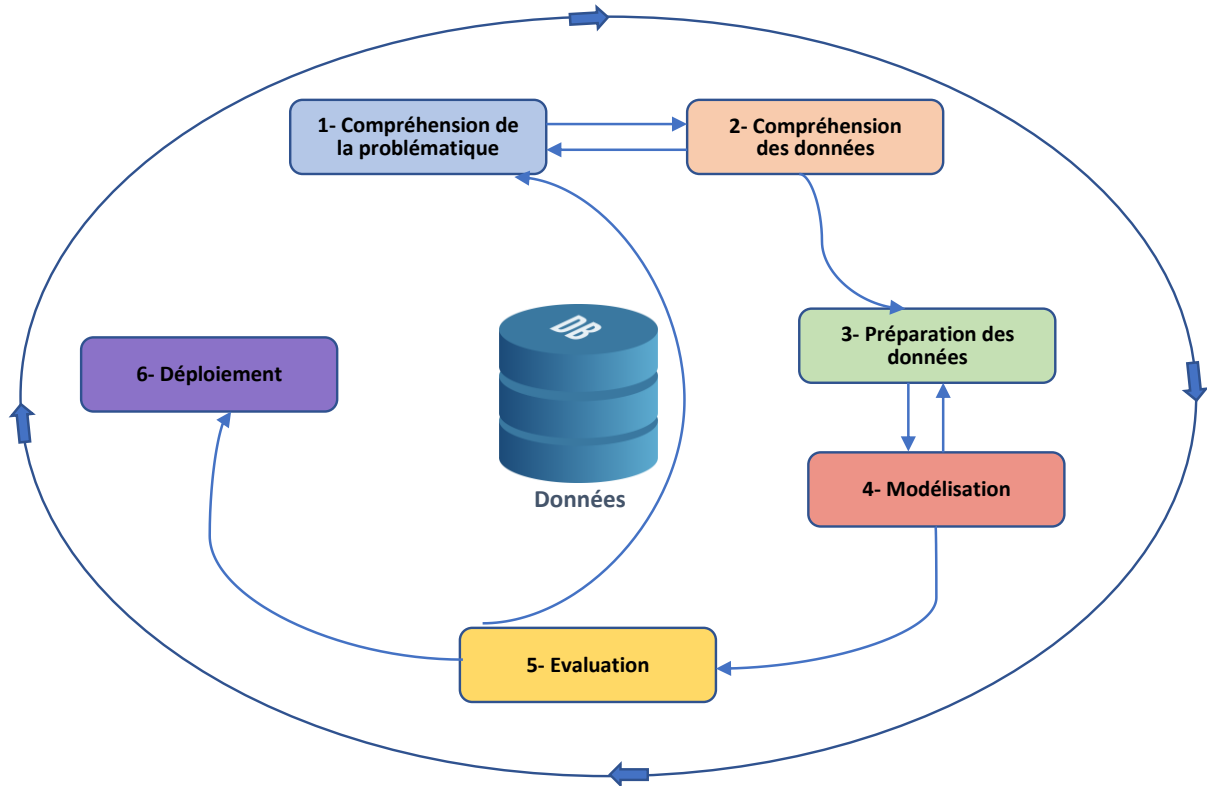


Figure 3.3 – Cycle de vie d'un projet data mining avec le modèle CRISP-DM.

### Les étapes du CRISP-DM :

**Compréhension de la problématique** : La première étape du CRISP est très importante, dans cette étape les objectifs et les exigences d'un point de vue commercial doivent être clairs et fixés précisément, ensuite les objectifs commerciaux sont traduits en objectifs de data mining, et un plan de projet préliminaire sera créé pour accomplir ces objectifs.

**Compréhension des données** : Cette étape commence par le recensement des données existantes, la description des données, l'exploration des données et l'évaluation de la qualité de ces données.

**Préparation des données** : Cette étape est la plus longue d'un projet d'ECD, elle monopolise généralement 80 % du temps consacré à l'ensemble du projet. Plusieurs outils d'extraction, de transformation, et de téléchargement (ETL) peuvent être utilisés dans cette étapes afin de concevoir et d'alimenter l'entrepôt de données. Généralement cette phase se fait en quatre étapes, la sélection des données, le nettoyage des données, la transformation et l'agrégation des données, et sont susceptibles d'être exécutées plusieurs fois et sans un ordre précis.

**Modélisation** : L'objectif de cette étape est la sélection et l'application des techniques de modélisation appropriées afin d'en extraire des modèles. La préparation de données est aussi nécessaire dans cette étape pour utiliser correctement certains algorithmes particuliers qui requièrent souvent des paramétrages différents et des données en entrées différentes. Par conséquent, il y aura dans cette étape un va et vient entre la phase de préparation de données et la modélisation. C'est ce qui est représenté par une boucle entre les deux étapes sur la figure 3.3. A la fin de cette étape plusieurs modèles data mining vont être générés qui répondent tous à la problématique et un test préalable de ces modèles est souvent réalisé à ce stade.

**Evaluation** : Les modèles créés à l'étape précédente passent par une étape de validation, elle consiste à évaluer si les résultats de ces modèles sont conformes aux objectifs de l'entreprise. Par conséquent, un ou plusieurs modèles présentant les meilleures performances sont découverts.

**Déploiement** : L'objectif de cette étape est d'organiser et de présenter les connaissances découvertes dans une forme compréhensible par le client. L'étape de déploiement peut être divisée en plusieurs activités : l'élaboration d'un plan de déploiement, et un autre de contrôle et de maintenance, la rédaction d'un rapport final et enfin la revue et l'évaluation de tout le projet.

L'adoption du modèle CRISP-DM pour atteindre les résultats de notre approche, sera détaillée dans le chapitre suivant.

Après ce bref résumé des points fondamentaux du processus de découverte de connaissances, et le modèle CRISP en particulier, nous allons maintenant nous concentrer sur les méthodes de fouille de données.

### **1.3.La fouille de données**

La fouille de données (data mining en anglais) est une étape centrale du processus d'ECD, qui implique l'utilisation d'outils sophistiqués d'analyse de données pour découvrir des modèles et des relations valides auparavant inconnus dans de grands ensembles de données. Ces outils peuvent inclure des modèles statistiques, des algorithmes mathématiques et des méthodes d'apprentissage automatique. Selon le résultat attendu du processus global d'ECD, les tâches de data mining peuvent être classées en deux catégories : prédictive et descriptive [97,98]. Les techniques prédictives utilisent des données historiques pour déduire quelque chose sur les

événements futurs. Les tâches d'exploration prédictive utilisent des données pour créer un modèle permettant de faire des prédictions ; les techniques descriptives visent à trouver des modèles dans les données qui fournissent des informations sur les relations cachées internes. Les tâches d'exploration descriptives caractérisent les propriétés générales des données et les représentent de manière significative. Bien qu'il existe une large gamme de méthodes de data mining, la figure 3.4 illustre la classification de celles qui sont principalement utilisées pour atteindre les objectifs de prédiction et de description.

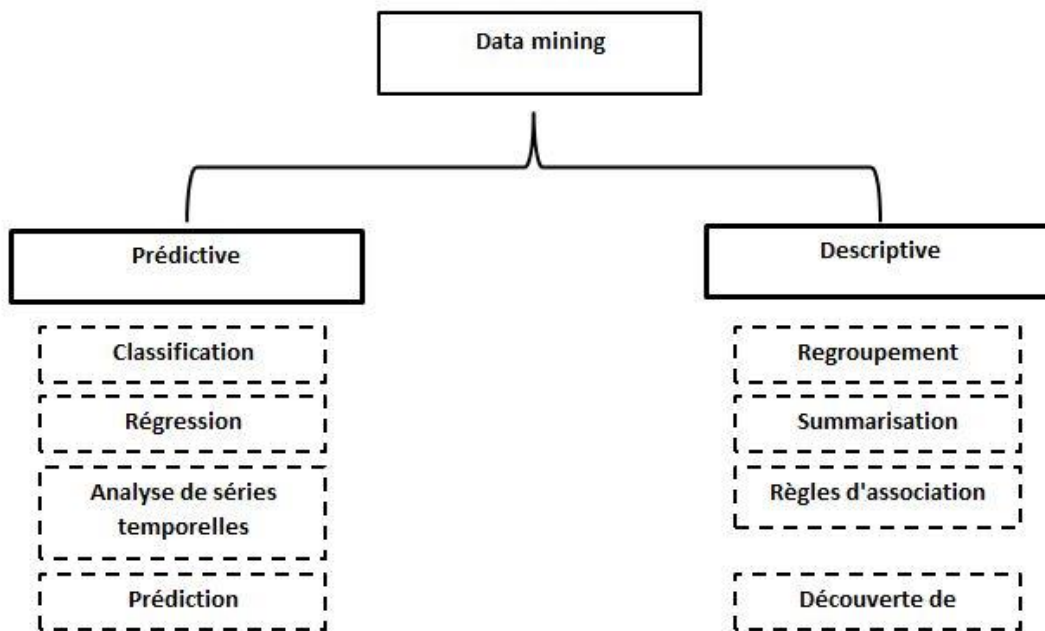


Figure 3.4 - Les différentes méthodes de data mining

Comme il est indiqué précédemment, le choix de la méthode de data mining dans le processus de découverte de connaissance est lié aux problèmes qui font l'objet de l'analyse, ces problèmes peuvent concerner la classification, le clustering, le séquençage, l'association et la prédiction. En d'autres termes, une fois que le rôle des systèmes de data mining est identifié en fonction de chaque domaine d'application, il sera possible de cibler les techniques de data mining convenables.

Il convient également de noter que plusieurs méthodes avec des objectifs différents peuvent être appliquées successivement pour atteindre un résultat souhaité. Par exemple, pour déterminer quels clients sont susceptibles d'acheter un nouveau produit, un responsable de marketing peut avoir besoin d'utiliser d'abord le clustering pour segmenter la base de données client, puis

appliquer une régression pour prédire le comportement d'achat de chaque cluster [99]. Nous décrivons ci-dessous quelques techniques clés de data mining.

- **Association** : L'association est l'une des méthodes de data mining les plus connues. C'est la technique qui consiste à trouver des modèles où un événement est étroitement lié à un autre événement. Les règles d'association extraites de cette méthode sont des modèles descriptifs de la forme  $X \rightarrow Y$ , où X est la partie conditionnelle d'une règle d'association, et Y représente la partie conséquente. L'extraction de règle d'association est une tâche vise à découvrir les règles d'association cachées et intéressantes, entre les éléments de la base de données, ayant le support  $\geq$  minsup (le seuil). Le support d'une règle d'association indique à quelle fréquence cette règle se produit dans les données. Cependant, un support plus élevé correspond à une corrélation plus forte entre les éléments de la même transaction. L'algorithme APRIORI [100] représente le plus connu des algorithmes de règle d'association.

Par exemple, la technique d'association est utilisée dans l'analyse du panier de consommation pour identifier les produits que les clients achètent fréquemment ensemble. Sur la base de ces données, les entreprises peuvent avoir une campagne de marketing correspondante pour vendre plus de produits afin de réaliser plus de profits.

Applications : analyse de données de panier de marché, marketing croisé, conception de catalogue, etc.

- **Clustering** : Etant donnée un ensemble d'éléments de données, les méthodes de clustering visent à organiser cet ensemble en un ensemble de groupes (clusters) de sorte que les éléments ayant des caractéristiques similaires soient regroupés dans le même cluster. Un cluster est donc une collection d'objet similaires entre eux et dissimilaires aux autres collections. Lorsqu'une nouvelle donnée est introduite, elle est classée dans le cluster le plus similaire. En effet, il existe de nombreux types de méthode de Clustering [101] :

- Les méthodes de partitionnement
- Les méthodes hiérarchiques
- Les méthodes basées sur la densité
- Les méthodes basées sur des grilles

- Les méthodes basées sur des modèles

L'algorithme K-means [102] est l'un des méthodes de partitionnement les plus connues, est l'un des algorithmes d'apprentissage non supervisés les plus simples qui résolvent des problèmes de clustering bien connus, et il est largement utilisé dans le CRM et le marketing.

- **Classification** : Les techniques de classification de data mining sont les plus couramment utilisées. Fondamentalement, la tâche de classification consiste à prendre chaque instance d'un ensemble de données et à l'assigner à une classe prédéfinie. Le processus de classification des données implique deux phases : l'apprentissage et la classification. En apprentissage, les données d'apprentissage (training data) sont analysées par un algorithme de classification, et dans la phase de classification, les données de teste sont utilisées pour évaluer les performances de classification du modèle.

Les types de modèles de classification :

- Classification par induction d'arbre de décision
- Réseaux de neurones
- Classification bayésienne
- Classification basée sur les associations
- Machines à vecteurs supports (SVM)

- **Prédiction** : Les techniques de prédiction consistent à estimer la valeur d'une variable continue dite variable dépendante ou variable cible, en fonction de la valeur d'une ou plusieurs autres variables dites variables indépendante. A titre d'exemple, dans un modèle de prédiction du prix d'appartement, la valeur cible (à prédire) est le prix d'appartement, tandis que la superficie, l'année de construction, l'étage et le quartier sont les variables indépendantes (explicatives).

- **Sequence mining** : Le sequence mining cherche à découvrir des motifs dans les flux de données dont les valeurs sont délivrées par des séquences dans une période prédéfinie. Son but est de modéliser les états du processus générant la séquence ou d'extraire et de rapporter la déviation et les motifs dans le temps [103]. Cette technique est particulièrement utilisée en biologie pour l'analyse de gènes et des protéines et dans le text mining.

## 1.4.Applications

L'utilisation de méthodes de data mining permet aux entreprises de divers secteurs d'acquérir un avantage concurrentiel à travers les connaissances acquises par ces méthodes. Des exemples de telles utilisations sont fournis ci-dessous.

- Le secteur bancaire : Les transactions bancaires nécessitent le stockage et le traitement de grandes quantités de données. De ce fait, Les banques peuvent utiliser la découverte des connaissances pour diverses applications, y compris :
  - Mesure du risque de crédit : En appliquant des techniques de prise de décision multi critères (*Analytical Hierarchy Proses*) les banques peuvent évaluer les facteurs qui affectent le risque de crédit, et donc construire des modèles de notation de crédit pour décider d'accorder ou non des crédits aux nouveaux candidats [104,105].
  - Détection de la fraude : La fraude est extrêmement coûteuse. En analysant les transactions passées qui ont ensuite été jugées frauduleuses, les banques peuvent identifier ces crimes [106,107].
  - Gestion prédictive du cycle de vie des clients : data mining aide les banques à prédire la valeur à vie de chaque client et à assurer un service approprié à chaque segment (par exemple, en offrant des produits et services exclusifs pour des consommateurs spécifiques) [108,109].
- Le secteur de vente au détail : Les données générées dans ce secteur comprennent généralement des données sur les clients, les produits, les fournisseurs et également des données sur les concurrents. La meilleure source d'information pour les détaillants est le point de vente, les cartes de crédit appartenant à l'entreprise, les cartes de fidélité, etc. Certaines applications ont été créés dans ce secteur, en adoptant les techniques de data mining :
  - Analyse du panier de la ménagère : L'extraction des règles d'association est une technique explicite qui permet aux détaillants d'extraire des informations nécessaires pour comprendre le comportement d'achats de l'acheteur, ce qui permet de prendre les bonnes décisions et donc d'augmenter les ventes [110]. Ainsi l'analyse du panier de la ménagère permet d'extraire des règles d'association qu'on peut les utiliser pour



affecter les produits d'une façon efficace, ce qui facilite la vente croisée et facilite les achats pour les clients.

- Prédiction des ventes : La prédiction des ventes est importante pour les détaillants dans la gestion des stocks. Des méthodes plus efficaces basées sur le data mining ont été proposées pour prévoir les ventes de détail en intégrant des informations concurrentielles, y compris les prix et les promotions [111].
  - Amélioration de la rentabilité de produit : Les détaillants peuvent améliorer la rentabilité de nouveaux produits et services en utilisant les techniques de data mining pour trouver les clients les plus susceptibles de répondre à une offre pour de tels produits ou services.
  - Segmentation de clients : L'utilisation de techniques de Clustering aide les entreprises à avoir une meilleure vision sur leur clientèle [112,113].
- Le secteur de Télécommunication : Dans l'environnement actuel des télécommunications, la concurrence est extrêmement féroce. Les services et les options des clients sont également devenus plus comparables et plus compétitifs. C'est la raison pour laquelle la fidélisation de la clientèle a tendance à s'éroder. La découverte des connaissances dans les télécommunications comprend les éléments suivants :
    - Analyse des enregistrements des appels : Les opérateurs de télécommunication accumulent des enregistrements d'appels détaillés. En analysant ces enregistrements par les techniques de data mining, les opérateurs de télécommunication peuvent déterminer les clients potentiels qui souscriront aux lignes de télécommunications résidentielles supplémentaires [114].
    - Gestion de la perte de client (Customer churn) : Les opérateurs mobiles souhaitent conserver leurs abonnés et satisfaire leurs besoins. Par conséquent, Des études antérieures sur les prédictions de perte de client dans l'industrie des télécommunications ont principalement appliqué des techniques d'exploration de données telles que les réseaux de neurones, les arbres de décision et l'analyse de clusters pour prédire le taux de désabonnement [115,116,117].

- Le secteur agroalimentaire : La gestion efficace des ressources alimentaires et la disponibilité alimentaire sont des principaux piliers de la sécurité alimentaire, y compris la production et l'importation. La découverte des connaissances dans ce secteur comprend les éléments suivants :
  - Prédire la quantité nécessaire de cultures pour satisfaire les besoins des citoyens pour les prochaines années [118].
  - Estimer la taille et le risque de la culture pour aider à planifier la décision de la chaîne d'approvisionnement comme la planification de la production [119].
  - Extraire de données pour évaluer et prédire l'alerte précoce aux risques liés à la sécurité alimentaire [120].
- Commerce électronique : Le secteur de commerce électronique peut utiliser la découverte des connaissances pour diverses applications, y compris :
  - Systèmes de recommandation et personnalisation : Les systèmes de recommandation sont utilisés par un nombre croissant de sites de commerce électronique pour aider les consommateurs à trouver des produits en fonction de leurs attentes. Plusieurs approches sont proposées sur la recommandation basées sur l'évaluation des consommateurs et donne des recommandations sur les produits en utilisant les techniques de filtrage collaboratif, de règles d'association, de Clustering et d'autres techniques [121].
  - Acquérir et fidéliser le client : En utilisant les données de navigation sur un site e-commerce et les historiques d'achats par des techniques d'association et de prédiction, les entreprises en ligne peuvent avoir une meilleure connaissance de comportement d'achat des clients, ce qui permet aux responsables de marketing de prévoir les besoins des clients et l'intérêt à acheter un produit particulier. En utilisant ce type de prédiction, les entreprises en ligne peuvent conserver leurs clients existants en offrant des rabais ou des offres, attirer et acquérir des clients.

## 2. Les techniques de data mining et la prédiction d'attrition

### (churn) : Etat de l'art

Comme il est mentionné dans les chapitres précédents, les techniques de data mining sont très importantes pour le CRM. Dans cette section du chapitre, nous identifions les techniques et les approches de data mining qui ont été développées pour prédire la perte de clientèle, puis nous présentons nos observations sur ces approches.

#### 2.1. Prédiction d'attrition / churn (Revue de littérature)

Le « churn » - en français, on dit « taux d'attrition », est un terme commercial utilisé pour décrire la perte de la clientèle. Il peut être défini comme la propension d'un client à cesser ses transactions commerciales avec une organisation [130]. Dans ce sens, l'identification des clients qui ont un grand potentiel d'abandonner la relation existante avec leurs entreprises, a fait l'objet d'une attention particulière de la part des entreprises de divers domaines. Dans le domaine financier, l'attrition est généralement présentée comme la clôture de compte. Par exemple, Özden Gür Ali, Umut Arıtürk [122] ont utilisé les données bancaires relatives aux événements des clients, et un ensemble de classificateurs binaires pour concevoir un framework dynamique de prédiction du churn qui prédire la perte des clients de la banque, et fournit un classement spécifique de ces clients en fonction de leur probabilité de quitter la banque [122]. Ainsi dans ce domaine, Niels Holtrop, et al [123] proposent une nouvelle approche pour la prédiction de churn dans l'industrie de l'assurance qui combine les principes d'anonymisation et de minimisation des données tout en conservant la forte capacité prédictive du modèle, puis ils comparent ses performances avec une sélection de modèles de référence. Dans le domaine des services de télécommunications, le churn est généralement considéré comme le changement d'opérateur de services. Bingquan, et al [124] présentent un nouvel ensemble de caractéristiques avec sept techniques de modélisation pour améliorer la précision de la prédiction de perte de clients. H. Faris[125] développe un modèle hybride de prédiction du churn en combinant deux approches : *Neighborhood Cleaning Rules* (NCL) et *Particle Swarm Optimization* (PSO) et les comparent avec le modèle de base et d'autres modèles traditionnels. En ce qui concerne les abonnements (abonnement à la télévision payante, abonnements aux journaux), un le phénomène de churn correspond au non-renouvellement d'abonnement après une date d'expiration. Par exemple,

Burez et Van den [126] ont développé un modèle de prédiction de churn et le teste en utilisant les données d'une société de télévision payante. Coussement et Van den [127] ont utilisé « Machines à vecteurs supports » pour prédire la probabilité qu'un client annule leur abonnement à un journal. Dans le commerce de détail d'épicerie, le churn a été considéré comme la défection partielle et progressive des clients. V.L. Miguéis, et al [128] construisent un modèle de prédiction basé sur trois techniques de classification pour prédire la défection partielle dans l'épicerie au détail. Ssu-Han Chen [129] ont proposé deux modèles prédictifs dans le commerce de détail pour détecter la perte partielle des clients. Dans les environnements en ligne, les clients perdus sont considérés comme ceux qui ont été fidèles aux services d'un site web puis ils ont arrêté de visiter ou de faire leurs achats sur ce site depuis une longue durée. Par exemple, [32] a proposé un mécanisme de prédiction individuelle de churn et le tester avec une base de données du monde réel d'un site de rencontres à Taïwan. Clemente-Císcar, et al [131], ont développé un modèle de prédiction de churn adapté au secteur du commerce électronique B2B, et ils ont évalué sa performance par rapport à d'autres modèles de prédiction.

Table 3.2 Les modèles de prédiction de churn.

|   | Secteur                          | Paramètre d'environnement                        | Type de défection  | Définition de churn   |
|---|----------------------------------|--|--------------------|---|
| Buckinx and van den Poel (2005) [157]         | Commerce de détail               | Environnement hors ligne / Cadre non contractuel | Partielle          | Le client change leur comportement d'achat  |
| Shin-Yuan Hung, et al. (2006) [132]           | Télécommunication                | Environnement hors ligne / Cadre contractuel     | Totale             | Un client perdu est défini comme un abonné qui a quitté volontairement l'opérateur.   |
| Jae-Hyeon Ahna, et al. (2006) [133]           | Télécommunication                | Environnement hors ligne / Cadre contractuel     | Partielle & Totale | Changement dans le statut d'un client   |
| Burez and van den Poel (2007) [156]           | Abonnement de télévision payante | Environnement hors ligne / Cadre contractuel     | Totale             | L'abonné ne renouvelle pas son contrat ou ne la paye pas.   |
| Xie, et al. (2009) [134]                      | Services financiers              | Environnement hors ligne / Cadre contractuel     | Totale             | Non mentionné   |
| Xiaobing Yu, et al. (2011) [138]              | E-commerce                       | Environnement en ligne / Cadre non contractuel   | Totale             | Le client cesse de faire des affaires.  |
| V.L. Miguéis, et al. (2012) [129]             | Commerce de détail               | Environnement hors ligne / Cadre non contractuel | Partielle          | Les clients qui n'ont rien acheté depuis une certaine période, ou ceux qui ont dépensé dans les périodes ultérieures une somme moins de 40% du montant dépensé au cours de la période de référence. |
| Bingquan Huang, et al. (2012) [125]           | Télécommunication                | Environnement hors ligne / Cadre contractuel     | Totale             | Non mentionné   |
| Kristof Coussement, et al. (2013) [135]       | Jeux en ligne                    | Environnement en ligne / Cadre non contractuel   | Totale             | Le joueur qui n'a pas joué pendant une période déterminée.  |
| M. Clemente-Císcar, al. (2014) [131]          | Commerce de détail               | Environnement hors ligne / Cadre non contractuel | Partielle          | Ensemble de définitions de churn.   |
| Faris (2014) [126]                            | Télécommunication                | Environnement hors ligne / Cadre contractuel     | Totale             | L'utilisateur quitte la société.  |
| A. T. Jahromi, et al. (2014) [188]            | E-commerce (B2B)                 | Environnement en ligne / Cadre non contractuel   | Totale             | Inactivité dans la seconde moitié de l'année.   |
| K. Kyoungok et al. (2014) [116]               | Télécommunication                | Environnement hors ligne / Cadre contractuel     | Totale             | Les clients qui quittent le service ou migrent vers les concurrents à une date proche d'une date précise.   |
| M.A.H. Farquod, et al (2014) [136]            | Services financiers              | Environnement hors ligne / Cadre contractuel     | Totale             | Transférer les fidélités d'un fournisseur de services à un autre.   |
| Ozden Gur Ali and Umut Arıturk (2014) [123]   | Services financiers              | Environnement hors ligne / Cadre contractuel     | Totale             | Un client est déclaré défectueux si la taille de son portefeuille tombe en dessous d'une valeur seuil spécifique et reste dans cette situation pendant six mois consécutifs.                        |
| Ssu-Han Chen (2016) [130]                     | E-commerce                       | Environnement en ligne / Cadre non contractuel   | Totale             | Client avec un petit nombre sessions pendant les 8-12 mois qui suivent son inscription sur le site e-commerce.  |
| Niccolò Gordinia, Valerio Vegliob (2017) [32] | E-commerce (B2B)                 | Environnement en ligne / Cadre non contractuel   | Totale             | Lorsque la société a été active (c'est-à-dire qu'elle a réalisé au moins une transaction dans l'année) mais n'a aucune activité (c'est-à-dire, achat) dans l'année.                                 |
| N. Holtrop, et al. (2017) [124]               | Assurance/Télécommunication      | Environnement hors ligne / Cadre contractuel     | Totale             | Un client est considéré comme perdu, s'il a été avec la société en début de l'année mais n'est plus à la fin de l'année.  |
| Cette étude                                   | E-commerce (B2C)                 | Environnement en ligne / Cadre non contractuel   | Partielle & totale | Déplacement du client d'un segment avec une valeur importante à l'instant T vers un autre segment de valeur inférieure à l'instant T+1.   |

La table 3.2 résume les modèles de prédiction de churn rapportés dans la littérature récente. Les caractéristiques distinctives de chaque étude en termes de champ d'application, de paramètre d'environnement, de type de défection, et de définitions de churn sont fournies.

En observant la table 3.2 qui résume quelques travaux principaux sur la prédiction de churn, nous pouvons extraire deux remarques majeures : (1) Type d'environnement : La majorité des études qui traitent l'attrition dans les environnements hors ligne sont effectuées dans un cadre contractuel, qui est caractérisé par l'existence d'un contrat entre l'entreprise et ses clients. Dans un tel cas, le churn est clairement définie et facile à prédire, il correspond souvent à la date d'annulation du contrat ou d'abonnement. (2) la deuxième remarque est liée au type de la défection (Partielle ou Totale) : la plupart de ces études considèrent la défection totale, alors que peu parmi elles utilisent des modèles de prédiction pour identifier la défection partielle [128], [157], [131]. De plus, chacune de ces études définissent le churn différemment, ce qui soulève la question suivante : laquelle est la plus appropriée ?

La table 3.2 (la colonne type de défection) révèle qu'aucune recherche antérieure n'a traité la prédiction des deux types de défection simultanément (partielle et totale) dans les environnements en ligne non contractuel. Elles se concentrent soit uniquement sur la défection totale, soit uniquement sur la défection partielle. Par conséquent, la présente étude contribue à la littérature existante de deux façons importantes. Tout d'abord, elle combine le modèle LRFM et les techniques de clustering pour arriver à définir le concept de défection partielle et totale dans un cadre non contractuel, en particulier dans l'industrie du commerce électronique. Deuxièmement, elle introduit des techniques de classification pour construire des modèles de prédiction permettant de prédire la défection partielle et totale afin de minimiser le risque de perte de clientèle.

Le tableau 3.2 révèle que la question de prédiction de churn a fait l'objet de recherches suffisantes dans le secteur du commerce électronique. De plus, toutes les analyses considèrent la défection totale. Pour découvrir la défection partielle et totale dans ce secteur, notre étude utilise dans un premier temps la segmentation des clients en combinant le modèle LRFM et l'algorithme de clustering (k-means) pour segmenter tous les clients en clusters homogènes selon leurs valeurs L, R, F et M, puis un LRFM pattern sera affecté à chaque cluster [137]. Le changement dans le modèle LRFM des clients (Déplacement d'un client d'un cluster avec une valeur importante à

l'instant T vers un autre cluster de valeur inférieure à l'instant T+1) peut être considéré comme un signal de défection partielle ou totale.

Contrairement aux recherches qui visent à ne conserver que les clients rentables [128], ou celles qui consacrent beaucoup d'efforts à l'ensemble de la clientèle [32], [138], [139], notre étude ne se focalise pas seulement sur les principaux clients, mais aussi sur ceux qui démontrent un changement positif dans leur comportement d'achat, même s'ils sont regroupés dans des clusters qui ne contribuent pas positivement aux bénéfices.

D'une part, la création d'un programme de rétention ciblant tous les types de clients coûtera trop cher pour l'entreprise. D'autre part, en adoptant une stratégie axée uniquement sur les clients rentables dans le but de les retenir, les entreprises, notamment celles travaillant dans le domaine du commerce électronique, peuvent perdre ceux qui ont le potentiel de devenir des clients rentables à cause de leur exclusion, ce qui conduira à une augmentation du taux de churn suivie d'une baisse au niveau de profits. En effet, ces clients ont vraiment besoin d'attention de la part de l'entreprise. Ils devraient être placés dans une autre catégorie au lieu d'être éliminés. C'est un point important car aucune entreprise ne veut rater l'opportunité de convertir un client précédemment insatisfait en un client fidèle. Comme il est mentionné ci-dessus, ces clients sont ceux qui démontrent un changement positif dans leur comportement d'achat, même s'ils sont regroupés avec les clients insuffisamment profitables. L'identification de ces clients sera discutée dans les sections suivantes.

Par exemple, dans une situation où l'objectif d'une entreprise de vente en ligne est de retenir que les clients rentables, elle devrait comprendre dans un premier temps, pourquoi ses clients la quitte et partent chez des concurrents. Une analyse de churn pour le segment de clientèle rentable montre que certains clients quittent le e-commerce parce que les frais de livraison ne sont pas gratuits. Par la suite, l'entreprise décide de réduire les coûts de livraison pour les clients les plus rentables afin de les garder et les inciter à continuer de faire des achats. Cependant, les clients moins rentables ne sont pas servis avec cette réduction. Ainsi, seuls les clients rentables sont satisfaits. Par conséquent, cibler uniquement les clients rentables n'est pas une stratégie optimale pour augmenter le taux de rétention, ainsi les pratiques ont démontré les limites de cette vision car un client qui était rentable dans le passé, n'implique pas qu'il continuera à l'être dans le futur [140].

## **Conclusion**

Dans ce chapitre, nous avons défini les notions d'extraction de connaissances à partir des données et les techniques de data mining, ainsi nous avons décrit différentes applications des techniques de data mining dans le CRM, et plus particulièrement dans les stratégies de rétention. Dans la deuxième partie de ce chapitre, nous avons passé en revue la littérature des modèles de prédiction d'attrition. En effet, aucune recherche antérieure n'a porté à la fois sur l'attrition partielle et totale dans les environnements non contractuel. Toutes les approches existantes se basent soit uniquement sur l'attrition partielle, soit uniquement sur l'attrition totale. Dans la mesure où ces approches ne considèrent qu'un seul type d'attrition, ces approches sont fondamentalement sous-optimales. Cette situation s'explique largement par le fait que la mise en œuvre d'une stratégie de rétention efficace et optimale nécessite une connaissance préalable de degré de risque, car il est conseillé de ne pas concentrer les efforts et les ressources dédiés à la gestion d'attrition de la même manière pour les deux niveaux de défection, mais il faut optimiser l'affectation des ressources en fonction de degré de défection [6].

Pour cela, dans le chapitre suivant nous allons essayer de concevoir une nouvelle approche qui contribue de façon importante à la littérature.



# Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision.

## Introduction

De nos jours, avec le nombre croissant de sites e-commerce, les clients ont beaucoup plus de choix qu'auparavant, ce qui permet de les inciter à répartir leurs achats entre plusieurs sites de commerce électronique, de faciliter la comparaison des produits et services concurrents, ainsi d'augmenter le risque de perte de clients (l'attrition). En outre, le secteur du commerce électronique souffre de la nature de la relation entre les entreprises et leurs clients (un cadre non contractuel) ; parce que dans un tel cas, les clients ont la possibilité de changer leurs comportements d'achat sans en informer le vendeur. Face à ces contraintes, la rétention des clients en les évitant de passer une partie de leurs achats à un concurrent est devenue un défi commun pour de nombreuses entreprises de commerce en ligne.

Pour certaines raisons, la satisfaction de la clientèle peut diminuer, ce qui conduit à une crise dans la relation avec le vendeur, et sera ensuite complètement détruite si un concurrent peut offrir un produit similaire. En d'autres termes, l'insatisfaction des clients augmente directement leurs intentions de recourir ou de commutation aux services de concurrents. De plus, l'attractivité des alternatives telles que les programmes de promotion ou les prix attractifs offerts par des concurrents affectent négativement la satisfaction du client, car elles peuvent attirer les clients quel que soit leur niveau de satisfaction [140,141]. Par conséquent, la satisfaction et l'attractivité des alternatives déterminent alors la force et la stabilité des relations [142,143,144]. De nombreuses recherches ont examiné la relation entre la satisfaction des clients et son effet sur la rétention de ces derniers. Alshurideh, et al [145] et Danesh, et al [146] ont constaté que la satisfaction de la clientèle a un effet direct sur la rétention de la clientèle. Rao, et al [147] soutiennent le lien entre la satisfaction et la rétention de la clientèle. Omotayo, et al [148] ont affirmé que la satisfaction est la prémisse nécessaire (ainsi que d'autres facteurs) pour la rétention de la clientèle. D'après ces conclusions, nous pouvons dire que plus les clients sont satisfaits, plus leur taux de rétention est grand. Si les entreprises sont en mesure de satisfaire les clients, cela créera une rétention des clients suivie par des profits [149]. En d'autres termes, parmi les

principaux facteurs qui augmentent la rentabilité d'une entreprise, est sa capacité à retenir ses clients fidèles, parce que les clients fidèles achètent fréquemment de grandes quantités et sont également moins sensibles au prix. Ainsi les clients fidèles sont beaucoup plus réceptifs que les nouveaux clients, ce qui diminue les coûts de vente. Ces clients peuvent aussi contribuer à l'augmentation des profits en partageant leur histoire positive avec leurs amis et leur famille, car ils sont plus susceptibles d'offrir un bouche-à-oreille favorable.

Des études antérieures ont montré qu'en adoptant des approches de rétention appropriées, les clients conservés produisent des revenus et une marge plus élevée que les nouveaux [150]. En fait, plusieurs études soulignent également l'importance d'augmenter les dépenses pour fidéliser et retenir les clients existants que d'en acquérir de nouveaux [151,139]. De plus, selon la recherche effectuée par Reichheld FF & Sasser WE [153], la rétention des clients est une stratégie utile pour réduire les frais administratifs et pour augmenter les profits, ils ont prouvé que même une augmentation de 5% au niveau de taux de rétention des clients pourrait entraîner une augmentation substantielle de la rentabilité des entreprises (de 25% à 85% dans l'ensemble des industries).

Comparé aux environnements hors ligne, le coût d'acquisition des clients dans les environnements en ligne comprend généralement de petites mesures : le coût dépensé par visite (le coût d'acquisition du visiteur), le pourcentage de ceux qui se transforment en prospects (le coût d'acquisition du prospect) et enfin le pourcentage des clients qui ont effectivement validé un achat (le coût de l'acquisition du client). En raison du coût élevé de l'acquisition de nouveaux clients dans le contexte e-commerce, de nombreuses relations clients deviennent non rentables pendant un certain temps [155], car l'action d'acquérir un nouveau client reste souvent bloquée dans la phase de départ (phase de conversion des visiteurs en prospects) de la relation commerciale. À ce stade, les entreprises doivent dépenser plus de ressources marketing pour convertir les prospects en clients. Les relations puissent après génèrent des profits dans les transactions ultérieures, lorsque le coût de la fidélisation des clients diminue [153]. En plus des coûts élevés consacrés à attirer de nouveaux clients dans le secteur de e-commerce, la facilité de passage d'un site marchand à un autre rend encore plus difficile la rétention des clients existants [154]. D'autre part, la dissémination des insatisfactions par le biais du bouche-à-oreille négatif se propage plus rapidement dans l'environnement de commerce électronique que dans l'environnement hors ligne.

Dans ce chapitre, nous abordons la question suivante qui devient de plus en plus importante pour les gestionnaires du commerce électronique. Comment pouvons-nous éviter la réduction de taux de rétention des clients ? En d'autres termes, comment pouvons-nous réduire le taux d'attrition des clients dans le contexte du commerce électronique ?

Il existe deux approches fondamentales pour résoudre ce problème: les «approches non ciblées» qui s'appuient sur des produits de qualité supérieure et sur une publicité de masse pour augmenter la fidélité de la marque et la rétention des clients, et les «approches ciblées» qui reposent sur l'identification des clients susceptibles de partir vers les concurrents et ensuite les cibler par des incitations directes ou de personnaliser un plan de service pour les inciter à rester et à continuer de réaliser des achats [32], [122], [156], [157], [188]. Dans notre travail, nous allons nous référer à la deuxième approche, pour cela, nous allons examiner si nous sommes capables d'identifier le moment où les clients commencent à interrompre leur relation avec le site e-commerce pour éviter la défection totale de ces clients, parce qu'à long terme la défection partielle peut ensuite conduire à la défection totale [128]. En outre, en ce qui concerne la défection partielle et totale, notre approche utilise une combinaison de techniques de Clustering de data mining et le modèle LRFM (longueur, récence, fréquence et montant) pour définir ces deux concepts dans le cadre du commerce électronique.

Le reste de ce chapitre est organisé comme suit : la section 2 fournit une brève présentation de quelques concepts qui sont indispensables à la compréhension et à la réussite de notre approche. Ensuite, l'approche de recherche basée sur le standard CRISP-DM est brièvement décrite dans cette section, suivie d'une étude empirique dans la section 3 pour démontrer comment cette approche de prédiction d'attrition fonctionne dans la pratique, alors que dans la quatrième section nous discutons et interprétons les résultats obtenus. Et à la fin du chapitre, nous terminons par une conclusion.

## 1. Méthode de recherche

L'objectif principal de notre travail est de concevoir une nouvelle approche dont le but principal est la réalisation d'un système d'aide à la décision basé sur le processus d'extraction de connaissances à partir de données pour prédire l'attrition dans les contextes de e-commerce. En se basant sur le modèle de référence CRISP-DM, nous présentons toutes les étapes indispensables pour la compréhension et la réussite de notre approche.

### 1.1.Phase de compréhension du problème

A mesure que le nombre de site e-commerce augmente, les entreprises de commerce électronique sont soumises à une concurrence croissante. Cela peut être expliqué par le fait que plus le nombre de site e-commerce augmente plus les clients auront beaucoup de choix, ce qui les encourage à diviser leurs achats sur plusieurs sites et à comparer les produits et les services de concurrents, ce qui augmente finalement le risque d'attrition (perte de client) suivie d'une baisse au niveau de profits des entreprises. Ces dernières doivent trouver le moyen de préserver leurs bénéfices. L'une des solutions proposées est d'adopter une stratégie de rétention dans laquelle les entreprises de commerce électronique peuvent prédire les clients fidèles en voie de quitter partiellement ou totalement le site marchand, afin que ces entreprises puissent anticiper un tel comportement défectueux. Une autre complication réside sur la difficulté de définir l'attrition, car les caractéristiques qui doivent être observées pour dire qu'un client a totalement ou partiellement défectueux ne sont pas clairement définies dans les cadres non contractuels où le domaine de e-commerce est fait parti. Cependant, comme nous l'avons déjà mentionné, nous nous concentrons sur deux types de clients : les clients qui appartiennent aux clusters représentent les principaux clients de l'entreprise, et ceux qui ont insuffisamment rentables, mais qui montrent un changement positif dans leur comportement d'achat. Par conséquent, il est nécessaire pour une entreprise, en tant que première étape essentielle, de développer des outils de profilage pour mieux identifier les différents types de profils de clients existant dans sa base de données clients, et ensuite faire de son mieux pour les préserver ou, au moins, les plus fidèles, car l'identification de ces segments peut être la base d'une stratégie efficace pour cibler et retenir les bons clients [156].

D'après ce bref descriptif de la problématique, nous fixons les objectifs commerciaux et ceux du data mining comme suit :

- **Les objectifs commerciaux :**

- Augmenter le taux de rétention de clients en évitant leur défection.
- Segmenter les clients en plusieurs segments selon des critères précis, afin de bien identifier les différences et les similitudes dans leur comportement pour prendre finalement les bonnes décisions et pour mener des actions marketing ciblées et personnalisées.

Il existe de nombreux autres problèmes commerciaux liés à ces deux principaux objectifs commerciaux :

1. L'inactivité d'un client dans un intervalle de temps supérieur à un nombre de jours donné est suffisante pour le classer comme un client défectueux ? Si oui, s'agit-il d'une défection partielle ou totale ?
2. Si non, est ce qu'il y a une définition inclusive d'attrition fondée sur d'autre critères ?
3. Quels sont les critères pour décider quel client doit être gardé ?
4. La valeur client présente un critère suffisant dans la prise de décision ?
5. Si oui, comment définissons-nous cette valeur ?
6. Comment optimiser les efforts et les ressources dédiés à la gestion d'attrition ?

Le bénéfice commercial attendu est de réduire le taux d'attrition. Cela va influencer positivement sur les ventes. Cette approche sera considérée comme une réussite si le système décisionnel final puisse prédire les clients qui sont en voie de défection afin de les inciter par des stratégies de rétention.

- **Les objectifs du data mining :**

L'objectif principal de data mining est d'utiliser l'historique des précédentes transactions et sessions pour développer un modèle de prédiction d'attrition adapté au secteur du commerce électronique B2C. Ce modèle prédictif sera précédé par un modèle de clustering qui va servir à identifier les différents types de profils clients et d'analyser la valeur de chacun de ces types.

L'objectif secondaire de data mining est d'identifier et d'examiner les variables prédictives qui contribuent davantage à la prédiction de la défection partielle et totale des clients.

Le critère de réussite du data mining est la qualité des modèles qui seront créés. Pour les modèles prédictifs cette qualité sera évaluée par la mesure Macro-averaging comprenant

l'efficacité moyenne, la précision<sub>M</sub>, Rappel<sub>M</sub> et F-score<sub>M</sub>. Pour le modèle de clustering, la qualité de segmentation des clients sera évaluée par l'intégration de deux méthodes : la méthode du coude (Elbow method) et la méthode de silhouette.

## 1.2.Phase de compréhension des données

Les données analysées dans cette étude ont été fournies par une entreprise de vente en ligne au Maroc spécialisés dans l'électronique, la mode, les appareils ménagers et les articles pour enfants. Lorsque les clients visitent le site e-commerce, le serveur MySQL enregistre leur connexion, déconnexion, processus d'achat et l'état final de chaque session. Un client peut effectuer quatre types d'événements, à savoir « Session avec produits consultés », « Session avec ajout au panier », « Session avec check-out » et « Session avec transaction ». En ce qui concerne le mode de paiement, les clients ont le choix entre quatre modes de paiement : paiement à la livraison, carte bancaire, virement bancaire et paiement en trois fois.

### - Collecte initiale de données :

L'entreprise de commerce en ligne de notre étude utilise les sources de données suivantes :

- **La base de données des clients :** Cette base de données contient les coordonnées, et les données démographiques sur les clients (nom, sexe, âge, adresse, ville, pays, profession, ..)
- **Les données sur les transactions :** Lorsque le client valide une commande, toutes les informations liées à cette action sont enregistrées (identifiant du client, référence de produit, date de transaction, montant, quantité achetée, le mode de paiement, adresse de livraison, ville de livraison ...).
- **Les données sur les sessions :** Toutes les sessions ouvertes par les clients sur le site e-commerce sont enregistrées même qu'elles n'ont pas terminées avec une transaction. Ces informations comprennent l'identification du client, la date de connexion et de déconnexion, la durée passée lors d'une session, le statut de session (session avec consultation de produits, session avec ajout au panier, session avec check-out et session avec transaction), l'adresse ip, le terminal utilisé, système d'exploitation utilisé, etc...

La société dispose aussi une base de données des produits qui contient des informations détaillées sur ses produits (référence produit, libellé, description, prix, quantité disponible,

catégorie). Mais dans notre étude nous allons focaliser seulement sur les coordonnées de clients, les habitudes de navigations et les données sur les achats.

- **Description et exploration de données :**

L'ensemble de données comprend les coordonnées, les habitudes de navigations et les habitudes de consommation de 2783 clients qui ont visité le site marchand de l'entreprise (voir table 4.1). Les évènements effectués par ces clients sur le site e-commerce pour la période du 1er novembre 2013 au 28 février 2015 ont été utilisés.

Table 4.1 – Les données utilisées et leurs attributions.

| Type de données        | Attribution de données  |
|------------------------|---|
| Données démographiques | Idclient, nom, sexe, age, adresse, profession, email, mobile phone, register date             |
| Données sur les accès  | Customer id, login ip, login date, login page   |
| Session                | Idsession, Idclient, session date, session state  |
| Transaction            | Idtransaction, Idclient, date_transaction, idproduit, nomproduit, description, prix, quantité |

Basé sur la connaissance du contenu des variables, il est maintenant possible de déterminer laquelle d'entre elles sera exclu dans les étapes prochaines. Cette exclusion initiale de certaines variables simplifiera le travail ultérieur avec l'ensemble de données. En outre, certaines variables de type symbolique (date et heures, page consultées) doivent être exploitées pour créer d'autres variables numériques, telles que la durée passée dans une session, le nombre total de session. Ainsi, l'exploration de la base de données de transactions et de sessions permet d'obtenir des récapitulatifs intéressants sur les clients, tels que le montant qu'ils dépensent, la fréquence des achats, la longueur de la relation, la récence, le nombre de sessions totales, le nombre de sessions abandonnées, le taux d'abandon de panier, etc...

**1.3.Phase de préparation des données (conception et alimentation du data warehouse)**

Comme nous avons déjà mentionné dans la description des étapes du modèle CRISP-DM, l'étape de préparation de données représente l'étape la plus importante et la plus coûteuse en

temps dans la conception des systèmes décisionnel global, elle représente selon les estimations, de 50 à 70% des efforts et des temps consacrés à un projet data mining.

Les outils d'extraction, de transformation et de téléchargement (ETL) sont utilisés dans cette étape pour sélectionner les données depuis l'ensemble des sources et de les fusionner. Ensuite l'opération de sélection d'enregistrement et des attributs doit être effectuée. La sélection des attributs dépend en grande partie des objectifs de data mining souhaités. Par exemple dans notre cas la portée de l'étude sera limitée aux clients qui ont effectué des transactions entre le 1er novembre 2013 et le 28 février 2015, et pour les attributs nous allons focaliser sur ceux qui peuvent nous décrire le comportement d'achat de client, à titre d'exemple : les montants dépensés, les dates de transactions, les dates de connexion et de déconnexion, l'état finale de chaque session, et les modes de paiement préférés pour chaque client. Il convient donc de définir des filtres afin d'exclure les clients qui n'ont pas effectué des transactions dans la période d'étude.

La figure 4.1 représente les tables constituant la base de données source.

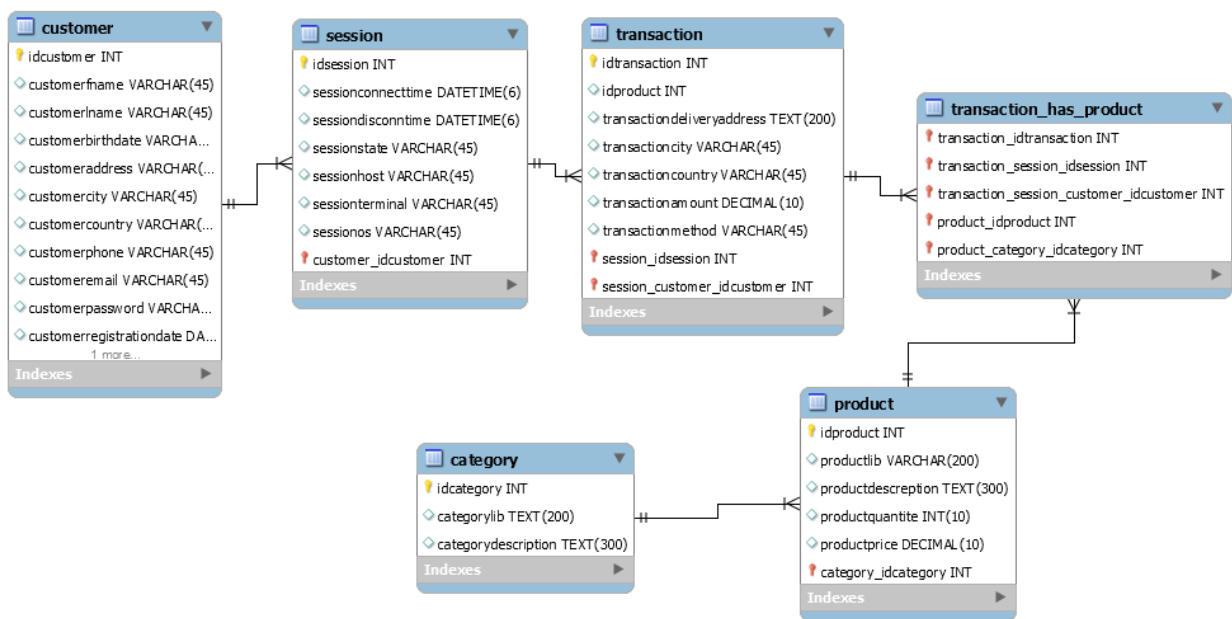


Figure 4.1- Schéma relationnel des tables.

A partir de la base de données source, nous avons utilisé les outils ETL de la plateforme KNIME ANALYTICS 3.4 pour créer le workflow illustré dans la figure 4.2, et qui va nous permettre de construire notre magasin de données (data mart) dont le schéma en étoile est montré



dans la figure 4.3. Les tables de 4.2 à 4.4 résument brièvement et expliquent ce workflow. La conception de ce magasin de donnée a été particulièrement focalisé sur le client pour extraire seulement les variables nécessaires à la réussite de notre approche.

Table 4.2 Tables de la base de données source

| Nom table       | Type        |
|-----------------|-------------|
| tab_client      | Master      |
| tab_Produit     | Master      |
| tab_Catégorie   | Master      |
| tab_Session     | Transaction |
| tab_Transaction | Transaction |

Table 4.3 Dimensions et table du fait du data mart

| Nom table              | Type      |
|------------------------|-----------|
| Dim_client             | Dimension |
| Dim_transaction        | Dimension |
| Dim_session            | Dimension |
| Dim_time               | Dimension |
| Fact_customer_behavior | Fait      |

Table 4.4 Les actions utilisées dans la création et l'alimentation du data mart.

| Workflow                          | Source                  | Cible                  | Actions       |
|-----------------------------------|-------------------------|------------------------|---------------|
| Chargement_dim_customer           | tab_client              | Dim_client             | Update/Insert |
| Chargement_dim_session            | tab_session             | Dim_session            | Insert        |
| Chargement_dim_transaction        | tab_transaction         | Dim_transaction        | Update/Insert |
| Chargement_dim_time               | tab_session             | Dim_time               | Update/Insert |
| Chargement_fact_customer_behavior | Jointure des dimensions | Fact_customer_behavior | Update/Insert |

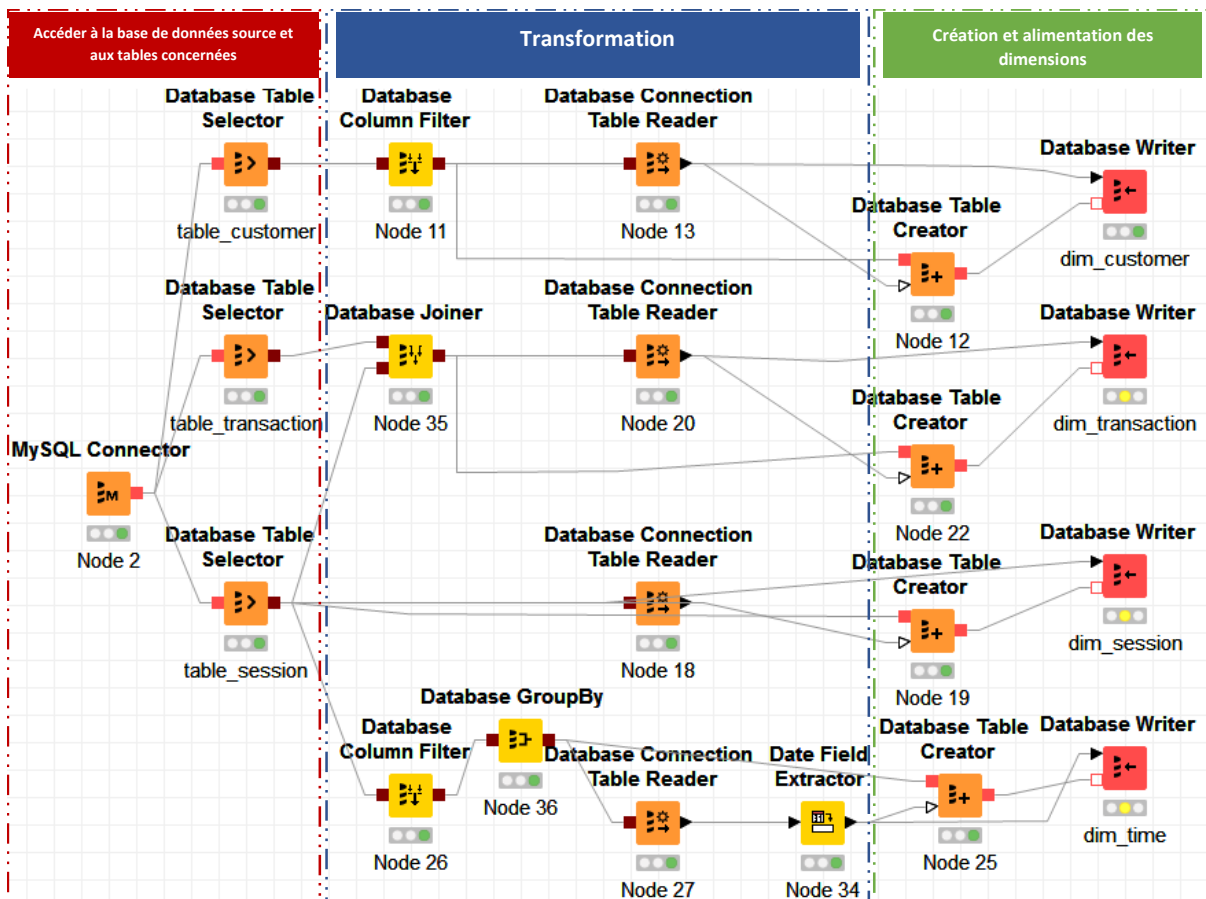


Figure 4.2- Le workflow utilisé pour la création et l'alimentation des dimensions de notre data mart (Kettle Analytics Platform).

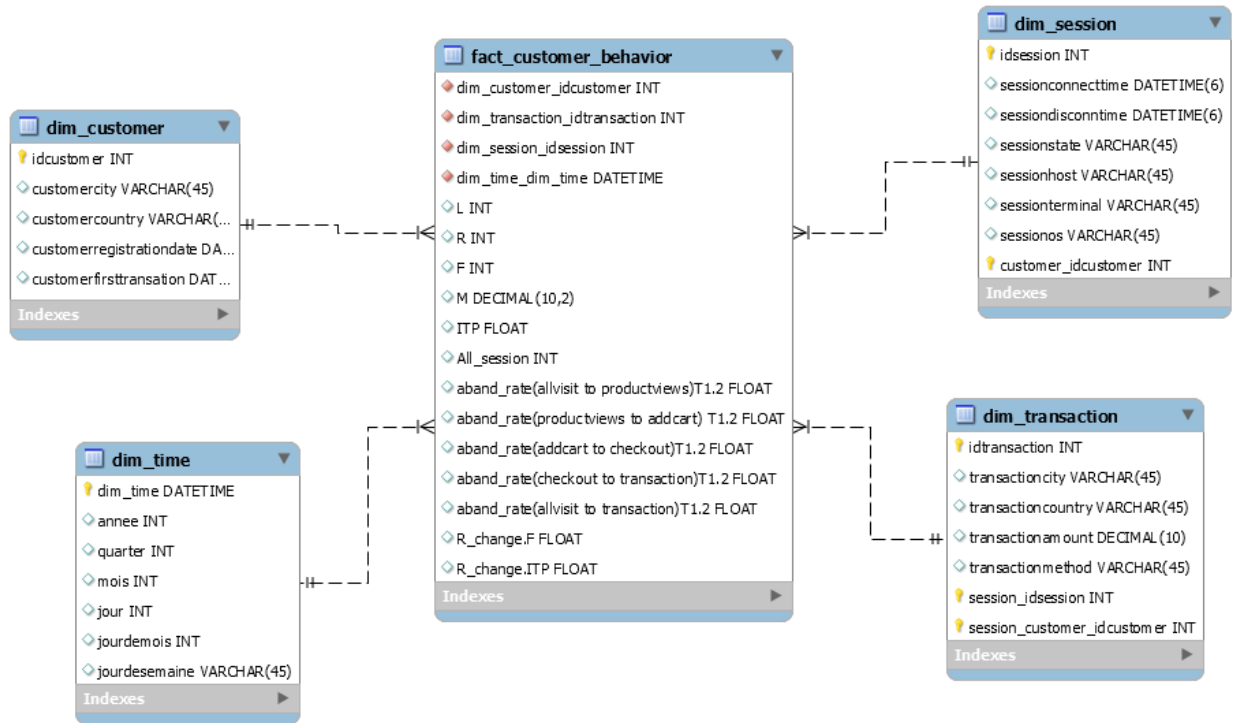


Figure 4.3- Modélisation des données en utilisant le schéma en étoile.

Nous rappelons que notre approche de prédiction d'attrition est divisée en deux sous étapes :

- La première étape : Sert à développer un modèle de segmentation basé sur l'algorithme k-means et sur le modèle LRFM pour identifier et analyser la valeur de chaque segment et aussi pour arriver à la définition d'attrition.
- La deuxième étape : Dans cette étape des modèles de prédiction basés sur des techniques de classification multi-classes seront développés pour prédire à la fois les deux degrés d'attrition (partielle et totale).

Pour atteindre ces deux objectifs, la table du fait de notre data warehouse doit fournir toutes les mesures nécessaires à la création de ces deux modèles (un modèle de clustering et trois modèles de prédiction). Pour cela nous allons détailler dans la section de modélisation ces deux étapes, en présentant les techniques, les méthodes et les mesures nécessaires pour la création de nos modèles.

## **1.4.Phase de modélisation**

### **1.4.1. Modèle de segmentation :**

#### **1.4.1.1. Segmentation des clients**

La segmentation du marché est le processus d'identification des groupes clés du marché général qui partagent des caractéristiques et des habitudes de consommation spécifiques [159-160]. Aujourd'hui, le modèle RFM (récence, fréquence et montant), proposé par Hughes [163], est l'une des méthodes les plus courantes pour segmenter et identifier les valeurs des clients dans les entreprises. En outre, les techniques de clustering ont été largement utilisées pour segmenter les clients lors de l'utilisation du modèle RFM [137],[158-159], [161-162]. Dans cette section, nous examinons k-means comme technique de clustering et le modèle LRFM comme version étendue du modèle RFM qui prend en compte la longueur de relation client (L) que nous utilisons pour la tâche de profilage de client.

#### **1.4.1.2. Les modèles RFM et LRFM**

Récence (R), Fréquence (F) et Montant (M) décrivent le comportement d'achat des clients dans le passé à partir des aspects suivants :

Récence : Quand le client a-t-il fait son dernier achat ?

Fréquence : À quelle fréquence le client effectue-t-il un achat ?

Montant : Combien d'argent le client dépense-t-il ?

La méthode RFM est considérée comme une méthode d'analyse de la clientèle très simple mais efficace pour résoudre ce genre de questions. La méthode RFM est donc une technique d'analyse qui peut être utilisée pour analyser les comportements des clients et définir le segment de marché. [161] décrit que le principal atout de la méthode RFM est, d'une part, d'obtenir une analyse comportementale des clients pour les regrouper en clusters homogènes et, d'autre part, de développer un plan marketing adapté à chaque segment de marché. L'analyse RFM améliore la segmentation du marché en examinant quand (R), à quelle fréquence (F), et l'argent dépensé (M) pour un article ou un service particulier. A. X. Yang [164] résume que les clients qui ont acheté le plus récemment, le plus souvent et qui ont dépensé le plus d'argent seraient beaucoup plus susceptibles de réagir aux futures promotions.

Certains chercheurs tentent de développer de nouveaux modèles RFM en y ajoutant des paramètres supplémentaires afin d'examiner s'ils obtiennent ou non de bons résultats par rapport au modèle RFM de base [165-168]. Par exemple, Chang and Tsay [168] proposent le modèle LRFM, en prenant en compte la longueur de la relation client, pour résoudre le problème du modèle RFM lié à la difficulté de distinguer entre clients qui ont une longue relation avec l'entreprise et ceux qui ont une relation courte. En outre, S. Chow et R. Holden [169] suggèrent que la fidélité et la rentabilité du client dépendent de la relation entre l'entreprise et ses clients. À cet égard, afin d'identifier les clients les plus fidèles, il est nécessaire de prendre en compte la longueur de relation du client (L), où L est défini comme le nombre de périodes (par exemple nombre de jours) du premier achat au dernier achat dans la base de données.

#### **1.4.1.3. La Méthode de clustering K-means**

K-means est l'algorithme le plus couramment utilisé pour regrouper  $n$  vecteurs basés sur des attributs en  $k$  partitions, où  $k < n$ , en fonction de certaines mesures. Le nom vient du fait que  $k$  clusters sont identifiés et le centre d'un cluster est la moyenne de tous les vecteurs dans ce cluster. L'algorithme commence par choisir aléatoirement  $k$  centroïdes initiaux, puis assigne les vecteurs au centroïde le plus proche en utilisant la distance euclidienne et recalcule les nouveaux centroïdes comme moyen des vecteurs de données assignés. Ce processus est répété plusieurs fois jusqu'à ce que les vecteurs ne modifient plus les clusters entre les itérations [170]. Chaque vecteur de l'espace correspond à un client dont les attributs sont les variables L, R, F et M. Ensuite, chaque client est positionné dans le cluster de centre le plus proche. Une fois les groupes de clients formés, la position des centres est recalculée et l'opération réitérée jusqu'à l'obtention d'un état stable.

Dans la technique k-means, le nombre de clusters est sélectionné aléatoirement, ce qui signifie que le résultat du clustering deviendra peu fiable si le nombre supposé de clusters est incorrect [171], ceci soulève la question fondamentale suivante : Comment choisir le meilleur nombre de clusters attendus ( $k$ ) ?

Certains types d'index de qualité de clustering efficaces peuvent aider à déterminer le meilleur nombre ( $k$ ). Dans notre approche, nous avons utilisé deux méthodes pour déterminer le nombre optimal de clusters pour k-means. Ces méthodes consistent à optimiser un critère, tel que la somme des carrés WCSS (Within-Cluster Sum of Squares) et la silhouette moyenne. Les

méthodes correspondantes sont respectivement nommées la méthode du coude (Elbow method) et la méthode de silhouette. Dans cette étude, la somme des erreurs au carré (SSE) et le coefficient de silhouette moyen présentées par les équations Eq.1 et Eq.2 seront ensuite combinés afin de mesurer la qualité du regroupement et de déterminer le nombre de clusters le plus optimal. Plus précisément, nous avons appliqué la technique k-means sous différentes valeurs de k (k=2,3....10), puis nous avons tracé les courbes de SSE et le coefficient de silhouette moyen (Avg-Silhouette-Width) par rapport aux nombres de clusters K pour les analyser et d'identifier le nombre optimal de clusters.

Le nombre de cluster optimal peut être donc trouvé, en cherchant la valeur de K où un genou, un pic ou un creux existe dans le graphe de la mesure d'évaluation lorsqu'il est tracé par rapport au nombre de clusters [172].

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} \|y_i - c_i\|^2 \quad (1)$$

Où K est le nombre de clusters,  $y_i$  est la valeur de la ième observation dans  $C_i$

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2)$$

Où  $a_i$  est la distance moyenne de l'objet i aux tous les autres objets de son cluster,  $b_i$  est le minimum des distances de l'objet i aux éléments relevant d'un autre cluster que le sien. (Ces deux méthodes sont détaillées respectivement dans les travaux de [173,174].

#### **1.4.1.4. L'intégration du modèle LRFM dans les techniques de data mining**

L'intégration de l'analyse LRFM dans les techniques de data mining fournit des outils d'intelligence plus puissants pour les gestionnaires de marché qui les aident à prédire le comportement des clients et à faire de meilleures recommandations pour satisfaire les besoins des clients [175]. Avec le développement d'approches innovantes de data mining, les chercheurs considèrent l'intégration des variables LRFM dans les techniques de modélisation, comme dans le clustering [137,159,175], la régression logistique et les arbres de décision [176], Machines à vecteurs supports SVM [177,178], les règles d'association [179] et aussi une suite de multiples modèles de data mining [180], etc. Dans ce contexte, notre approche consiste d'une part, à combiner la méthode K-means et modèle LRFM pour diviser les clients d'un site e-commerce en groupes homogènes selon leurs valeurs L, R, F et M, et en d'autre part, ces variables seront utilisées avec d'autres mesures pour construire nos modèles de prédiction de churn. Les

enregistrements transactionnels pour chaque client doivent donc être transformés en un format utilisable pour le modèle LRFM.

La définition du modèle LRFM utilisé dans cette étude est présentée dans la table 4.5.

Les statistiques descriptives pour les variables (LRFM) dans T1 sont fournies dans la table 4.6.

Table 4.5 - Définition du modèle LRFM

| Nom d'attribut | Le contenu des données  |
|----------------|---|
| Longueur (L)   | Fait référence au nombre de jours entre le premier et le dernier achat                              |
| Récence (R)    | Indique le nombre de jours entre le premier jour de la période d'étude et le jour du dernier achat. |
| Fréquence (F)  | Fait référence au nombre de transactions observées au cours de la période analysée                  |
| Montant (M)    | Désigne le montant total dépensé par le client au cours de la période analysée. (MAD)               |

Table 4.6 - Statistiques descriptives de la longueur, la récence, la fréquence et le montant

| Variables     | Max       | Min   | Moyenne | Ecart type |
|---------------|-----------|-------|---------|------------|
| Longueur (L)  | 813       | 2     | 656.68  | 192.87     |
| Récence (R)   | 241       | 1     | 164.77  | 76.05      |
| Fréquence (F) | 17        | 1     | 8.67    | 4.99       |
| Montant (M)   | 13,723.00 | 87.00 | 4431.15 | 4327.72    |

Basé sur le travail de [168], nous utilisons les valeurs moyennes des quatre variables (L, R, F et M) de chaque cluster pour les comparer aux valeurs moyennes totales de tous les clusters. Si la valeur moyenne (L, R, F, M) d'un cluster est supérieure à la moyenne totale, une barre supérieure apparaît. Cependant, si la valeur moyenne (L, R, F, M) d'un cluster est inférieure à la moyenne totale, une barre inférieure apparaît. (Par exemple,  $\bar{R}$ : valeur de R est élevée, ce qui signifie que

le client a récemment fait un achat,  $\underline{R}$ : valeur de R est inférieure, le client n'a pas acheté sur le site en ligne depuis longtemps).

Sur la base de l'étude effectuée par S. H. Ha et S. C. Park [180], Chang et Tsay [168] ont proposé une classification de client en regroupant les seize combinaisons du modèle LRFM en cinq types de groupes de clients à savoir : (1) les clients principaux, (2) les clients potentiels, (3) les clients perdus, (4) les nouveaux clients et (5) les consommateurs de ressources. Plus précisément, les principaux clients comprennent les clients dont les LRFM patterns sont  $\overline{L}\overline{R}\overline{F}\overline{M}$ ,  $\overline{L}\overline{R}\overline{F}\underline{M}$ , ou  $\overline{L}\overline{R}\underline{F}\overline{M}$ . Les clients potentiels sont composés de  $\overline{L}\underline{R}\overline{F}\overline{M}$ ,  $\overline{L}\underline{R}\overline{F}\underline{M}$ , et  $\overline{L}\underline{R}\underline{F}\overline{M}$ . Les clients perdus :  $\underline{L}\underline{R}\overline{F}\overline{M}$ ,  $\underline{L}\underline{R}\overline{F}\underline{M}$ ,  $\underline{L}\underline{R}\underline{F}\overline{M}$ , ou  $\underline{L}\underline{R}\underline{F}\underline{M}$ . Les nouveaux clients :  $\underline{L}\overline{R}\underline{F}\underline{M}$ ,  $\underline{L}\overline{R}\overline{F}\underline{M}$ ,  $\underline{L}\overline{R}\underline{F}\overline{M}$ , ou  $\underline{L}\overline{R}\overline{F}\overline{M}$ . Enfin, les clients consommateurs de ressources sont ceux qui appartiennent à l'un de ces deux modèles  $\overline{L}\overline{R}\underline{F}\underline{M}$  ou  $\overline{L}\underline{R}\underline{F}\underline{M}$ . Dans notre cas, lorsque les différentes combinaisons de LRFM sont identifiées pendant la période t, nous concentrons premièrement sur celles qui représentent les principaux et les nouveaux clients (aucune entreprise ne veut rater l'opportunité de fidéliser les nouveaux clients), deuxièmement, nous prenons en compte les clients appartenant aux autres combinaisons restantes à la période t, et qui sont convertis à la période t+1 en clients principaux. Plus précisément, les clients qui nous intéressent sont ceux dont le modèle LRFM appartient aux modèles suivants :

(i)-  $\overline{L}\overline{R}\overline{F}\overline{M}$ ,  $\overline{L}\overline{R}\overline{F}\underline{M}$ ,  $\overline{L}\overline{R}\underline{F}\overline{M}$ ,  $\underline{L}\overline{R}\overline{F}\overline{M}$ ,  $\underline{L}\overline{R}\overline{F}\underline{M}$  et  $\underline{L}\overline{R}\underline{F}\overline{M}$  pendant la période t.

(ii)- Les clients qui n'appartiennent pas aux modèles listés en (i) au court de la période t, et qu'ils ont transformé dans période t+1 à l'un des modèles mentionnés en (i). Cette transformation est principalement due au changement positif de leur comportement d'achat.

D'une autre part, les clients qui ont été regroupés au cours de la période t dans les segments qui ne contribuent pas aux bénéfices (les clients perdus, potentiels, et les consommateurs de ressources), et qui sont restés au court la période t+1 dans le même segment ou, ils sont transformés en un segment de valeur inférieure (ex, un client potentiel vers un client perdu) seront ignorés.

#### 1.4.1.5. La défection partielle et totale dans les cadres non contractuels

Parmi les premiers obstacles majeurs auxquels font face les modèles de prédiction de churn dans les cadres non-contractuels, il y a la difficulté de définir le phénomène de churn, car les



caractéristiques qui doivent être observées pour dire qu'un client a totalement ou partiellement défectueux ne sont pas clairement définies [131]. Pour résoudre ce problème, le modèle LRFM et la méthode de clustering k-means sont combinés. Cette étude, propose une nouvelle procédure, en joignant les valeurs quantitatives des variables L, R, F et M pendant une période T dans l'algorithme k-means afin d'identifier les différents types de profils de clients (les différents modèles LRFM). Nous définissons ensuite le changement au niveau du modèle LRFM d'un client principal ( $\bar{L} \bar{R} \bar{F} \bar{M}$ ,  $\bar{L} \bar{R} \bar{F} \underline{M}$ ,  $\bar{L} \bar{R} \underline{F} \bar{M}$ ) ou d'un nouveau client ( $\underline{L} \bar{R} \bar{F} \bar{M}$ ,  $\underline{L} \bar{R} \bar{F} \underline{M}$ ,  $\underline{L} \bar{R} \underline{F} \bar{M}$ ), au client potentiel ( $\bar{L} \underline{R} \bar{F} \bar{M}$ ,  $\bar{L} \underline{R} \bar{F} \underline{M}$ ,  $\bar{L} \underline{R} \underline{F} \bar{M}$ ), ou au client à faible coût de consommation de ressources ( $\bar{L} \underline{R} \underline{F} \underline{M}$ ) comme une défection partielle. De même principe, si un client change son modèle LRFM à partir de l'un de deux types de client suivants : les clients principaux ( $\bar{L} \bar{R} \bar{F} \bar{M}$ ,  $\bar{L} \bar{R} \bar{F} \underline{M}$ ,  $\bar{L} \bar{R} \underline{F} \bar{M}$ ) ou les nouveaux clients ( $\underline{L} \bar{R} \bar{F} \bar{M}$ ,  $\underline{L} \bar{R} \bar{F} \underline{M}$ ,  $\underline{L} \bar{R} \underline{F} \bar{M}$ ) vers les clients perdus ( $\underline{L} \underline{R} \bar{F} \bar{M}$ ,  $\underline{L} \underline{R} \bar{F} \underline{M}$ ,  $\underline{L} \underline{R} \underline{F} \bar{M}$ ,  $\underline{L} \underline{R} \underline{F} \underline{M}$ ), ou vers les clients à forte consommation de ressources ( $\bar{L} \underline{R} \underline{F} \underline{M}$ ), dans ce cas nous parlons d'une défection totale. Cela indiquerait que le changement négatif du modèle LRFM au fil du temps est un signal précoce d'une défection partielle ou totale. Tandis que les clients qui restent fidèles à leurs modèles existants sont susceptibles de rester. À cette fin, comme la montre la figure 4.4, nous considérons deux sous-périodes égales T1 et T2. T1 est utilisée pour identifier les différents groupes de clients (les différents modèles LRFM) et d'affecter chaque client à son groupe le plus approprié, alors que T2 est utilisée pour prédire la défection partielle et totale. La figure 4.5 illustre notre approche proposée pour définir la défection partielle et totale, et le processus complet est résumé dans la figure 4.6.

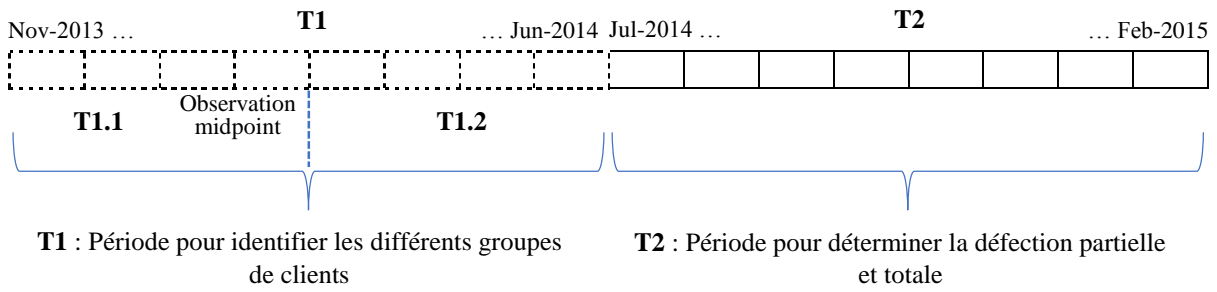


Figure 4.4- Période d'observation.

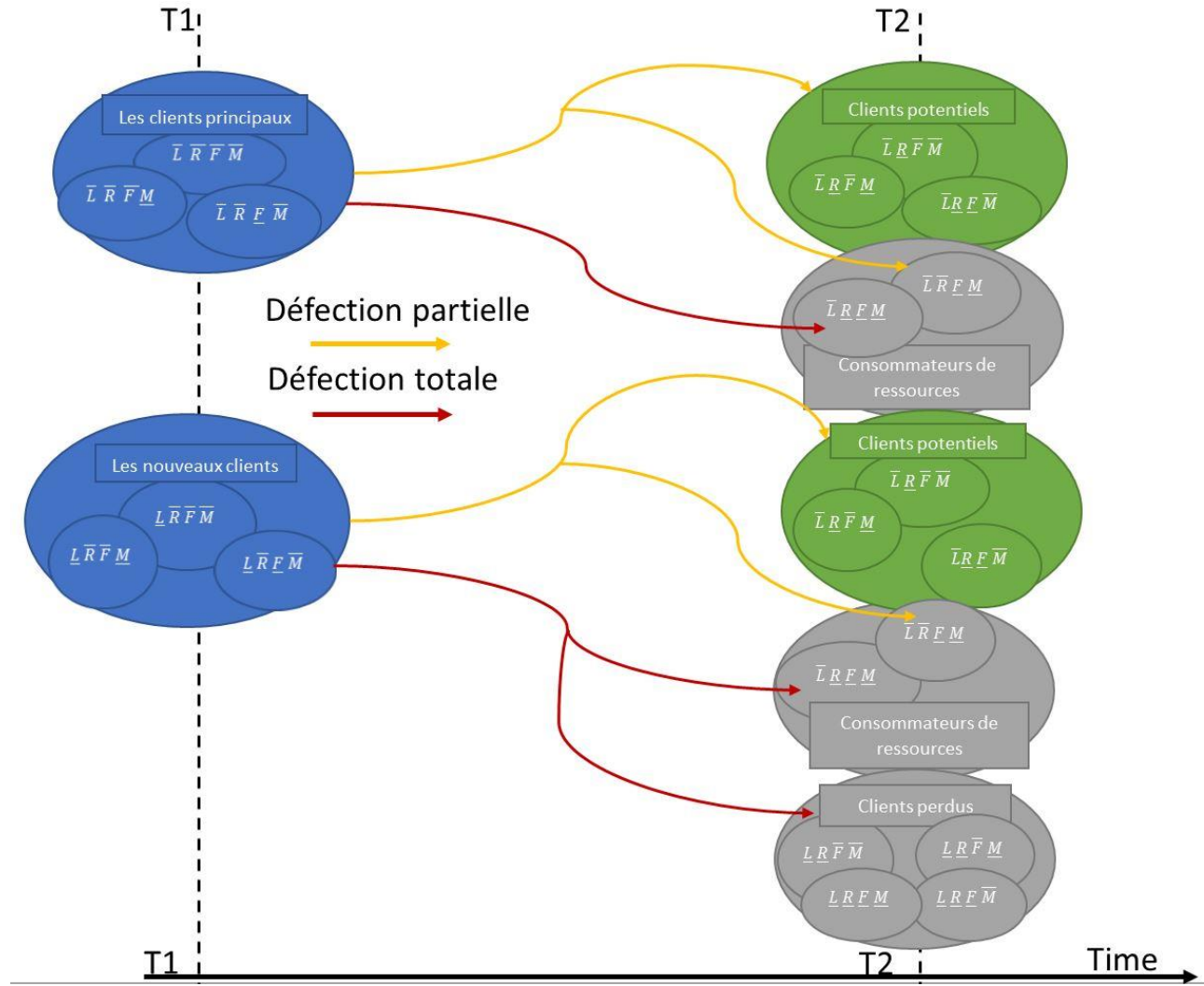


Figure 4.5 - Changement du modèle LRFM des clients au fil du temps pour définir le churn (perte de clientèle).

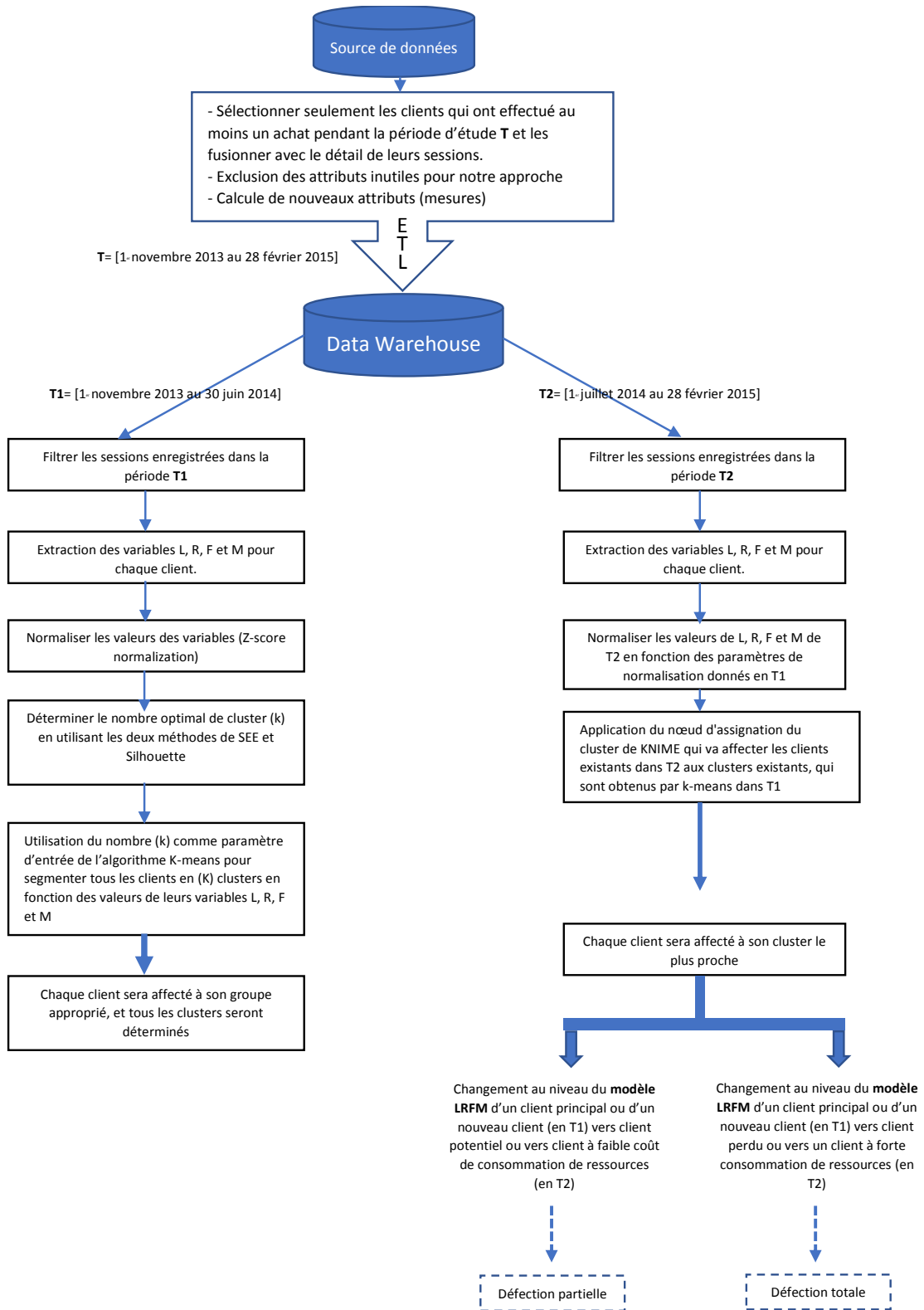


Figure 4.6 - La méthodologie proposée basée sur le modèle LRFM et la technique K-means pour définir et identifier le churn dans les contextes non-contractuels.

#### 1.4.1.6. K-means et le modèle LRFM pour la segmentation des clients

La première période de huit mois, de novembre 2013 à juin 2014 (T1), est utilisée pour identifier les différents groupes de clients (les différents modèles LRFM) (voir Figure 4.4). Par conséquent, 2691 clients ont visité le site Web de commerce électronique durant cette période.

Selon la méthodologie proposée décrite dans la figure 4.6, nous avons établi deux workflow avec la plateforme KNIME Analytics 3.4, le premier sert à déterminer le nombre optimal de cluster (k), et le deuxième sert à segmenter l'ensemble de clients en K groupes et d'identifier les différents modèles de LRFM. Par conséquent, nous trouvons sept groupes de clients qui ont un comportement LRFM différent. Le nombre optimal de cluster (k=7) est obtenu à partir des méthodes « Elbow » et « Silhouette ». La figure 4.7 montre les courbes de l'indice SSE et le coefficient de silhouette moyenne en fonction du nombre de clusters de k-means. Un genou distinct dans le SSE et un pic remarquable dans le coefficient de silhouette sont présents lorsque le nombre de clusters est égal à 7.

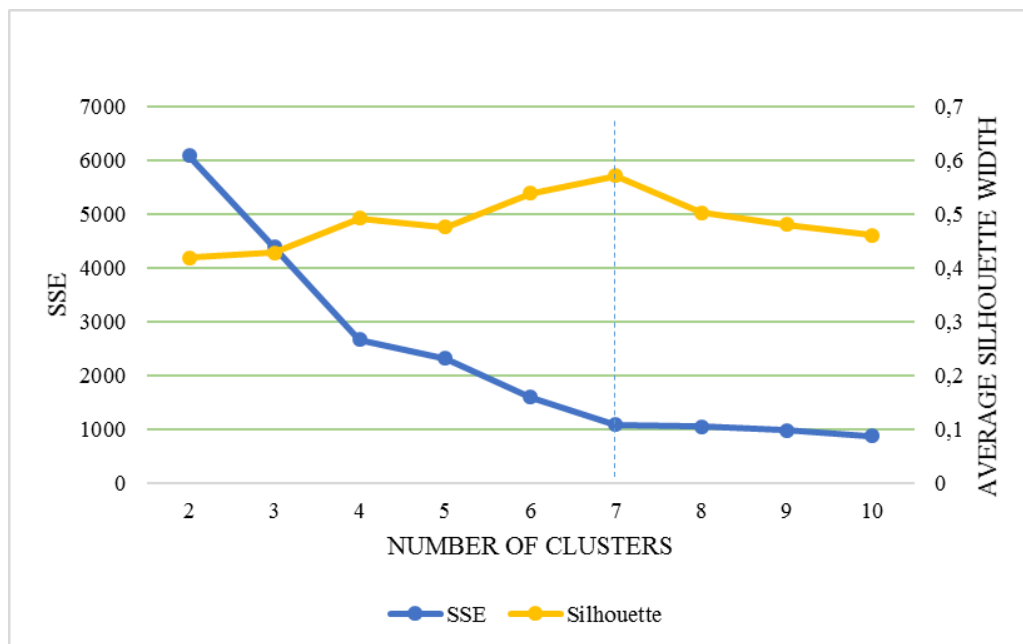


Figure 4.7 - Le résultat des méthodes « Elbow » et « Silhouette » pour déterminer le nombre optimal de clusters

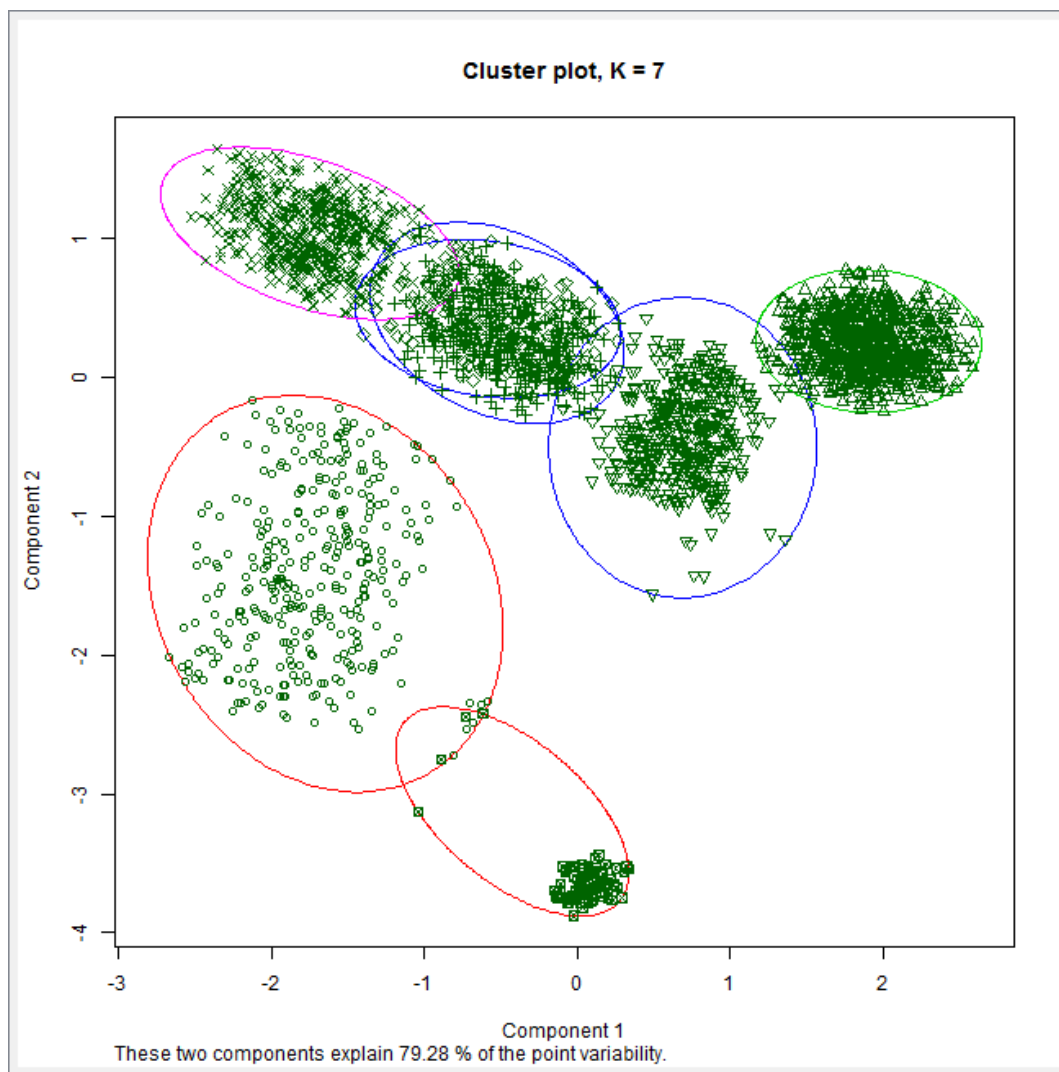
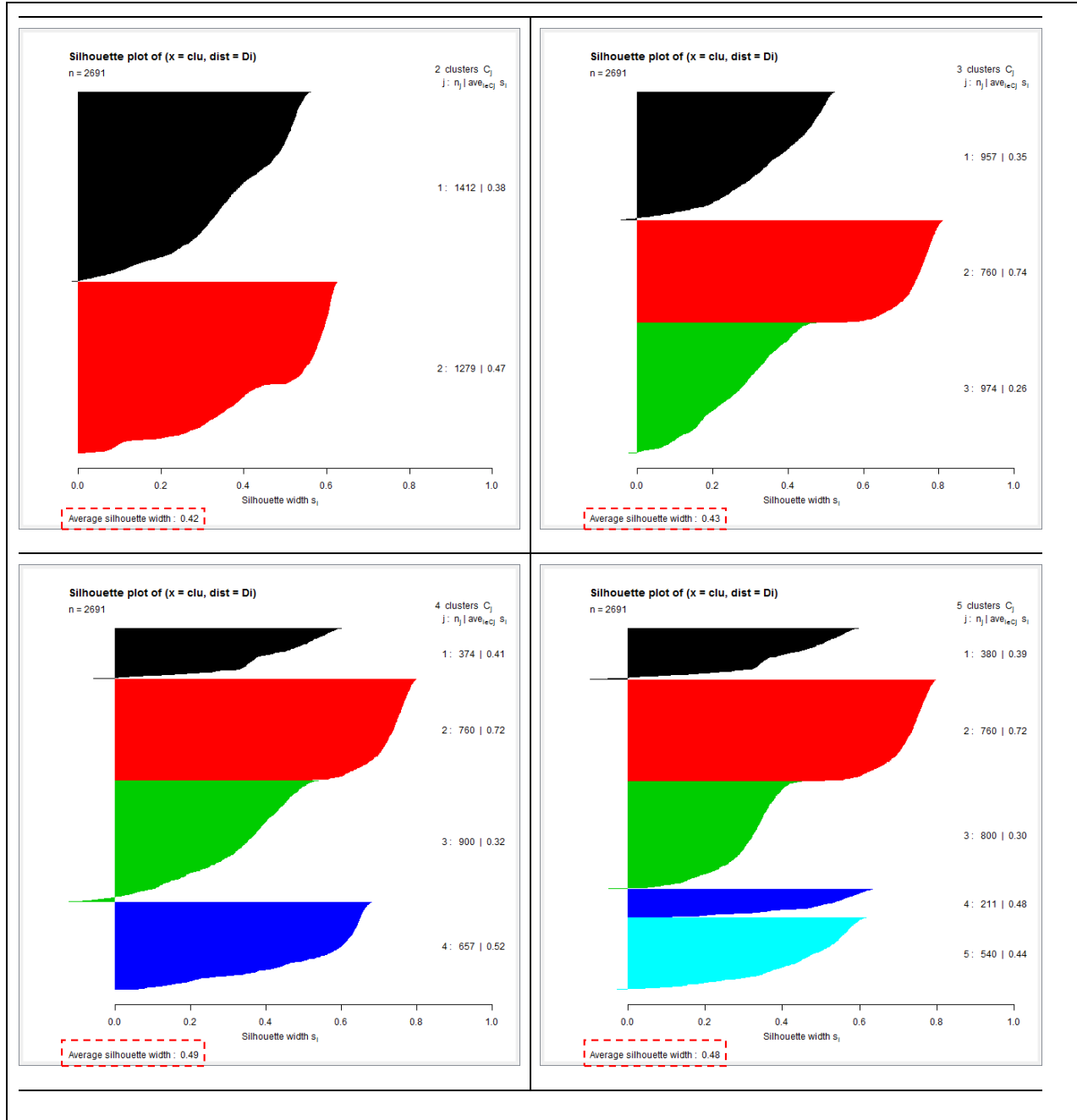
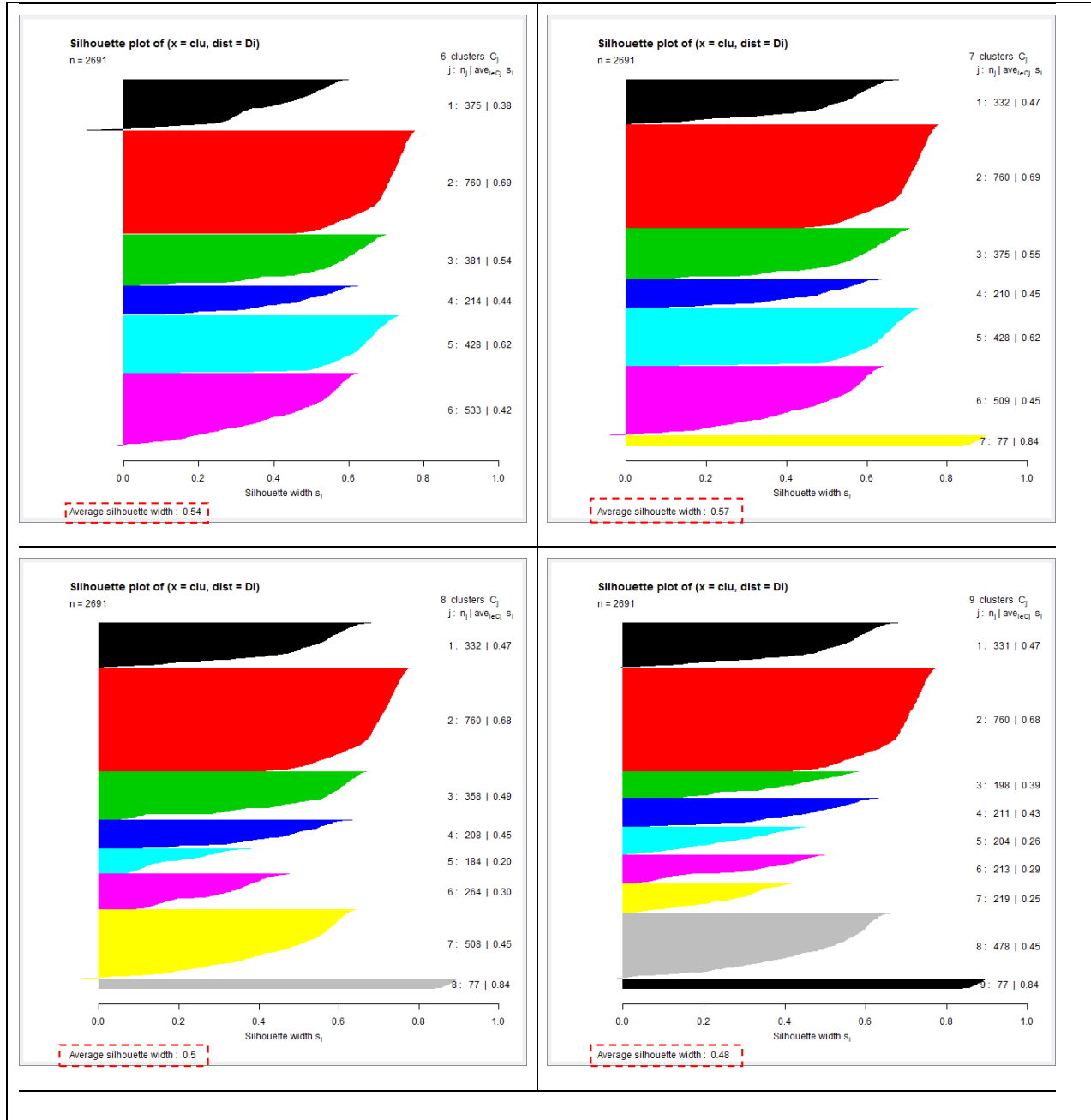


Figure 4.8 – Présentation des 7 clusters trouvés par la méthode K-means (k=7).







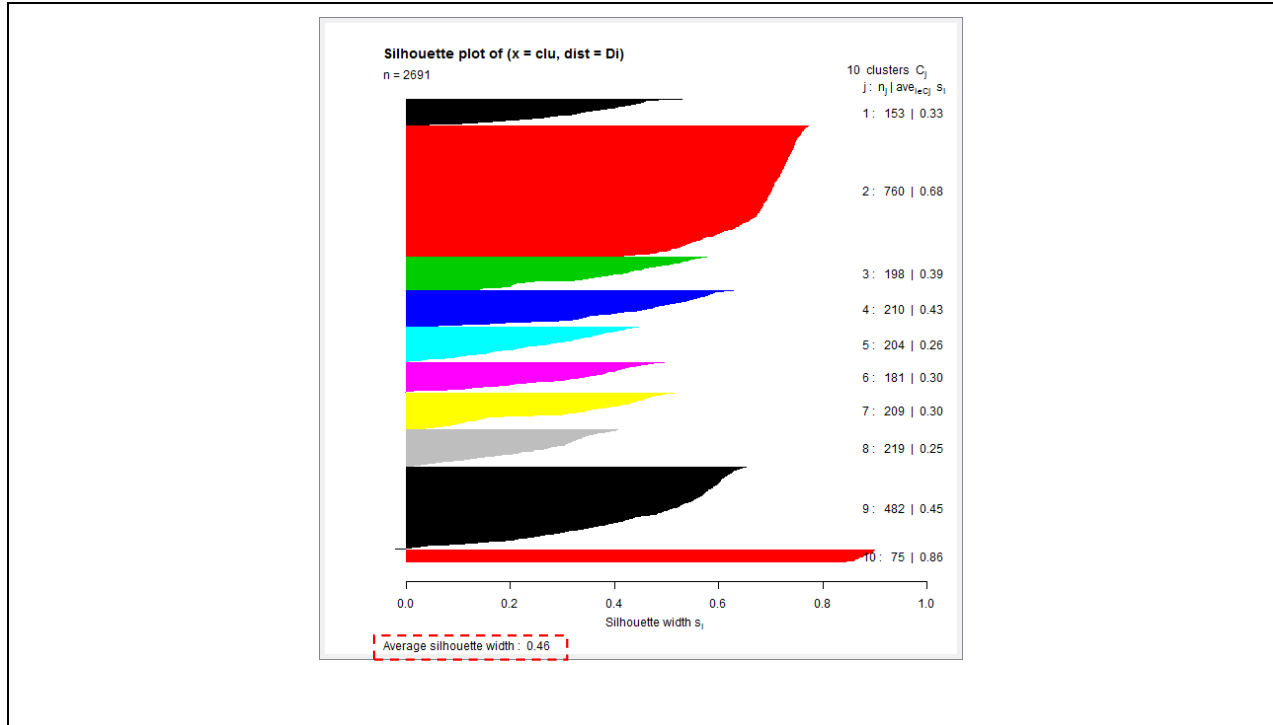


Figure 4.9 - Graphique de silhouette pour k allant de 2 à 10.

La table 4.7 est un résumé du résultat de regroupement de ces sept clusters, chacun avec le nombre correspondant de clients, longueur moyenne (L), récence moyenne (R), fréquence moyenne (F), moyenne monétaire (M) et la dernière colonne montre le modèle LRFM pour chaque cluster. La plupart des clients se trouvent dans les clusters 1, 3 et 5. Alors que le groupe 6 comprend le nombre minimum de clients (seulement 77 clients).

Table 4.7 - Statistiques descriptives de sept clusters obtenus par la méthode k-means en T1

| Cluster   | Count | Mean(L) | Mean(R) | Mean(F) | Mean(M)  | Pattern  |
|-----------|-------|---------|---------|---------|----------|--|
| Cluster_0 | 332   | 282,31  | 211,26  | 12,41   | 8204,28  | $\underline{L} \overline{R} \overline{F} \overline{M}$   |
| Cluster_1 | 760   | 752,87  | 48,31   | 4,47    | 857,62   | $\overline{L} \underline{R} \underline{F} \underline{M}$ |
| Cluster_2 | 375   | 707,71  | 206,59  | 13,34   | 2187,08  | $\overline{L} \overline{R} \overline{F} \underline{M}$   |
| Cluster_3 | 509   | 742,81  | 210,91  | 14,65   | 10817,28 | $\overline{L} \overline{R} \overline{F} \overline{M}$    |
| Cluster_4 | 210   | 741,10  | 209,95  | 6,97    | 8266,72  | $\overline{L} \overline{R} \underline{F} \overline{M}$   |
| Cluster_5 | 428   | 699,51  | 212,89  | 3,93    | 1063,69  | $\overline{L} \overline{R} \underline{F} \underline{M}$  |

|           |    |       |        |      |        |  |
|-----------|----|-------|--------|------|--------|--|
| Cluster_6 | 77 | 35,52 | 214,66 | 2,91 | 405,12 | $\underline{L} \overline{R} \underline{F} \underline{M}$ |
|-----------|----|-------|--------|------|--------|--|

À partir de la table 4.7, les clients de Cluster\_0 ont une valeur faible de L ( $\underline{L} \overline{R} \overline{F} \overline{M}$ ). La baisse valeur de L indique que ces clients n'ont pas encore établi une longue relation avec le site marchand de l'entreprise (des nouveaux clients). En outre, malgré la valeur inférieure de L, on observe que la valeur moyenne de R, F et M est supérieure à la moyenne globale, ce qui pourrait indiquer que ces clients achètent récemment et fréquemment avec un montant dépensé élevé. Ces clients sont considérés comme les nouveaux clients de haute valeur. En plus de Cluster\_0, les clients de Cluster\_6 ( $\underline{L} \overline{R} \underline{F} \underline{M}$ ) sont de nouveaux clients incertains, ils ont la plus grande valeur de R par rapport aux autres clusters.

Le Cluster\_1 est caractérisé par une valeur plus élevée de L, et des valeurs R, F et M faibles par rapport à la valeur moyenne globale de ses variables ( $\overline{L} \underline{R} \underline{F} \underline{M}$ ). Les clients dans ce cluster appartiennent aux anciens clients, qui montrent peu d'intérêt aux articles et services proposés par le site marchand en ligne. Ce manque d'attention est déterminé par le faible nombre de transactions effectuées par ces clients, la faible valeur de récence et leur faible contribution aux bénéfices. Ils ont perdu le contact avec le site marchand de l'entreprise, et ceci est dû au fait qu'ils n'ont pas effectué des achats sur le site depuis longtemps. Peut-être que ces clients n'étaient pas satisfaits des services et des produits qu'ils ont reçus, ou qu'ils ont été attirés par les concurrents, ou qu'ils ont perdu confiance envers le site marchand. Ce groupe de clients s'appelle les clients défectueux.

En plus du cluster 1, les clients du cluster 5 ( $\overline{L} \overline{R} \underline{F} \underline{M}$ ) sont en voie de défection (défection partielle), car le nombre de transactions effectuées par ces clients et leur contribution à l'entreprise restent faibles même s'ils ont une relation étroite avec le site e-commerce ; la seule différence entre le cluster\_1 et cluster\_5 est que ce dernier groupe a récemment fait des achats sur le site e-commerce.

Contrairement aux clients défectueux, les clients dans les clusters 2, 3 et 4 sont les principaux clients du site marchand en ligne. En observant directement les symboles de L, R, F et M dans la table 4.7, les clients font partie du cluster 3 sont considérés comme des clients fidèles de grande valeur, car ils ont la meilleure valeur moyenne d'argent dépensé, la très grande fréquence d'achat, une grande valeur de récence, et, plus important, ils ont établi une relation longue et étroite avec

le site e-commerce. Sans aucun doute, ces 509 clients sont les plus rentables et les plus fidèles à leur site marchand en ligne à long terme.

Les clients de Cluster\_2 ( $\bar{L} \bar{R} \bar{F} \bar{M}$ ) sont considérés comme des acheteurs à haute fréquence, et en voie de migrer vers le Cluster\_3.

Pour le Cluster\_4, à l'exception de la baisse valeur moyenne de F, les valeurs moyennes de L, R et M sont très élevées. Les clients de ce cluster sont considérés comme des dépensiers, car même s'ils ont effectué peu de transactions, ils ont réussi à dépenser beaucoup d'argent, ce qui signifie que ce type de clients a l'habitude d'acheter des articles coûteux.

Comme il est mentionné précédemment, nous nous concentrons initialement sur les clients appartenant aux principaux clients (Cluster 2, 3 et 4), et ceux qui font partie des nouveaux clients de haute valeur (Cluster\_0). Il faut noter que ces deux types de clients représentent 51,23% de la base de données clients disponibles. Deuxièmement, nous prenons en compte ceux appartenant aux autres clusters restants, et qui sont convertis en clients principaux au cours de la deuxième période de huit mois (T2).

Dans la deuxième période T2 (de juillet 2014 à février 2015), nous introduisons le nœud « *Cluster Assigner* » (qui affecte les clients qui ont effectué des transactions dans T2, aux clusters existants obtenus par k-means dans T1) pour déterminer le client qui a changé leur modèle LRFM du client principal en T1 vers un client défectueux dans le période T2. L'application de notre définition de défection partielle et totale décrite dans la section 1.4.1.5 donne 254 défections partielles (17,81% = 254/1426) et 363 défections totales (25,45% = 363/1426), où 1426 représente le nombre de clients sous enquête (Cluster\_2 + Cluster\_3 + Cluster\_4 + Cluster\_0 = 1426).

Enfin, dans la dernière étape du processus de CRISP-DM, nous avons utilisé l'outil décisionnel Power BI de Microsoft pour la génération des rapports interactifs.

Une analyse plus détaillée sur le résultat du modèle de segmentation client est présentée dans la figure 4.10 sous forme de rapport qui contient plusieurs visualisations, et qui utilise les dimensions suivantes: dim\_client, dim\_transaction, dim\_time..

# Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision

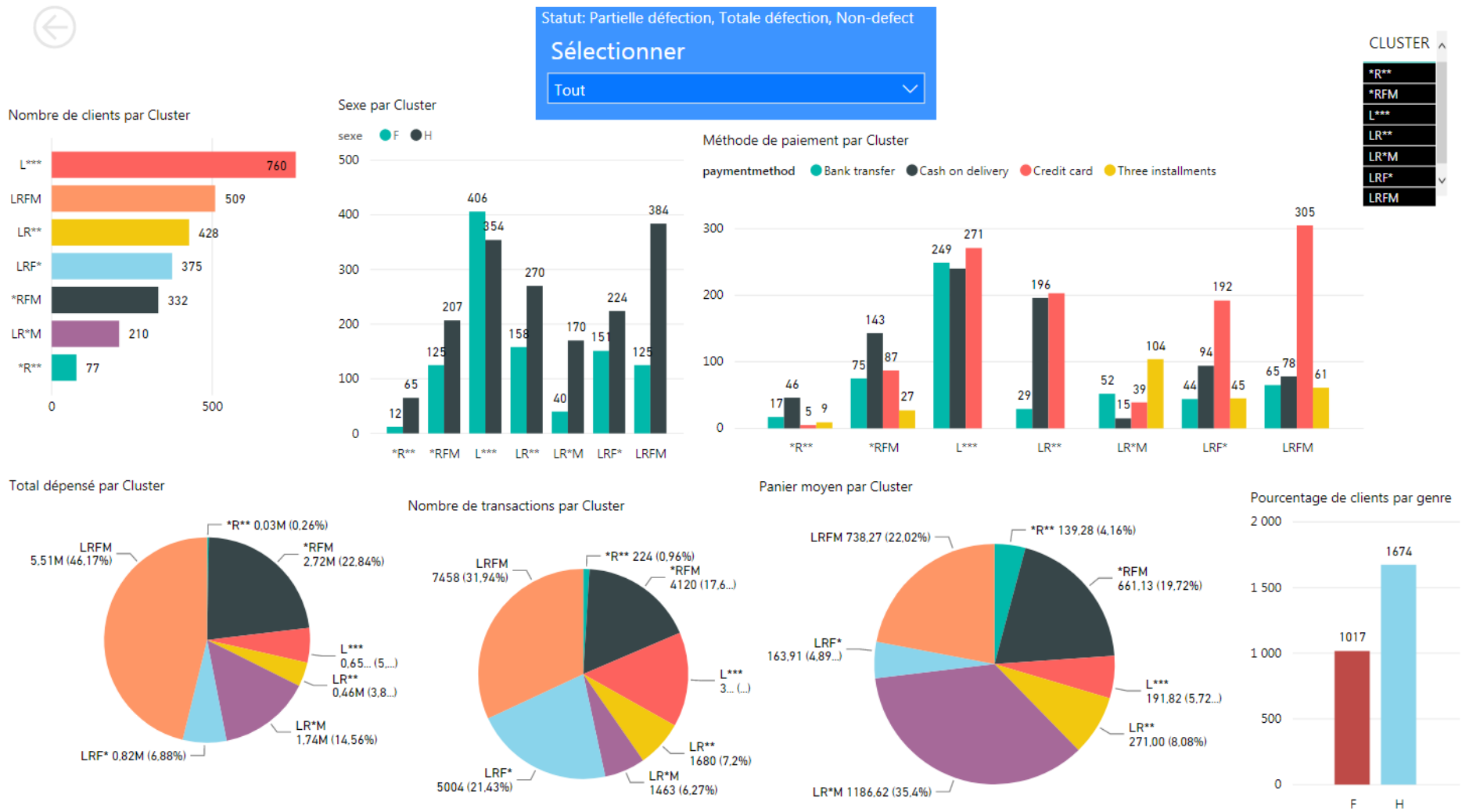


Figure 4.10 : Vue d'ensemble sur les sept clusters.

La figure 4.10 est un rapport qui contient plusieurs visualisations qui dynamisent les données de sept clusters générés par l'algorithme K-means, et donne une représentation visuelle qui permet aux managers d'avoir une compréhension intuitive des différents segments de clients. Par exemple, à partir des visualisations présentées dans cette figure, nous voyons que, à l'exception des clients appartiennent au cluster L\*\*\*, le nombre de clients masculins est toujours supérieur à celui des femmes. La majorité des clients appartiennent aux clusters \*RFM et \*R\*\* préfèrent le paiement en espèces à la livraison pour payer leurs achats. Le paiement par carte de crédit est le mode le plus utilisé par les clusters LRFM et LRF\*. Enfin, le paiement en trois tranches reste le mode privilégié par les clients du cluster LR\*M (les dépensiers).

Nous pouvons aussi conclure que, plus la relation entre le site marchand et les clients (L) est longue; plus les clients font confiance au système de paiement électronique proposé par le magasin en ligne. Ceci s'applique aux clusters LRFM, LRF\*, LR\*\* et L\*\*\* dans lesquels les clients utilisent souvent la carte bancaire comme moyen de paiement. De plus, les clusters qui ont une faible valeur de L, comme les clusters \*R\*\* et \*RFM préfèrent payer leurs achats avec le mode « Cash à la livraison ». Peut-être que les clients qui ont récemment rejoint le site marchand n'ont pas encore fait confiance au systèmes de paiement électroniques de l'entreprise. La méthode de paiement représente donc, l'un des indicateurs clés d'un nouvel indice de confiance qui permet d'évaluer le niveau de confiance des clients envers le site e-commerce de l'entreprise. L'entreprise en ligne doit donc encourager les clients à payer leurs achats en ligne en utilisant des cartes marocaines ou étrangères afin de créer un climat de confiance avec ces clients, car si l'entreprise parvient à établir un lien de confiance avec ses clients, cela l'aidera à promouvoir un sentiment de satisfaction et à encourager une relation à long terme.

En outre, pour chaque cluster, le montant total dépensé, le nombre total de transactions, et la valeur du panier moyen sont affichés dans le rapport. Ces trois valeurs sont représentées dans trois graphiques en secteurs respectivement dans la partie basse du rapport.

En termes de montant total dépensé, clusters LRFM et \*RFM représentent les clusters qui génèrent les grandes parties des revenus 46.17% et 22.84% respectivement. En ce qui concerne le nombre de transactions, les clusters LRFM, LRF\*, \*RFM représentent les clusters qui ont enregistré les plus grands nombres de transactions. Alors que pour le panier moyen, le cluster

LR\*M a la valeur la plus élevée 1186.62 dh (35.4%), car d'après son modèle LRFM, il est clair que les clients de ce cluster sont des dépensiers.

La liste déroulante "Statut" située à l'entête du rapport, peut être utilisée pour filtrer les clients. Si l'utilisateur sélectionne "Partially churned", alors pour chaque cluster, seuls les clients définiés comme "des clients partiellement défectueux" par notre approche de défintion d'attrition seront comptés (voir figure 4.11).

La figure 4.12 montre comment les visualisations appartenant d'un même rapport peuvent être filtrées, exploiter d'autres visualisations et interagir entre elles. Par exemple, pour visualiser les résultats concernant seulement le cluster LRFM, il suffit de cliquer dans la légende CLUSTER sur le pattern LRFM. Par conséquent, les résultats de ce cluster sont mis en surbrillance dans le rapport et le reste des résultats sera désactivé.

Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision

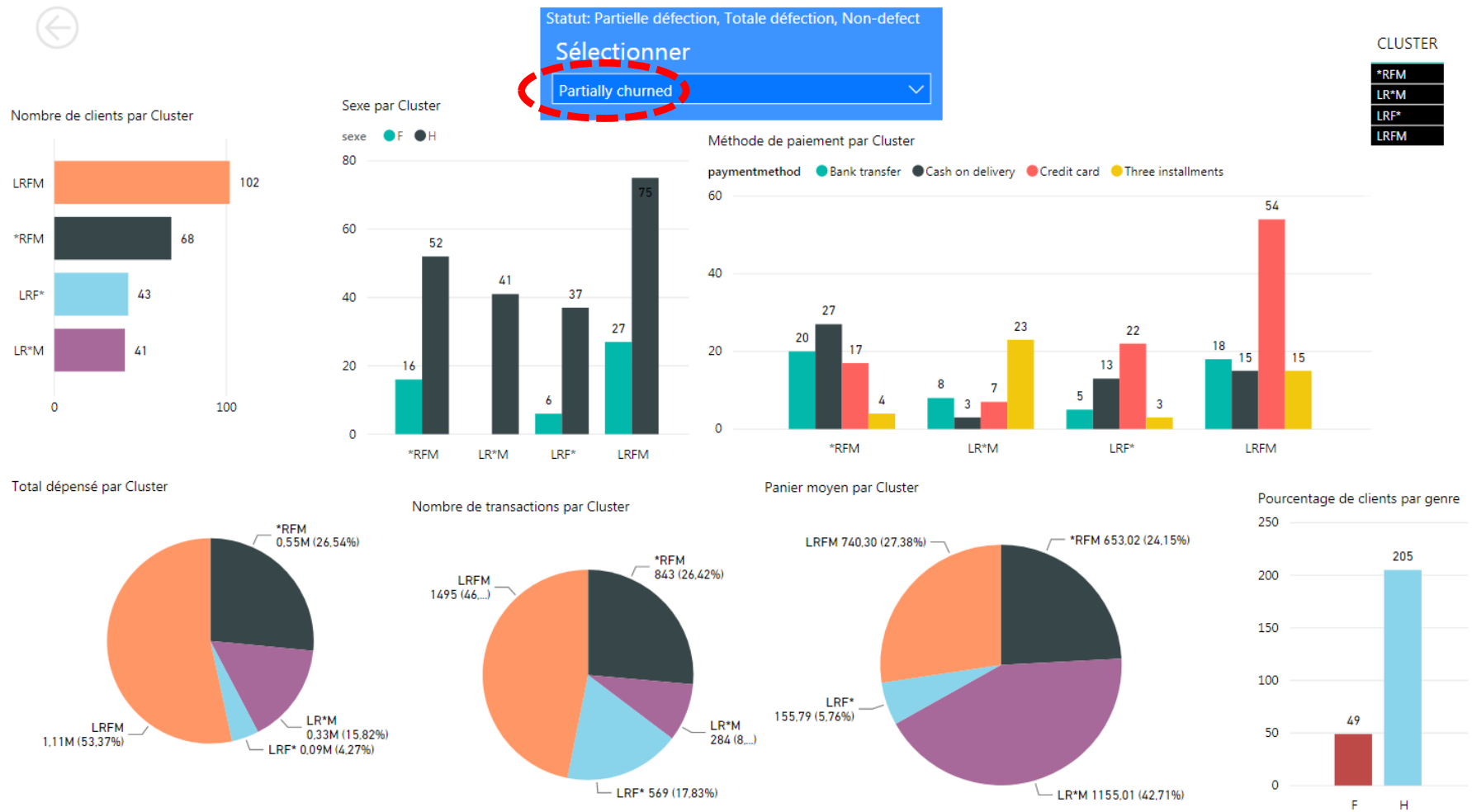


Figure 4.11 : Exemple de filtrage, vue d'ensemble sur les clients définis comme des clients partiellement défectueux.

# Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision

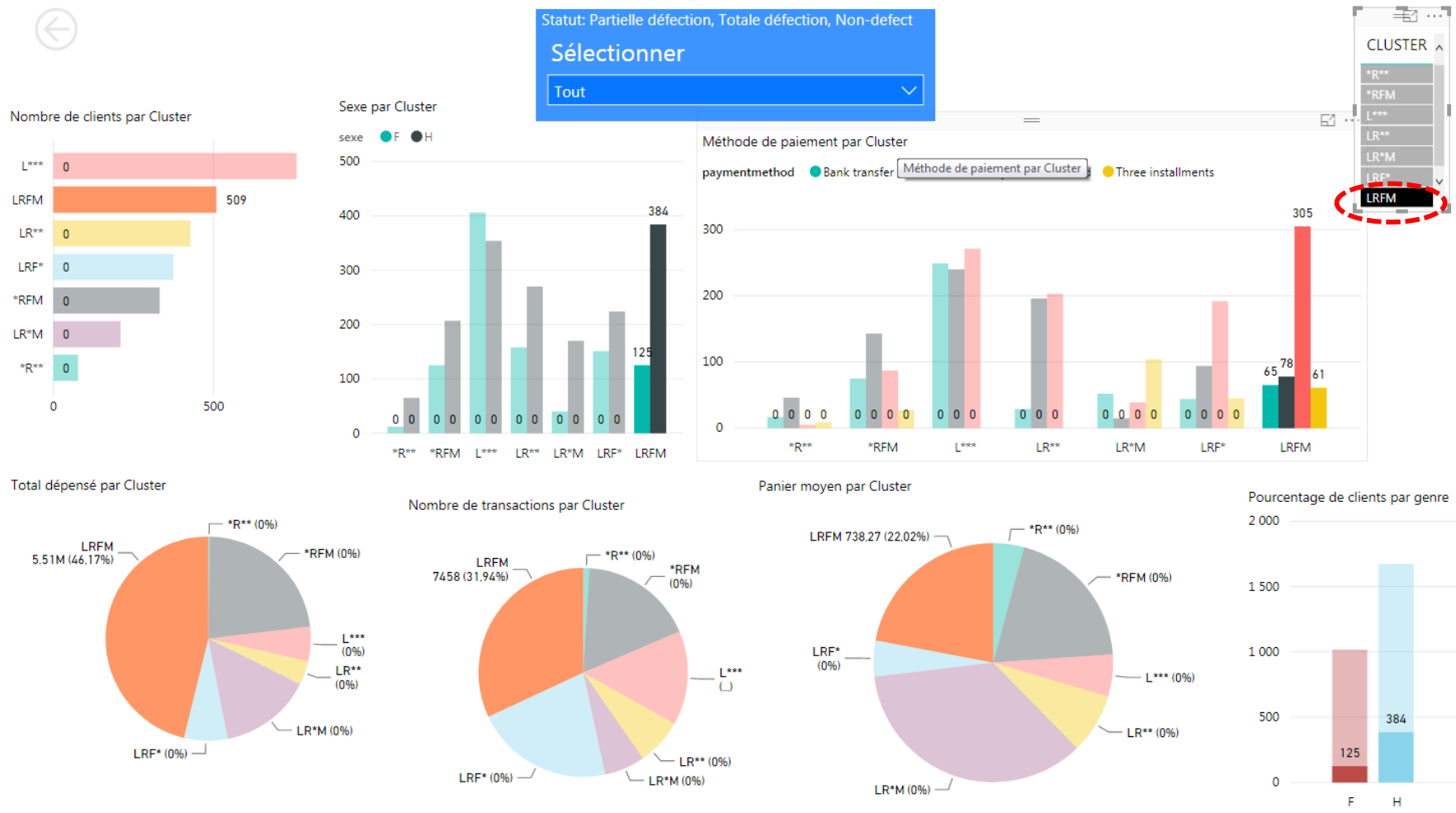


Figure 4.12 : Vue d'ensemble sur le cluster LRFM.



## **1.4.2. Modèle de prédiction :**

### **1.4.2.1. Les techniques de classification**

L'objectif de notre travail consiste à développer un modèle prédictif d'attrition dans un cadre non contractuel, qui permettrait de distinguer entre les clients susceptibles d'être partiellement ou totalement défectueux dans un futur proche, et ceux qui susceptibles de rester fidèles. Notre modèle sera basé sur les caractéristiques des clients et sur leurs historiques de transactions et de navigations sur le site e-commerce. Pour atteindre cet objectif, trois modèles sont proposés, le premier est basé sur des techniques d'arbre de décision (DT), le second sur des réseaux de neurones artificiels (ANN) et le troisième sur un ensemble d'arbres de décision. Nous notons que tous nos modèles sont construits en utilisant la Plateforme KNIME Analytics 3.4. Les parties suivantes représentent une courte description de ces techniques de data mining utilisées pour la tâche de prédiction.

#### **1.4.2.1.1. Les réseaux de neurones artificiels**

Contrairement aux méthodes statistiques conventionnelles, les réseaux neuronaux artificiels (Artificial Neural Networks, ANN) n'ont pas besoin d'hypothèses sur les variables, ils sont bien adaptés pour traiter des problèmes complexes non structurés, c'est-à-dire des problèmes sur lesquels il n'est pas possible de spécifier a priori la forme des relations entre variables.

Les réseaux neuronaux peuvent être distingués en perceptron monocouche et en perceptron multicouche (MLP) dont chacun a des avantages et des inconvénients. Dans notre étude, nous utilisons la structure MLP qui connue par sa puissance de réaliser diverses applications. Ce choix est fait pour la rapidité et la facilité de la construction de ce genre de structure de réseaux de neurones et encore par le fait que notre problème présente un nombre limité de variables d'entrées. Un réseau MLP est généralement composé d'un ensemble fini de cellules (neurones), organisées en couches successives. La première couche comprenant plusieurs neurones est appelée couche d'entrée, la dernière couche est la couche de sortie et les couches intermédiaires sont les couches cachées. Les neurones dans différentes couches sont connectés par des fonctions de transfert (sigmoïde, linéaire, seuil, tangente hyperbolique, ...) qui sont utilisées comme fonctions d'activation dans la perception multicouche. Un simple MLP avec une couche masquée est présenté à la figure 4.13. Chaque flèche représente une connexion entre deux neurones et

indique la voie de circulation de l'information. Chaque connexion a un poids, un nombre entier qui contrôle le signal entre les deux neurones. Si le réseau génère une sortie "bonne ou souhaitée", il n'est pas nécessaire d'ajuster les poids. Cependant, si le réseau génère une sortie « médiocre ou non souhaitée » ou une erreur, le système modifie les poids afin d'améliorer les résultats attendus.

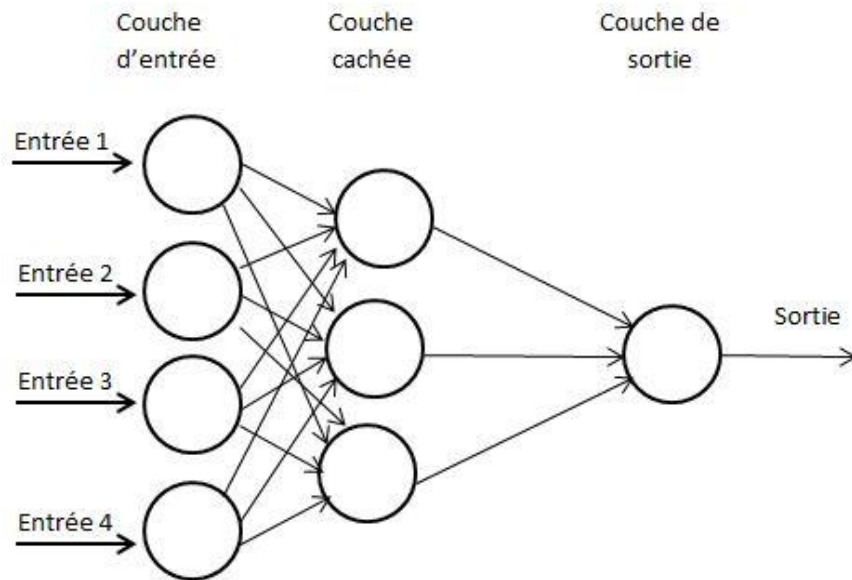


Figure 4.13 - Exemple d'architecture d'un MLP à 5 variables d'entrée, 3 neurones pour la couche cachée et un neurone pour la couche de sorties.

#### 1.4.2.1.2. Les arbres de décision

L'arbre de décision (Decision Tree, DT) est l'une des techniques de data mining les plus utilisées pour la découverte des connaissances, et il est généralement utilisé à des fins de classification et de prédiction [181]. La simplicité et la facilité d'interprétation des résultats générés par cette technique sont les principales raisons de sa popularité auprès des décideurs dans les milieux d'affaires, par rapport aux autres techniques de prédiction [182]. Il est considéré comme un outil d'aide à la décision représentant plusieurs choix sous la forme graphique d'un arbre. Le développement d'un arbre de décision comprend généralement deux étapes distinctes : la construction d'arbres et l'élagage des arbres. Au début, les techniques commencent à rechercher dans l'ensemble d'apprentissage un attribut offrant le meilleur gain d'information au niveau du nœud racine, puis divisant l'arbre en sous-arbres. La même procédure est utilisée pour partitionner récursivement le sous-arbre suivant la même règle, puis le partitionnement s'arrête

lorsque les extrémités des branches « les feuilles de l'arbre » sont atteintes. Une fois l'arbre créé, les règles peuvent être extraites en traversant l'arbre jusqu'à ce qu'un nœud feuille soit atteint. Plusieurs algorithmes tels que ID3, C4.5, C5.0, CHAID et CART sont utilisés pour produire les arbres. Dans cette étude, nous utilisons l'algorithme C4.5 en raison de sa capacité de gérer les deux types d'attributs continus et discrets, les valeurs manquantes et les données bruitées.

#### 1.4.2.1.3. Ensemble d'arbres de decision

Malgré les avantages de la méthode d'arbre de décision mentionnée ci-dessus, elle souffre aussi d'un problème lié au biais et à la variance [183]. Un grand biais avec des arbres simples et une grande variance avec des arbres complexes. Par conséquent, les méthodes d'ensemble (Decision Tree Ensemble, DTE), qui consistent à combiner plusieurs arbres de décision pour produire de meilleures performances prédictives que d'utiliser un seul, sont devenues parmi les meilleures et les plus récentes solutions à ce type de problème. Le principe général (voir [184]) est de construire une collection d'arbres de décision (prédicteurs), pour ensuite agréger l'ensemble de leurs résultats (prédictions). En classification, l'agrégation de résultats obtenus par N prédicteurs consiste par exemple à faire un vote majoritaire parmi les labels des classes fournis par ces N prédicteurs [185]. En régression, agréger les résultats de N prédicteurs revient par exemple à en faire la moyenne : chaque prédicteur fournit une prédiction  $P_l$ , et la prédiction finale est alors  $\frac{1}{N} \sum_{l=1}^N P_l$ . (3)

À cet égard, nous utilisons les deux nœuds *Tree Ensemble Learner* et *Tree Ensemble Predictor* de la plateforme *Knime* pour construire notre troisième modèle qui sera basé sur un ensemble d'arbres de décision. Au premier lieu, nous allons utiliser le nœud *Tree Ensemble Learner* pour construire un ensemble d'arbres de décision, comme une variante des techniques de forêts aléatoires (*Random forest*). Chacun de ces modèles d'arbre de décision sera formé sur un sous-ensemble différent de lignes et/ou colonnes, sélectionné au hasard à chaque itération. Le modèle de sortie est alors un ensemble de modèles d'arbres de décision formés différemment. Les paramètres d'apprentissage des arbres de décision sont similaires à ceux du *Random Forest* décrit par [185]. Par la suite nous allons introduire le nœud « *Tree Ensemble Predictor* » qui permet d'appliquer l'ensemble des arbres de décision (construit par le premier nœud) à chaque enregistrement de données et d'utiliser finalement le vote majoritaire pour la prédiction finale.

### **1.4.2.2. Les variables utilisées pour la tâche de prediction**

#### **1.4.2.2.1. Les variables indépendantes (prédictives)**

Une grande partie des études existantes relatives à la prédiction du churn met l'accent sur l'intégration de deux types de données clients, qui sont les données comportementales et les données démographiques. L'utilisation de données démographiques est excellente lorsque nous n'avons pas accès à d'autres informations sur nos clients. Mais dans un monde en ligne, où nous pouvons suivre chaque action client dans notre boutique en ligne, ceci nous permet d'avoir l'accès à une très grande quantité d'informations sur les comportements d'achats et les habitudes de navigations, que nous pouvons les utiliser pour offrir des expériences d'achat uniques et plus pertinentes. En outre, selon plusieurs études telles que Coussement et Van den Poel [127], Guadagni et Little [186], Rossi et al., [187], et Tamaddoni Jahromi et al., [188], les données démographiques (sexe, âge, adresse, profession, etc.) ont moins d'impact sur la prédiction du churn. Pour cela, notre étude se basera uniquement sur les données comportementales au niveau du client individuel (les variables indépendantes), ceci va nous permettra de garder nos modèles de prédiction dans leur forme la plus simple et, d'autre part, de maximiser leur pouvoir prédictif.

Comparé aux méthodes de transaction traditionnelles, le plus grand avantage du commerce électronique est que toutes les données de navigation de toutes les visites effectuées par les clients sur le site de commerce électronique sont stockées dans les serveurs. A partir de ces données comportementales et transactionnelles (page vue, séquence de visites, processus d'achat, nombre de transactions, etc ...) nous pouvons extraire de nombreux variables et indicateurs, et de les utiliser comme des variables prédictives par nos modèles, pour améliorer leur pouvoir de distinction entre les clients totalement défectueux, partiellement défectueux, et ceux qui restent fidèles dans leurs comportements. La table 4.8 montre comment ces variables prédictives ont été calculé dans la phase de conception de data warehouse.

La table 4.9 résume toutes les variables comportementales prédictives examinées par des recherches antérieures dans les environnements hors ligne et en ligne.

Table 4.8 - Les variables prédictives utilisées dans notre étude

| Type de variable     | Nom variable                             | Description  |
|----------------------|--|--|
| Récence              | R  | Nombre de jours entre le premier jour de la période d'étude et le jour du dernier achat dans la période de calibration ( $0 \leq R \leq T1$ ).   |
| Fréquence            | F  | Nombre de transaction enregistrées durant la période de calibration (T1).  |
|                      | R_change.F                               | Changement relatif du nombre de transactions effectuées dans la seconde moitié de la période de calibration (T1.2), noté F.T1.2, par rapport à la première moitié de la même période (T1.1), noté F.T1.1, soit $R\_change.F = (F.T1.2 - F.T1.1) / F.T1.1$        |
| Montant              | M  | Montant total des achats pendant la période de calibration (T1).   |
| Longueur de relation | L  | Nombre de jours entre le premier et le dernier achat.  |
| Inter-achat          | IPT                                      | Nombre moyen de jours entre les achats.  |
|                      | R_change.IPT                             | Changement relatif du temps d'inter-achat dans la seconde moitié de la période de calibration (T1.2), noté IPT.T1.2 par rapport à la première moitié de la période de calibration (T1.1), noté IPT.T1.1, soit $R\_change.IPT = (IPT.T1.2 - IPT.T1.1) / IPT.T1.1$ |
| Méthode de paiement  | Mopayment                                | Indique le mode de paiement le plus utilisé dans les trois dernières transactions.   |
| Taux d'abandon       | Last_session_abandoned                   | Indique si la dernière session a été abandonnée à l'étape de Checkout (oui, non).  |
|                      | aband_rate(allvisit to productviews)T1.2 | Le pourcentage de sessions qui ont abandonné le processus d'achat à l'étape "Vues de produits" dans la seconde moitié de la période d'étalonnage T1.2.   |
|                      | aband_rate(productviews to addcart) T1.2 | Taux d'abandon pour un client lors du passage de l'étape "Vues de produits" à l'étape "Ajouter au panier" dans la seconde moitié de la période de calibration T1.2.  |
|                      | aband_rate(addcart to checkout)T1.2      | Taux d'abandon pour un client lors du passage de l'étape "Ajouter au panier" à l'étape "Checkout" dans la seconde moitié de la période de calibration T1.2.  |
|                      | aband_rate(checkout to                   | Taux d'abandon pour un client lors du passage de l'étape Checkout à  |

|  |   |   |
|--|---|---|
|  | transaction)T1.2                        | l'étape Transaction dans la seconde moitié de la période de calibration T1.2.                 |
|  | aband_rate(allvisit to transaction)T1.2 | Le pourcentage total de sessions qui ont abandonné le processus d'achat dans la période T1.2. |

La table 4.9 - montre que les variables R, F et M sont celles les plus utilisées pour prédire le churn dans les environnements en ligne. D'un autre côté, le taux d'abandon, la longueur de la relation (L), le temps moyen d'inter-achat (IPT) et le mode de paiement (Mopayment) sont des variables rarement utilisées dans les recherches antérieures. Par conséquent, afin d'évaluer leur importance dans la prédiction du churn, nous les prendrons en compte.

Table 4.9 - Les prédicteurs comportementaux de la défection et le type de variable cible utilisées dans les recherches antérieures

|                     |   | Variables prédictives |   |   |   |         |                     |                     |                | Variable cible                                   |
|---------------------|---|-----------------------|---|---|---|---------|---------------------|---------------------|----------------|--|
|                     |   | L                     | R | F | M | IPT/IAT | Catégories produits | Méthode de paiement | Taux d'abandon | Variable type                                    |
| Offline environment | Buckinx and van den Poel (2005) [157]         | X                     | X | X | X | X       | X                   | X                   |                | <b>Binaire:</b> (Churner, Non-churner)           |
|                     | V.L. Miguéis, et al (2012) [128]              | X                     | X | X | X |         |                     |                     |                | <b>Binaire:</b> (Partially churned, Non-churner) |
|                     | Mozer, et al. (2000) [152]                    | X                     |   | X | X | X       | X                   | X                   |                | <b>Binaire:</b> (Churner, Non-churner)           |
| Online environment  | Keaveney and Parthasarathy (2001) [192]       |                       |   | X | X |         |                     |                     |                | <b>Binaire:</b> (switchers, continuers)          |
|                     | K. Coussement and, K. W. De Bock (2013) [135] | X                     | X | X | X | X       |                     |                     |                | <b>Binaire:</b> (Churner, Non-churner)           |
|                     | A. T. Jahromi, et al. (2014) [188]            |                       | X | X | X |         |                     |                     |                | <b>Binaire:</b> (Churner, Non-churner)           |
|                     | Ssu-Han Chen (2016) [129]                     |                       | X |   |   | X       |                     |                     |                | <b>Binaire:</b> (Churner, Non-churner)           |
|                     | N. Gordinia and V.                            | X                     | X | X | X |         | X                   |                     |                | <b>Binaire:</b> (Churner, Non-churner)           |

|  |                        |   |   |   |   |   |  |   |   |  |
|--|------------------------|---|---|---|---|---|--|---|---|--|
|  | Vegliob<br>(2017) [32] |   |   |   |   |   |  |   |   |  |
|  | Cette étude            | X | X | X | X | X |  | X | X | <b>Multi-classe:</b> (Partially-churned, Totally-churned, Non-churner) |

#### 1.4.2.2.2. La variable cible (prédite)

La variable cible dans l'étude actuelle est nommée «statut». C'est une variable catégorielle qui a trois valeurs: 'Partially-churned', 'Totally-churned' et 'Non-Churner', elle est prédite en fonction de l'historique des événements des clients sur le site Web de commerce électronique.

### 1.5. Phase d'évaluation

L'évaluation expérimentale de nos modèles représente la dernière étape du processus de prédiction. Elle tente généralement d'évaluer l'efficacité d'un modèle prédictif, à savoir sa capacité de prendre les décisions de prédiction.

#### 1.5.1. Les mesures de performances

La table 4.9 révèle que toutes les études d'analyse de churn existantes, se concentrent sur les modèles de classification binaires. Cette étude contribue à la littérature en ne mettant pas l'accent sur les modèles de classification binaires où le modèle prédit l'état d'un client en tant que 'Churner' ou 'Non-churner'. Notre étude abordera, toutefois, le cas des problèmes de classification multi-classes où la variable cible, classe un client particulier soit en tant que client poursuivant son modèle fidèle d'achat (Non-churner), ou en tant que client partiellement défectueux (*Partially-churned*) ou en tant que client totalement défectueux. (*Totally-churned*).

Toutes les mesures utilisées pour évaluer nos modèles, sont calculées à partir de la matrice de confusion. En effet, Une matrice de confusion est un outil de visualisation couramment utilisé dans l'apprentissage automatique supervisé pour évaluer la performance d'un classifieur. Elle contient des informations sur les classifications réelles et celles prédites par un système de classification. Généralement, chaque colonne de la matrice représente les instances de la classe prédite, tandis que chaque ligne représente les instances de la classe réelle. Un exemple de matrice de confusion pour un modèle de classification binaire est illustré dans la table 4.10.

Table 4.10 - Matrice de confusion

|               |   | Classe prédite                                  |   |
|---------------|---|---|---|
|               |   | P   | N   |
| Classe réelle | P | <b>TP</b> True Positives<br>(VP Vrais positif)  | <b>FN</b> False Negatives<br>(FN Faux négatifs) |
|               | N | <b>FP</b> False Positives<br>(FP Faux positifs) | <b>TN</b> True Negatives<br>(VN Vrais négatifs) |

Les mesures extraites à partir de la matrice de confusion sont calculées comme suit :

$$\text{L'exactitude} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{L'erreur} = \frac{FP+FN}{TP+TN+FP+FN} \quad (5)$$

$$\text{La précision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Le rappel} = \frac{TP}{TP+FN} \quad (7)$$

$$\text{La spécificité} = \frac{TN}{TN+FP} \quad (8)$$

$$\text{F-score} = \frac{((1+\beta^2) * Precision * Rappel)}{((\beta^2 * Precision)) + Rappel}, \quad \text{avec } \beta^2 = 1 \quad (9)$$

Les symboles et les termes utilisés dans ces équations sont résumés dans la table suivante :

Table 4.11 - Mesures de performance

| Mesure                       | Signification   |
|------------------------------|---|
| TP ( <i>True Positive</i> )  | Nombre d'exemples de classe $C_i$ correctement reconnus                           |
| TN ( <i>True Negative</i> )  | Nombre d'exemples correctement reconnus qui n'appartiennent pas à la classe $C_i$ |
| FP ( <i>False Positive</i> ) | Nombre d'exemples qui ont été incorrectement assignés à la classe $C_i$           |



|                              |   |
|------------------------------|---|
| FN ( <i>False Negative</i> ) | Nombre d'exemples qui n'ont pas été reconnus comme des exemples appartiennent à la classe |
| Exactitude                   | Indice d'exactitude globale d'un classificateur (% des <b>prédictions correcte</b> )      |
| Précision                    | % de prédictions positives qui sont correctes   |
| Rappel (sensibilité)         | % d'exemples positifs prédits comme positives   |
| Spécificité                  | % d'exemples négatifs prédits comme négatives   |
| F-score                      | Traduit l'équilibre entre la précision et le rappel.                                      |

Un classificateur parfait serait décrit comme 100% de sensibilité et 100% de spécificité. Dans la grande majorité des cas, on se limite à la précision et au rappel.

En termes de problèmes de classification multi-classes, il existe différentes mesures de performance pour vérifier les performances d'un classifieur multi-classes. En général, ce sont les mêmes utilisées dans le cas d'un problème binaire.

La table 4.12 montre un exemple typique du résultat d'une matrice de confusion pour un problème à trois classes, où  $N^{ij}$  représente le nombre d'exemples avec comme classe réelle  $i$  et qui sont prédites comme classe  $j$  ( $i=1, 2,3, j=1, 2,3$ ). Généralement, dans le cas d'un problème à  $C$  classes la matrice de confusion obtenue dispose de  $C \times C$  positions, contenant les  $C$  bonnes classifications (les éléments sur la diagonale principale) et  $C(C - 1)$  erreurs possibles (les éléments hors de la diagonale).

Table 4.12 - Matrice de confusion : cas de problème à trois classes

|        |                    | Prédite            |                    |                    |
|--------|--------------------|--------------------|--------------------|--------------------|
|        |                    | Class <sub>i</sub> | Class <sub>j</sub> | Class <sub>k</sub> |
| Réelle | Class <sub>i</sub> | $N^{ii}$           | $N^{ij}$           | $N^{ik}$           |
|        | Class <sub>j</sub> | $N^{ji}$           | $N^{jj}$           | $N^{jk}$           |
|        | Class <sub>k</sub> | $N^{ki}$           | $N^{kj}$           | $N^{kk}$           |

Les équations de 10 à 17 résument les mesures de performance les plus souvent utilisées dans le cas d'un problème de classification multi-classes basées sur  $TP_i$  : instances de Class<sub>i</sub>

correctement prédites,  $TN_i$  : instances d'autres classes (j, k) correctement prédites,  $FP_i$  : instances d'autres classes (j, k) prédites comme instances de  $Class_i$  et  $FN_i$  : instances de  $Class_i$  incorrectement prédites. L'indice  $M$  et  $\mu$  représente respectivement la macro-moyenne et la micro-moyenne [189].

Pour une classe individuelle  $Class_i$ , l'évaluation est définie par  $TP_i$ ,  $FN_i$ ,  $FP_i$  et  $TN_i$ . Les autres mesures  $Précision_i$ ,  $Rappel_i$  et  $F-score_i$  sont calculées à partir des chiffres de  $Class_i$ . La qualité globale de prédiction est généralement évaluée par deux méthodes : Macro-averaging (macro moyenne) ou Micro-averaging (micro moyenne). Macro-averaging est la moyenne des mêmes mesures qui sont calculées pour chaque classe (représentée avec l'indice  $M$ ), la deuxième méthode a une idée similaire à la Macro-averaging mais calcule une mesure de performance à partir de la somme de  $TP$ ,  $FN$ ,  $TN$ ,  $FP$  de toutes les classes (représentée par des indices  $\mu$ ) [189]. Cependant, la deuxième méthode (Micro-averaging) ne fournit pas une mesure précise des performances lorsque les instances ne sont pas équitablement réparties sur les classes (le cas où la plupart des instances appartiennent à une classe). Contrairement à la Micro-averaging, Macro-averaging fournit une mesure de performance significative malgré que les données ne soient pas distribuées de manière égale entre toutes les classes (le cas des classes déséquilibrées) [190]. Pour cela, nous avons utilisé la Macro-averaging comme une mesure de performance pour évaluer la qualité de prédiction de nos modèles multi-classe.

En outre, nous considérons une cross-validation à 10 folds dans laquelle les données initiales sont divisées au hasard en dix parties égales, neuf parties sont utilisées comme données d'apprentissage pour construire le modèle de prédiction, tandis que la partie restante est réservée comme ensemble de test. L'opération se répète ainsi 10 fois pour qu'en fin de compte chaque partie soit utilisée exactement neuf fois pour l'entraînement et une fois pour le test. Le taux moyen d'exactitude est enfin calculé pour estimer la qualité de prédiction.

$$\text{Exactitude Moyenne} = \frac{\sum_{i=1}^n \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{n} \quad (10)$$

$$\text{Taux d'erreur moyen} = \frac{\sum_{i=1}^n \frac{FP_i + FN_i}{TP_i + TN_i + FP_i + FN_i}}{n} \quad (11)$$

$$\text{Précision}_{\mu} = \frac{TP_1 + \dots + TP_n}{TP_1 + \dots + TP_n + FP_1 + \dots + FP_n} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i)} \quad (12)$$

$$\text{Rappel}_\mu = \frac{TP_1 + \dots + TP_n}{TP_1 + \dots + TP_n + FN_1 + \dots + FN_n} = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)} \quad (13)$$

$$\text{Fscore}_\mu = \frac{2 * \text{Précision}_\mu * \text{Rappel}_\mu}{\text{Précision}_\mu + \text{Rappel}_\mu} \quad (14)$$

$$\text{Précision}_M = \frac{\text{Précision}_1 + \dots + \text{Précision}_n}{n} = \frac{\sum_{i=1}^n \frac{TP_i}{TP_i + FP_i}}{n} \quad (15)$$

$$\text{Rappel}_M = \frac{\text{Rappel}_1 + \dots + \text{Rappel}_n}{n} = \frac{\sum_{i=1}^n \frac{TP_i}{TP_i + FN_i}}{n} \quad (16)$$

$$\text{F-score}_M = \frac{2 * \text{Précision}_M * \text{Rappel}_M}{\text{Précision}_M + \text{Rappel}_M} \quad (17)$$

## 2. Résultats et discussion

Afin d'évaluer la qualité des prédictions de trois modèles de prédiction d'attrition comprenant l'arbre de décision (DT), le réseau de neurones artificiels (ANN) et l'ensemble d'arbre de décision (DTE), les mesures de Macro-averaging et une cross-validation à 10 folds sont considérées. Les tables 4.14 à 4.16 illustrent la matrice de confusion de chacun des modèles construits. La fraction réelle des trois classes : 'Partially-churned', 'Totally-churned' et 'Non-churner' dans l'ensemble de données est de 17,81%, 25,45% et 56,73% respectivement.

Table 4.13 - Exactitude de prédiction des trois modèles avec une cross-validation à 10 folds.

|        | DT    | ANN   | DTE   |
|--------|-------|-------|-------|
| fold 1 | 93.71 | 90.21 | 93.71 |
| fold 2 | 95.10 | 95.10 | 98.60 |
| fold 3 | 95.80 | 97.20 | 98.60 |
| fold 4 | 93.01 | 90.91 | 95.10 |
| fold 5 | 92.31 | 91.61 | 95.10 |
| fold6  | 96.50 | 97.90 | 99.30 |

|            |              |              |              |
|------------|--------------|--------------|--------------|
| fold7      | 96.48        | 93.66        | 97.18        |
| fold8      | 95.07        | 92.25        | 97.18        |
| fold9      | 92.25        | 93.66        | 97.18        |
| fold10     | 95.07        | 92.25        | 96.48        |
| <b>Avg</b> | <b>94.53</b> | <b>93.48</b> | <b>96.84</b> |

La table 4.13 montre les performances de prédiction des trois modèles, qui sont calculées sur la base d'une validation croisée 10-folds. En moyenne, les trois modèles de prédiction fournissent un indice d'exactitude supérieur à 93%. En comparant les différentes techniques de classification, DTE offre dans toutes les parties de test (10 folds) les meilleurs résultats en termes de pourcentage de clients correctement classés. Tandis que la performance prédictive moyenne des modèles DT et ANN est très proche. L'indice d'exactitude (bien prédits), est parfois assez trompeur et insuffisant pour confirmer la qualité de la prédiction [191]. Pour cela, des mesures de performance supplémentaires, telles que le rappel et la précision, sont nécessaires pour identifier le modèle de prédiction du churn le plus performant. Par conséquent, sur la base des tables de matrice de confusion, nous calculons les valeurs Rappel<sub>i</sub>, Precision<sub>i</sub> et F-score<sub>i</sub> pour chaque classe C<sub>i</sub> afin d'évaluer la performance par rapport à chacune des trois classes de notre ensemble de données. Les résultats détaillés sont présentés dans les tables 4.17,4.18 et 4.19.

Table 4.14 - La matrice de confusion (par Artificial Neural Networks).

| Status \ Prediction<br>(Status) | Partially-churned | Totally-churned | Non-churned |
|---------------------------------|-------------------|-----------------|-------------|
| Partially-churned               | 192               | 26              | 36          |
| Totally-churned                 | 12                | 350             | 1           |
| Non-churned                     | 17                | 1               | 791         |

Table 4.15 - La matrice de confusion (par Simple Decision Tree).

| Status \ Prediction<br>(Status) | Partially-churned | Totally-churned | Non-churned |
|---------------------------------|-------------------|-----------------|-------------|
|---------------------------------|-------------------|-----------------|-------------|

|                   |     |     |     |
|-------------------|-----|-----|-----|
| Partially-churned | 218 | 13  | 23  |
| Totally-churned   | 17  | 346 | 0   |
| Non-churned       | 25  | 0   | 784 |

Table 4.16 - La matrice de confusion (par Decision Tree Ensemble).

| Status \ Prediction<br>(Status) | Partially-churned | Totally-churned | Non-churned |
|---------------------------------|-------------------|-----------------|-------------|
| Partially-churned               | 218               | 14              | 22          |
| Totally-churned                 | 6                 | 357             | 0           |
| Non-churned                     | 3                 | 0               | 806         |

Les tableaux 4.17 à 4.19 résument les valeurs de rappel, précision et F-score pour chacune des trois classes pour les trois techniques de classification avec une cross-validation à 10 folds.

Table 4.17 - L'exactitude globale, les mesures de Précision, Rappel et F-score pour les trois classes (Réseaux de neurones).

| ANN               | Exactitude globale | TP  | FP | TN   | FN | Recall | Precision | F-score |
|-------------------|--------------------|-----|----|------|----|--------|-----------|---------|
| Partially-churned | 93.48 %            | 192 | 29 | 1143 | 62 | 0.756  | 0.869     | 0.808   |
| Totally-churned   |                    | 350 | 27 | 1036 | 13 | 0.964  | 0.928     | 0.946   |
| Non-churned       |                    | 791 | 37 | 580  | 18 | 0.978  | 0.955     | 0.966   |

Table 4.18 - L'exactitude globale, les mesures de Précision, Rappel et F-score pour les trois classes (Arbre de décisions).

| DT                | Exactitude globale | TP  | FP | TN   | FN | Recall | Precision | F-score |
|-------------------|--------------------|-----|----|------|----|--------|-----------|---------|
| Partially-churned | 94.53 %            | 218 | 42 | 1130 | 36 | 0.858  | 0.838     | 0.848   |

|                 |     |    |      |    |       |       |       |
|-----------------|-----|----|------|----|-------|-------|-------|
| Totally-churned | 346 | 13 | 1050 | 17 | 0.953 | 0.964 | 0.958 |
| Non-churned     | 784 | 23 | 594  | 25 | 0.969 | 0.971 | 0.97  |

Table 4.19 - L'exactitude globale, les mesures de Précision, Rappel et F-score pour les trois classes (Ensemble d'arbres de décisions).

| DTE               | Exactitude globale | TP  | FP | TN   | FN | Recall | Precision | F-score |
|-------------------|--------------------|-----|----|------|----|--------|-----------|---------|
| Partially-churned | 96.84 %            | 218 | 9  | 1163 | 36 | 0.858  | 0.960     | 0.906   |
| Totally-churned   |                    | 357 | 14 | 1049 | 6  | 0.983  | 0.962     | 0.973   |
| Non-churned       |                    | 806 | 22 | 595  | 3  | 0.996  | 0.973     | 0.985   |

Table 4.20 – Les valeurs de la mesure Macro-averaging pour les trois modèles.

|                                  | Exactitude moyenne | Recall <sub>M</sub> | Precision <sub>M</sub> | F-score <sub>M</sub> |
|----------------------------------|--------------------|---------------------|------------------------|----------------------|
| Artificial Neural Networks (ANN) | 0.957              | 0.899               | 0.917                  | 0.907                |
| Simple Decision Tree (DT)        | 0.964              | 0.927               | 0.924                  | 0.925                |
| Decision Tree Ensemble (DTE)     | 0.979              | 0.946               | 0.965                  | 0.955                |

Les résultats présentés dans les tables 4.17, 4.18 et 4.19 indiquent que pour les trois classes de notre base de données, le modèle DTE offre les meilleurs résultats en termes d'exactitude globale, précision, rappel et F-score.

De plus, comme indiqué dans la table 4.20, le modèle DTE montre de meilleures performances prédictives que les deux autres modèles en termes de mesures Macro-averaging. Comparé aux DT et ANN, l'augmentation est respectivement de + 1,90 + 4,63 points en termes de Recall<sub>M</sub>, tandis que l'amélioration de Precision<sub>M</sub> est respectivement de +4,06 + 4,76 et + 2,93, + 4,80 en termes de F-Score<sub>M</sub>. D'après ces résultats, nous concluons que le modèle construit par

DTE est celui qui garantit la meilleure distinction entre les clients totalement défectueux, partiellement défectueux et ceux qui restent fidèles. En outre, la performance des modèles basé sur les méthodes d'ensembles (ensemble de classifieurs) est plus robuste et meilleure que les modèles basés sur un seul classifieur.

Enfin, nous examinons les variables prédictives qui contribuent davantage à la prédiction de la défection partielle et totale des clients en utilisant les trois modèles. Pour cela, nous créons un workflow qui permet de calculer l'importance de chacune de ces variables, l'idée de base est : pour calculer l'importance de la variable  $k$ , nous l'excluons dans la phase d'apprentissage du modèle. Si la prédiction du modèle diminue à l'absence de cette variable, cela indique que cette dernière est importante dans le processus de prédiction. La même procédure est utilisée pour toutes les variables prédictives et enfin, la variable d'importance pour chaque variable prédicteur est normalisée de 0 à 1 pour obtenir un indicateur d'importance de chaque variable. Les résultats sont présentés dans la table 4.21. Par conséquent, avoir une connaissance des variables prédictives d'attrition les plus importantes dans le secteur du commerce électronique, serait d'un intérêt particulier pour les décideurs et les directeurs de marketing. Car elles fournissent des informations exploitables permettent de cibler les clients les plus susceptibles de mettre fin à leur relation avec le site e-commerce par des incitations personnalisées afin de minimiser le risque d'attrition [32].

La table 4.21 montre l'importance des différentes variables prédictives utilisées dans la présente étude pour les trois modèles DT, ANN et DTE, où la première colonne indique le rang d'importance, la deuxième le nom de la variable et la troisième colonne montre l'importance normalisée d'une variable en fonction du taux de précision offert par le modèle quand il n'utilise pas cette dernière lors de sa phase d'apprentissage.

Table 4.21 - Importance des différentes variables prédictives utilisées pour la prédiction de churn.

| Rank | DTE                                  |                               | DT                                   |                               | ANN                                  |                               |
|------|--------------------------------------|-------------------------------|--------------------------------------|-------------------------------|--------------------------------------|-------------------------------|
|      | Variable                             | Normalisation de l'importance | Variable                             | Normalisation de l'importance | Variable                             | Normalisation de l'importance |
| 1    | aband_rate(checkout-transaction)T1.2 | 1.000                         | aband_rate(checkout-transaction)T1.2 | 1.000                         | aband_rate(checkout-transaction)T1.2 | 1.000                         |
| 2    | aband_rate(productviews-addcart)T1.2 | 0.562                         | aband_rate(productviews-addcart)T1.2 | 0.287                         | aband_rate(productviews-addcart)T1.2 | 0.652                         |
| 3    | aband_rate(allv-transaction)T1.2     | 0.210                         | aband_rate(addcart-checkout)T1.2     | 0.181                         | aband_rate(allv-transaction)T1.2     | 0.277                         |
| 4    | aband_rate(addcart-checkout)T1.2     | 0.181                         | aband_rate(allv-productviews)T1.2    | 0.070                         | aband_rate(addcart-checkout)T1.2     | 0.241                         |
| 5    | L                                    | 0.038                         | R                                    | 0.064                         | L                                    | 0.143                         |
| 6    | aband_rate(allv-productviews)T1.2    | 0.038                         | Last_session_abandoned               | 0.058                         | R                                    | 0.098                         |
| 7    | IPT                                  | 0.029                         | aband_rate(allv-transaction)T1.2     | 0.053                         | R_change.F                           | 0.089                         |
| 8    | R_change.IPT                         | 0.029                         | L                                    | 0.053                         | IPT                                  | 0.071                         |
| 9    | F                                    | 0.019                         | IPT                                  | 0.053                         | R_change. IPT                        | 0.071                         |
| 10   | M                                    | 0.010                         | R_change.F                           | 0.053                         | F                                    | 0.071                         |
| 11   | Mode-ofpayment                       | 0.010                         | M                                    | 0.047                         | aband_rate(allv-productviews)T1.2    | 0.063                         |
| 12   | Last_session_abandoned               | 0.010                         | F                                    | 0.035                         | M                                    | 0.027                         |
| 13   | R                                    | 0.010                         | Mode-ofpayment                       | 0.029                         | Mode-ofpayment                       | 0.027                         |
| 14   | R_change.F                           | 0.000                         | R_change. IPT                        | 0.000                         | Last_session_abandoned               | 0.000                         |

Il ressort clairement des classements d'importance des variables prédictives qui décrivent les taux d'abandon dans les étapes du processus d'achat en ligne comme: « *aband\_rate(checkout-*



*transaction*)T1.2 », « *aband\_rate(productviews-addcart)*T1.2 » et « *aband\_rate(addcart-checkout)*T1.2 » sont, en effet, des variables puissantes et contribuent d'une façon très importante dans la prédiction de deux degrés de churn dans le domaine du commerce électronique. Cela est confirmé par la similitude dans le classement d'importance de ces trois variables, où nous les trouvons classer parmi les quatre premières variables dans tous les trois modèles et semblent surpasser les autres variables. La principale différence entre le modèle DTE et les autres modèles est, plus clairement liée à la variable qui décrit la récence du dernier achat (R) ; elle apparaît en bas du classement pour DTE et assez importante dans les deux autres modèles.

Toutefois, en comparant les résultats de notre étude avec d'autres recherches récentes qui traitent la prédiction de churn dans les environnements en ligne A. T. Jahromi et al[188], N. Gordini et V. Vegliob [32], et K. Coussement and K. W. De Bock [135], il apparaît d'une part, qu'en terme d'importance de variables utilisées dans la phase d'apprentissage, les variables de récence et de fréquence semblent moins importantes pour la prédiction de churn; Cela contraste les attentes que nous avons formulées à partir des recherches existantes qui soulignent avec force le pouvoir prédictif des variables R et F du modèles RFM. Ceci est dû au fait que les clients impliqués par nos modèles de prédiction sont ceux qui représentent les principaux clients et les nouveaux clients à forte valeur, plus précisément, les clients appartenant aux clusters  $\bar{L} \bar{R} \bar{F} \bar{M}$ ,  $\bar{L} \bar{R} \bar{F} \underline{M}$ ,  $\bar{L} \bar{R} \underline{F} \bar{M}$  et  $\underline{L} \bar{R} \bar{F} \bar{M}$ . En observant la table 4.7, nous voyons que les valeurs moyennes de F et R de ces quatre clusters sont presque proches. De plus, les statistiques descriptives présentées dans la table 4.6 indiquent que l'écart-type de F et R est faible, ce qui signifie que les valeurs de ces deux variables sont peu dispersées autour de la moyenne. Par conséquent, cela rend leur contribution à la distinction entre les clients totalement défectueux, partiellement défectueux et ceux qui sont fidèles moins importantes.

Une autre explication pourrait résider dans le fait que ces études n'ont pas exploité totalement le volume important de données généré par les environnements en ligne, mais elles sont restées limitées aux variables qui caractérisent les environnements hors ligne telles que la récence, la fréquence et le montant. En effet, à partir des événements effectués par les clients sur le site marchand, nous pouvons extraire facilement de nombreuses variables prédictives qui ont un pouvoir explicatif dans la compréhension de leurs comportements, et dans l'analyse de leur expérience d'achat qui commence par la consultation du produit et se termine avec la validation

de transaction. Par exemple, nous pouvons extraire des variables qui décrivent pour chaque client le taux d'abandon de sessions aux différentes étapes de processus d'achat, la durée moyenne d'une session, les produits les plus consultés, nombre de page vues par session, etc....

D'une autre part, contrairement aux recherches antérieures qui mettait l'accent soit uniquement sur la prédiction de défection partielle, soit uniquement sur la prédiction de défection totale dans les contextes non-contractuels, notre approche permet de suivre à la fois ces deux types de défection, ce qui permet aux gestionnaires de marketing d'optimiser l'affectation des ressources en fonction de degré d'attrition.

Une utilisation vitale de la visualtion dans la phase de déploiement est le produit final de tout système décisionnel. Un rapport interactif sous forme de plusieurs visualisations est illustré dans la figure 4.14. Il fournit des informations sur les trois classes de clients qui sont prédites par le modèle prédictif. Ces information concerne le pourcentage de chaque classe, les taux moyens d'abandon dans chaque étape de paiement en ligne, ainsi ce rapport donne la probabilité moyenne de l'attrition partielle, de l'attrition totale et la probabilité de rester fidèle dans le futur.

Ce rapport permet aux managers de faire un filtre selon le niveau d'attrition, et donc d'avoir une vue plus précise sur chaque classe. Par exemple si le décideur veut lister seulement les clients qui vont quitter totalement le site e-commerce dans le futur, il suffit juste d'utiliser le menu situé en haut à droite du rapport et de sélectionner la classe correspondante. Cela va afficher seulement les clients concernés, ensuite la sélection d'un client va automatiquement donner sa probabilité de quitter totalement le site e-commerce, sa probabilité de quitter partiellement le site et aussi sa probabilité de rester fidèle. Un exemple est illustré dans la figure 4.15.

Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision

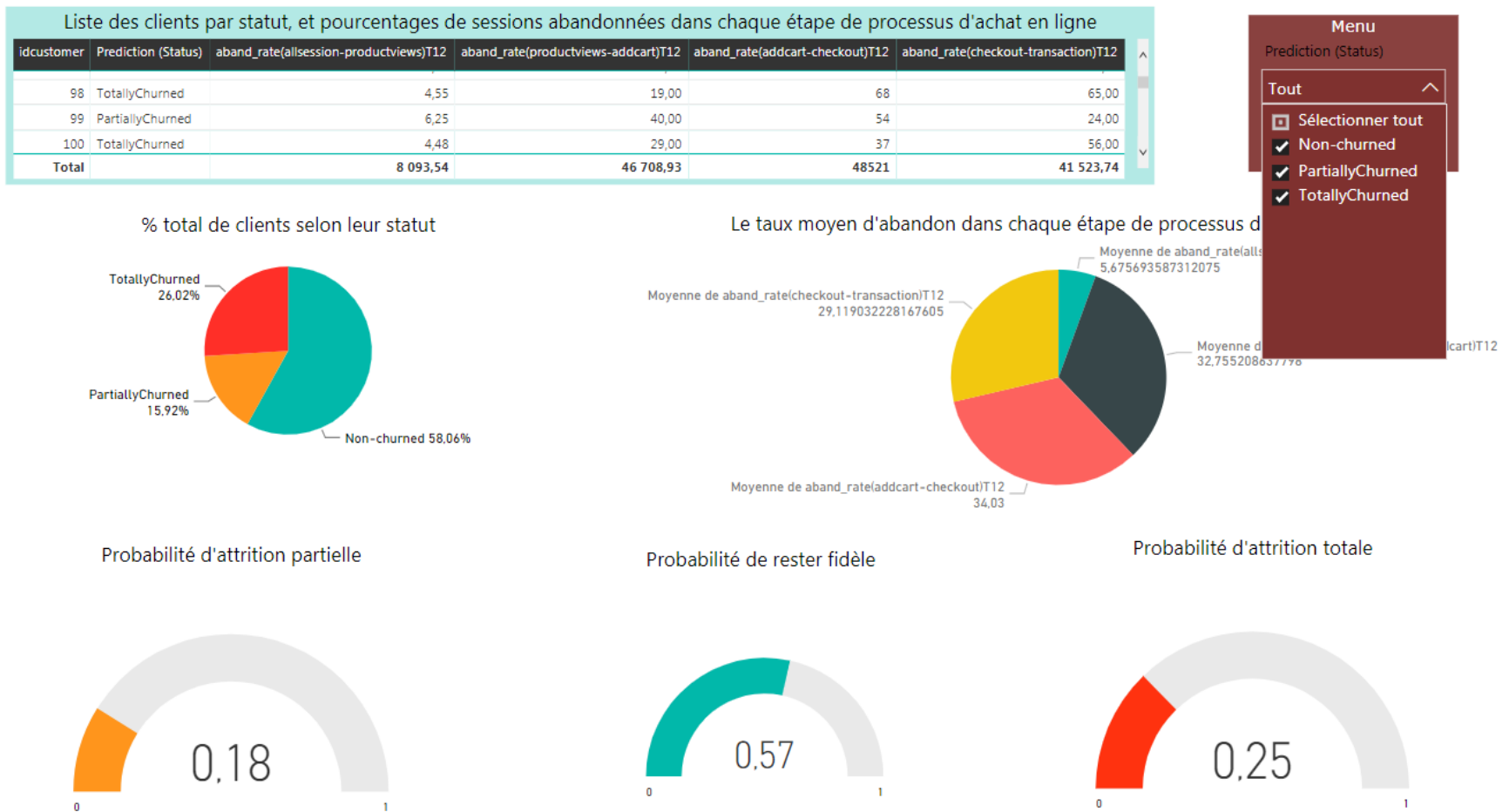


Figure 4.14- Vue d'ensemble sur les trois classes de clients en termes de taux d'abandon et de probabilité d'attrition.

Chapitre 4 : Présentation et implimentation de notre système d'aide à la décision

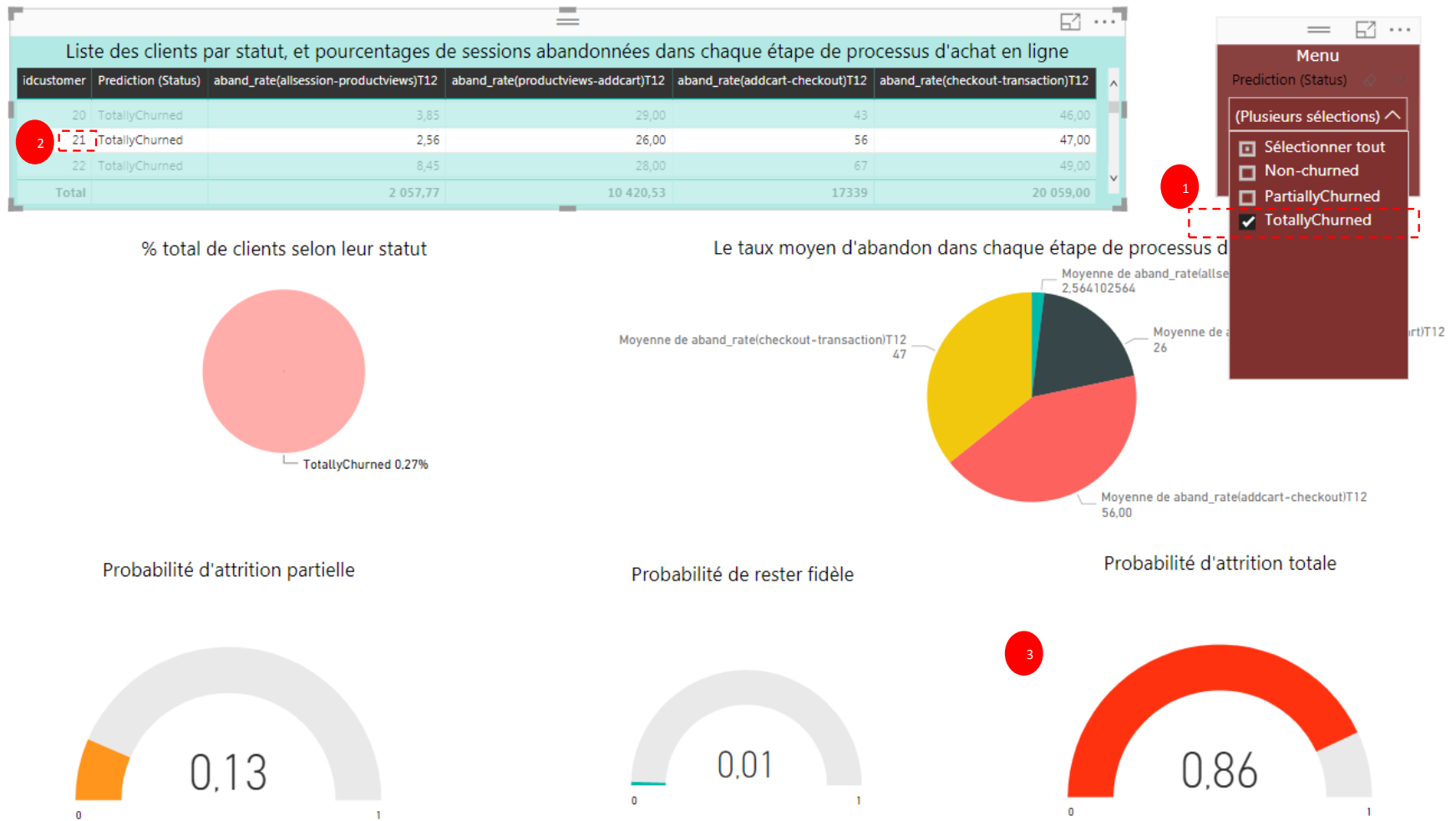


Figure 4.15- Exemple de résultat du rapport lorsque nous choisissons un clients précis.

## Conclusion

Afin de résoudre le problème crucial de la définition du churn dans les cadres non-contractuels, nous avons considéré deux sous périodes égales T1 et T2. Dans T1, le modèle LRFM et la technique de clustering (k-means) sont combinés comme première étape afin d'identifier les différents types de profils clients (les modèles LRFM). Par conséquent, nous trouvons sept groupes de clients qui ont un comportement LRFM différent ( $\underline{L} \overline{R} \overline{F} \overline{M}$ ,  $\overline{L} \underline{R} \underline{F} \underline{M}$ ,  $\overline{L} \overline{R} \overline{F} \underline{M}$ ,  $\overline{L} \overline{R} \underline{F} \overline{M}$ ,  $\overline{L} \underline{R} \underline{F} \underline{M}$ ,  $\underline{L} \overline{R} \underline{F} \underline{M}$ ), puis nous définissons le changement de modèle LRFM d'un client au fil du temps (de T1 vers T2) comme un signal précoce d'une défection partielle ou totale. À notre avis, la méthodologie proposée pour la définition du churn peut être un outil de décision utile pour les entreprises opérant dans des cadres non contractuels, où les clients et les entreprises n'ont aucun contrat entre eux. Après avoir résolu le problème de la définition de churn partiel et total, nous avons proposé trois modèles prédictifs basés sur les réseaux de neurones, les arbres de décision et les méthodes d'ensemble pour la prédiction du phénomène de churn dans les cadres non-contractuels, particulièrement dans le secteur du commerce électronique.

Pour tester les modèles proposés dans un contexte réel, nous avons utilisé comme étude de cas un site marchand en ligne, où les enregistrements de comportement des clients de la période allant du 1er novembre 2013 au 28 février 2015 ont été utilisés. Les résultats obtenus montrent que les trois modèles proposés peuvent fournir une prédiction au niveau individuel de la probabilité d'une perte partielle ou totale des clients dans un proche avenir, ce qui nous permettrait de suivre et cibler les clients concernés par des stratégies de rétention afin de réduire le taux d'attrition. Par conséquent, cette prédiction est très utile pour les responsables marketing puisqu'elle va grandement les aider à mettre en place de nouvelles solutions d'incitation adaptées (des actions de rétention) en fonction du degré de la défection (Partielle ou totale) pour les convaincre de rester.

En termes de problèmes de prédiction multi-classes, les performances de nos modèles sont évaluées et mesurées en utilisant la cross-validation à 10-fold et les mesures de Macro-averaging incluant l'efficacité moyenne,  $Precision_M$ ,  $Recall_M$  et  $F-score_M$ . Une analyse comparative de différents modèles est également présentée, les résultats de cette analyse comparative montrent que le modèle basé sur les méthodes d'ensemble (Decision Tree Ensemble) est plus puissant par

rapport aux autres modèles (Artificial Neural Networks et Decision Tree) en termes de qualité de prédiction. Ainsi les variables pouvant contribuer davantage à prédire la défection partielle et totale des clients dans le secteur du commerce électronique ont été identifiées.

Enfin, et comme le but de tout système d'aide à la décision, des rapports interactifs qui jouent un rôle crucial dans la phase de déploiement ont été créés sous forme de plusieurs visualisations.

# Conclusion et perspectives

## Revue des contributions

Dans un contexte concurrentiel, les entreprises cherchent à conserver leurs clients fidèles que d'attirer de nouveaux. Pour ce fait, les clients fidèles susceptibles de quitter partiellement ou totalement leurs relations avec l'entreprise dans un proche avenir, doivent être identifiés afin que l'entreprise puisse les approcher avec des incitations personnalisées ou avec autres offres de rétention sur mesure. C'est à cet objectif que souhaitent répondre les stratégies de rétention. Dans ce travail de recherche, nous nous sommes intéressés à l'application des techniques de data mining dans le contexte de la Gestion de la Relation Client en e-commerce. L'objectif principal de cette thèse est de comprendre et prédire le comportement des clients dans le cadre d'une stratégie de rétention. Pour adopter cette stratégie de rétention, nous devrions être équipés des modèles prédictifs capables d'identifier, avec précision les clients qui sont en voie de défection dans une période future déterminée.

Basé sur une gamme d'études réalisées sur le sujet de la perte de clients (chapitre 3), nous révélons que : (1) bien que plusieurs modèles prédictifs aient été développés pour modéliser l'attrition dans les cadres contractuels (banques, assurance, télécommunication), le cadre non-contractuel en général et le secteur de e-commerce en particulier ont reçu moins d'attention à cet égard. (2) le type d'attrition étudié : la plupart des approches existantes se basent, soit uniquement sur l'attrition partielle, soit uniquement sur l'attrition totale. (3) toutes les études de modélisation d'attrition existantes se concentrent sur les modèles de classification binaires où le modèle prédit l'état d'un client en tant que client défectueux ou non.

Pour répondre aux problématiques posées dans la thèse, nous avons traité le phénomène d'attrition dans un cadre non-contractuel, précisément dans le secteur de commerce électronique. Pour cela, nous avons développé une approche à deux étapes dans laquelle un modèle de segmentation basé sur l'algorithme k-means et sur le modèle LRFM et un autre modèle de prédiction basé sur des techniques de classification multi-classes sont combinés pour définir et prédire à la fois les deux degrés d'attrition (partielle et totale).

Pour avoir introduit, nous avons donné une introduction générale sur les contextes, les motivations, les objectifs et les contributions de cette thèse.

Dans les chapitres 1 et 2, nous avons parlé du concept de commerce électronique : ses différentes définitions, son évolution, ses différentes formes, ses avantages et ses inconvénients pour l'entreprise et le client. Nous avons détaillé toutes les étapes entrant dans un processus de transaction électronique ainsi que les deux principaux mécanismes utilisés pour sécuriser l'acte d'achat sur internet. Nous avons aussi jeté la lumière sur l'état des lieux du commerce électronique au Maroc et sur les organismes professionnels qui le représentent. Ensuite, nous avons passé à la présentation des systèmes de Gestion de la Relation Client (CRM) qui sont indispensables pour que le commerce électronique puisse réaliser les résultats souhaitables. Nous avons commencé par la description du concept CRM dans son sens général. En nous basant sur des recherches antérieures, nous avons essayé d'adopter les théories classiques du marketing relationnel dans le contexte de commerce électronique. Nous avons ensuite discuté la nécessité de la mise en œuvre d'un CRM analytique qui permet d'extraire, de stocker, d'analyser et d'exploiter les données brutes de l'entreprise dans le but d'améliorer la connaissance et la compréhension du client. La matière première de ce CRM analytique est la donnée et sa principale composante est la fouille de données (data mining).

L'état de l'art (chapitre 3) a présenté différentes techniques de data mining pour le CRM, en particulier celles utilisées dans les stratégies de rétention qui ont comme objectif de prédire les clients susceptibles de quitter leur relation avec l'entreprises dans un proche avenir. Les techniques de data mining : le clustering, la classification, la régression et les règles d'associations sont souvent utilisées dans les systèmes de CRM. Après cette étude bibliographique, nous avons identifié deux approches applicables pour l'objectif de la thèse : les «approches non ciblées» qui s'appuient sur des produits de qualité supérieure et sur une publicité de masse pour augmenter la fidélité de la marque et la rétention des clients, et les «approches ciblées» qui se reposent sur l'identification des clients susceptibles de partir vers les concurrents, afin de les cibler par des incitations directes et pour les inciter à rester et à continuer de réaliser des achats. Dans la présente thèse, nous avons adopté la deuxième approche.

Les contributions de nos travaux présentées dans le chapitre 4 sont résumées comme suit :



- Nous avons abordé la prédiction de churn dans les cadres **non-contractuels** où la perte des clients n'est pas facile à **définir** et à **tracer**.
- Nous avons développé une nouvelle **méthodologie** qui s'appuie sur la combinaison de deux modèles : Un modèle de segmentation qui se base sur l'algorithme k-means et un modèle LRFM qui décrit le comportement d'achat des clients dans le passé. Cette méthodologie a permis d'arriver à **définir** le concept de **défection partielle** et **totale** dans les cadres non contractuels, en particulier dans l'industrie du commerce électronique.
- Nous nous sommes concentrés sur les modèles de classification **multi-classes** où la variable cible classe un client particulier soit en tant que client partiellement défectueux, totalement défectueux, soit en tant que client poursuivant son comportement d'achat fidèle.
- Nous avons testé et **validé** l'efficacité de notre approche sur une base de données réelle d'un site e-commerce à l'aide d'une **validation croisée** à 10-fold.
- Nous avons **établi** un workflow qui permet d'obtenir un **indicateur d'importance** de différentes variables prédictives utilisées dans la phase d'apprentissage de nos modèles afin d'identifier celles qui sont importantes pour la prédiction de churn dans les contextes e-commerce.

Les résultats de notre approche indiquent la performance prometteuse des modèles proposés dans la prédiction des clients en voie de défection. Une analyse comparative de trois modèles proposés, montre que la performance du modèle DTE (Decision Tree Ensemble) basé sur les méthodes d'ensembles (ensemble de classifieurs) est plus robuste et efficace que les modèles basés sur un seul classifieur (Artificial Neural Networks et Decision Tree). Nous avons aussi conclu que, comparées aux variables classiques RFM (comportementales), les variables qui décrivent les décisions d'achats et les décisions d'abandon de panier sur les sites e-commerce sont les meilleures dans la distinction entre les clients susceptibles d'être partiellement ou totalement défectueux dans un futur proche, et ceux qui sont susceptibles de rester fidèles. Cela contraste les attentes que nous avons formulées à partir des recherches existantes qui soulignent avec force le pouvoir explicatif / prédictif des variables RFM.

## **Limitations et perspectives**

Nos résultats indiquent également certaines limites et problèmes pour d'autres futures recherches et études.

En premier lieu, cette étude est limitée au secteur du commerce électronique. Il est difficile de l'appliquer dans les environnements hors ligne, car elle se repose sur l'analyse de comportement de navigation des clients sur le site e-commerce.

En deuxième lieu, nous n'avons pas utilisé assez de variables dans la phase de clustering. Cependant, d'autres études peuvent utiliser d'autres variables telles que les variables liées à la catégorie de produit, ou celles provenant des réseaux sociaux.

En dernier lieu, le pouvoir prédictif du modèle est significativement influencé par le choix de la technique de classification. Cependant, dans les travaux futurs, d'autres techniques de classification, telles que les algorithmes génétiques, les classifieurs naïf de Bayes, les approches d'ensemble approximatif et la logique floue, seront utilisées.

## Bibliographie

- [1] Hung C, Tsai C. Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. *Expert Syst. Appl.*, 34(1): 780-787, 2008
- [2] Chang EC, Huang SC, Wu HH . Using K-means method and spectral clustering technique in an outfitter's value analysis. *Qual Quant.*, 44(4): 807-815, 2010
- [3] Allenby GM, Arora N, Ginter JL. On the heterogeneity of demand. *J. Mark. Res.*, 35: 384-389, 1998
- [4] Athanassopoulos, A. Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research*, 47(3), 191–207,2000
- [5] Colgate, M., & Danaher, P. Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science*, 28(3), 375–387,2000
- [6] John Hadden, Ashutosh Tiwari, Rajkumar Roy, Dymitr Ruta, Computer assisted customer churn management: State-of-the-art and future trends, *Computers & Operations Research*, Volume 34, Issue 10, Pages 2902-2917, 2007
- [7] Holsapple, C. Singh, M., Electronic Commerce:from a Definitional Taxinomy Towards a Knowledge-Management View. In *journal of Organizational Computing and Electronic Commerce*, 10(3), pp 149-170, 2000.
- [8] J. D. Sommers, "Electronic commerce opportunities on the Internet," , 1995. Available: <http://mweiser.bus.okstate.edu/students/jsommers/t5.htm>
- [9] D. Cook, C. Chung, and C. W. Holsapple, "Information flow first, material flow next," *APICS—The Performance Advantage*, vol. 5, pp. 38–39, Jan, 1995.
- [10] P. Kumar, "MGMT 539—Electronic commerce course home page," Iowa State University, Ames, [cited May 10, 1998]. Available: <http://www.public.iastate.edu/~PremCourses/ecomml.htm>
- [11] L. M. Applegate, C. W. Holsapple, R. Kalakota, F. J. Radermacher, and A. B. Whinston, "Electronic commerce: Building blocks of new business opportunity," *Journal of Organizational Computing and Electronic Commerce*, vol. 6, no. 1, pp. 1–10, 1996.
- [12] N. R. Adam and Y. Yesha, "Electronic commerce: An overview," in *Electronic Commerce*, N. R. Adam and Y. Yesha, Eds. Berlin, Germany: Springer-Verlag, pp. 5–12, 1995
- [13] R. Kalakota and A. B. Whinston, *Frontiers of Electronic Commerce*. Reading, MA: Addison-Wesley, 1996.
- [14] J. D. Becker, T. Farris, and P. Osborn, "Electronic commerce and rapid delivery: The missing 'logistical' link," in *Proc. of the 4th AIS Americas Conf. on Information Systems*, pp. 272–274, 1998.
- [15] Baldwin, R.W. and Chang, C.V., Locking the e-safe, *IEEE Spectrum*, 34, 2, 40–46, 1997.
- [16] McKinsey: Report "Half the World Is Unbanked", 2009, [https://wagner.nyu.edu/files/faculty/publications/Half\\_the\\_world\\_is\\_unbanked.pdf](https://wagner.nyu.edu/files/faculty/publications/Half_the_world_is_unbanked.pdf)
- [17] WorldPay: "The Global Payments Report", 2015, <http://offers.worldpayglobal.com/rs/850-JOA-856/images/GlobalPaymentsReportNov2015.pdf>
- [18] Alain DUCASS, Jean-Marc KWADJANE, « Le commerce électronique en Afrique Maroc, Tunisie, Sénégal et Côte d'Ivoire, Recommandations pour l'intégration régionale en

Méditerranée », 2015.  
[http://www.ipemed.coop/adminIpemed/media/fich\\_article/1460888627\\_ipemed-commerce-en-afriquebd.pdf](http://www.ipemed.coop/adminIpemed/media/fich_article/1460888627_ipemed-commerce-en-afriquebd.pdf)

- [19] Agence nationale de réglementation des télécommunications (ANRT), OBSERVATOIRE DES ABONNEMENTS A INTERNET AU MAROC, Situation au 2ème trimestre 2017, [https://www.anrt.ma/sites/default/files/publications/2017\\_t2\\_tb\\_internet\\_fr.pdf](https://www.anrt.ma/sites/default/files/publications/2017_t2_tb_internet_fr.pdf)
- [20] Centre monétique interbancaire, Activité Monétique Au 30 Juin 2017, [https://www.cmi.co.ma/sites/default/files/activite\\_monetique\\_marocaine\\_au\\_30\\_juin\\_2017\\_4.pdf](https://www.cmi.co.ma/sites/default/files/activite_monetique_marocaine_au_30_juin_2017_4.pdf)
- [21] Chris RYGIELSKI, Jyun-Cheng WANG et David C. YEN : Data mining techniques for customer relationship management. *Technology in Society*, 24(4):483 – 502, 2002.
- [22] Peppers, D., Rogers, M. *Le One-to-One: Valorisez votre Capital-client*. Les Editions d'Organisation, Paris, 1998.
- [23] Büren, A., Schierholz, R., Kolbe, L., Brenner, W. Customer Knowledge Management - Improving Performance of Customer Relationship Management with Knowledge Management. *Proceedings of the 37th Hawaii Conference on System Science (HICSS)*, Big Island, Hawaii, 2004.
- [24] C. X. Ling, T. Chen, Q. Yang, and J. Cheng. Mining Optimal Actions for Profitable CRM. In *Proceedings of 2002 IEEE International Conference on Data Mining*, 2002.
- [25] S. Brown. *Customer Relationship Management, La Gestion de La Relation Client*. PriceWaterHouseCoopers, 2006.
- [26] METAGroup (2001). *Integration: Critical issues for implementing of CRM solutions*. Stamford, CT: META Group Inc.
- [27] Artun, Ö., Levin, D., 2015. *Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data*. John Wiley & Sons, Inc, Hoboken, NJ, USA.
- [28] Provost, F., Fawcett, T.. *Data Science for Business: What you Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, Inc, Sebastopol, CA, USA, 2013,
- [29] Yong Cao, Thomas S. Gruca Reducing Adverse Selection Through Customer Relationship Management. *Journal of Marketing*: October 2005, Vol. 69, No. 4, pp. 219-229, 2005
- [30] Kotler, P. *Marketing management. Analysis, planning, implementation, and control* (8th ed.). Englewood Cliffs, NJ: Prentice-Hall, 1994,
- [31] Hennig-Thurau, T. and Klee, A. The impact of customer satisfaction and relationship quality on customer retention: A critical reassessment and model development. *Psychology and Marketing*, 14: 737–764, 1997
- [32] N. Gordini and V. Veglio, “Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry”, *Industrial Marketing Management*, Volume 62, Pages 100-107, April 2017.
- [33] Gerpott, T. J., Rams, W., and Schindler, A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy*, 25, 4, 249–269, 2001

- [34] Kim, M. K., Park, M. C., and Jeong, D. H. The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy*, 28, 2, 145–159,2004
- [35] Jerry Fjermestad & Nicholas C. Romano, Jr, *Electronic Customer Relationship Management (Advances in Management Information Systems)*, AMIS Vol. 3, ISBN 0-7656-1327-1, ISSN 1554-6152,2006
- [36] Winer, R. A framework for customer relationship management. *California Management Review*, 43(4), 89-105,2001.
- [37] Reichheld, F.F. *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*. Boston: Harvard Business School Press,1996,
- [38] Eugene Sivadas and Jamie L. Baker-Prewitt, An examination of the relationship between service quality, customer satisfaction, and store loyalty, *International Journal of Retail & Distribution Management* Volume 28. pp. 73-82, 2000.
- [39] Yen-Chun Chen, Yung-Cheng Shen & Shuling Liao, An integrated model of customer loyalty: an empirical examination in retailing practice, *The Service Industries Journal*, Volume 29 - Issue 3, 2009.
- [40] Chaffey, Dave. *E-business and E-commerce Management: Strategy, Implementation and Practice*. 4th ed. Harlow, England: FT Prentice Hall, Print, 2009
- [41] Chan, S., & Lam, J. *Customer Relationship Management on Internet and Mobile Channels: An Analytical Framework and Research Directions*. Idea Group Inc, 2009.
- [42] Lee, S., & Shu, W. An integrative and complementarity-based model for the design and adoption of customer relationship management technologies. *Proceedings of the Seventh American Conference on Information Systems*, 854-856, 2001.
- [43] Andaleeb, S. & Anwar, S. Factors influencing customer trust in salespersons in a developing country. *Journal of International Marketing*, 4(4), 35-52, 1996
- [44] Brink, A., Berndt, A. *Customer Relationship Management and Customer Service*. Juta and Co.Ltd, Lansdowne , pp. 120–132, 2004.
- [45] Kotler, P., Keller, K.L., *Marketing management (13th end)*. New Jersey: Pearson Education Inc, Upper Saddle River: NJ, 2009.
- [46] Oliver, R., & Swan, J. Equity and disconfirmation perceptions as influences on merchant product satisfaction. *Journal of Consumer Research*, 16, 372-383, (1989).
- [47] Kotler, P., Armstrong,G., *PrinciplesofMarketing(ThirteenthEdition)*.Pearson education, Upper Saddle River, 2010
- [48] Stenberg, Esa, *Customer value management – Concept creation and implementation – Case wärtsilä Diesel Power plants*. Turku School of Economics and Business Administration, Turku, Finland, 1997.
- [49] Safa, N.S., Ghani, N.A. & Ismail, M.A., ‘An artificial neural network classification approach for improving accuracy of customer identification in e-commerce’, *Malaysian Journal of Computer Science* 27(3), 171–185, 2014,
- [50] Oliver, R.L., *Satisfaction: A Behavioral Perspective on the Consumer*. McGraw Hill, New York, 1997.
- [51] Reicheld, F., *Loyalty Rules! : How to day’s leaders build lasting relationships*. Harvard Business School, Boston. Retrieved, 2001. from <http://scholar.google.com>.

- [52] Segoro, W., Mooring factor, customer satisfaction, and loyalty: a research on cellular providers in West Java, Indonesia. *J.Glob.Bus.Adm.* 4 (1), 60–69, 2012.
- [53] Kotler, P., Armstrong, G., *Principles of Marketing (Thirteenth Edition)*. Pearson education, Upper Saddle River, 2010.
- [54] Godes, D., Mayzlin, D., Using online conversations to study word-of-mouth communication. *Mark.Sci.* 23(4), 545–560, 2004.
- [55] Kotler, P., Keller, K., *Marketing Management (Fourteenth Edition)*. Prentice Hall, New Jersey, 2012.
- [56] Bodet, G., Customer satisfaction and loyalty in service: two concepts, four constructs, several relationships. *J. Retail. Consum. Serv.* 15 (3), 156–162, 2008.
- [57] Tontini, G., Silveira, A., Identification of satisfaction attributes using competitive analysis of the improvement gap. *Int. J. Oper. Prod. Manag.* 27 (5), 482–500, 2007.
- [58] Grace, D., O’Cass, A., An examination of the antecedents of repatronage intentions across different retail store formats. *J. Retail. Consum. Serv.* 12 (4), 227–243, 2005.
- [59] Wang, S.W., Do global airline alliances influence the passenger's purchase decision? *J. Air Transp. Manag.* 37, 53e59, 2014.
- [60] Hussain, R., Nasser, A.A., Hussain, Y.K., Service quality and customer satisfaction of a UAE-based airline: an empirical investigation. *J. Air Transp. Manag.* 42, 167e175, 2015.
- [61] Bolton, R.N., Kannan, P.K., & Bramlett, M. Implications for Loyalty Program Membership and Service Experiences for Customer Retention and Value. *Journal of Academy of Marketing Science*, (28/28), 95-108. 2000.
- [62] Fornell, J.E., & Wernerfelt. *Customer relations & rapport: Professional development series*. Australia: South Western Thomson Learning. 2002.
- [63] R. E. Anderson and S. S. Srinivasan, “satisfaction and e-loyalty: a contingency framework,” *Psychology and Marketing*, vol. 20, no. 2, pp. 123 – 138, 2003.
- [64] Ching-I Teng, Customization, immersion satisfaction, and online gamer loyalty, In *Computers in Human Behavior*, Volume 26, Issue 6, Pages 1547-1554, ISSN 0747-5632, 2010.
- [65] Torsten Lihra, op-cit, p12-13, 2009.
- [66] Allard, K. Graves, L. Gluck, M. May, M. McAteer, S, *Proactive Personalization - Jupiter*, N°8 p 2, ,(1999)
- [67] Anahita Bagherzad Halimi et All, *The Contribution Of Personalization To Customers’ Loyalty Across The Bank Industry In Sweden*, *International Conference on Social Science and Humanity IPEDR vol.5*, Singapore, p383, (2011)
- [68] Pierre Desmet, *Marketing Direct Concepts Et Méthodes* ,3e édition ,dunod,p55 ,(2005)
- [69] Ostrom, A., & Iacobucci, D. Consumer trade-offs and the evaluation of services. *Journal of Marketing*, 59(1), 17–28, 1995.
- [70] Hsin Hsin Chang, Su Wen Chen, The impact of customer interface quality, satisfaction and switching costs on e-loyalty: Internet experience as a moderator, *Computers in Human Behavior*, Volume 24, Issue 6, 17, Pages 2927-2944, 2008.
- [71] Srinivasan, S. S., Anderson, R., & Ponnnavolu, K. Customer loyalty in e-commerce: An exploration of its antecedents and consequences. *Journal of Retailing*, 78(1), 41–50, 2002

- [72] Schafer, B.J.; Konstan, J.; and Riedl, J. Recommender systems in e-commerce. Proceedings of the 1st ACM, Conference on Electronic Commerce, 158–166, 1996
- [73] Reichheld, F. F. and Schefer, P., “E-Loyalty: Your Secret Weapon on the Web,” Harvard Business Review, pp. 105-113, 2000
- [74] Adhi Setyo Santoso, Arlo Erdaka (2015), Customer Loyalty in Collaborative Consumption Model: Empirical Study of CRM for Product-Service System-Based e-Commerce in Indonesia, In *Procedia Computer Science*, Volume 72, Pages 543-551, ISSN 1877-0509, 2015.
- [75] Mohd Azizul Sulaiman, Mohd Amli Abdullah @ Baharum, Arifi Ridzuan, Customer Relationship Management (CRM) Strategies Practices in Malaysia Retailers, *Procedia - Social and Behavioral Sciences*, Volume 130, 15 , Pages 354-361, May 2014.
- [76] Bang Nguyen. The dark side of customer relationship management: Exploring the underlying reasons for pitfalls, exploitation and unfairness, *Journal of Database Marketing & Customer Strategy Management*, 19(1), pp 56–70, 2012.
- [77] Neil Wood, Andrew Green, Michael Starkey. Social CRM as a business strategy, *Journal of Database Marketing & Customer Strategy Management*, 18(1), pp 50–64, 2011.
- [78] Hultén, Bertil, Customer segmentation: The concepts of trust, commitment and relationships, *Journal of Targeting, Measurement and Analysis for Marketing*, 15-4, 1479-1862, 2007
- [79] Thomas. R, Skinner. L. Total Trust and Trust Asymmetry: Does Trust Need to Be Equally Distributed in Interfirm Relationships? *Journal of Relationship Marketing*. 9, pp. 43–53, 2010.
- [80] Sirdeshmukh, D. Singh, J. & Sabol, B. Consumer trust, value, and loyalty in relational exchanges. *Journal of Marketing*. 66 (1), pp. 15-37, 2010.
- [81] Andrews, K.M. and Delahay, B.L. Influences on Knowledge Processes in Organizational Learning: The Psychosocial Filter. *Journal of Management Studies*, 37, 797-810, 2000.
- [82] Morgan, R. M., & Hunt, S. D. The commitment-trust theory of relationship marketing. *Journal of Marketing*. 58 (3), pp. 20–38, 1994.
- [83] Lee, D., & Ahn, J. An exploratory study on the different factors in customer satisfaction with e-commerce between in the United States and in Korea. Proceedings of the Second International Conference on Telecommunication and Electronic Commerce, 1999.
- [84] Hart, P., Saunders, C.S., Emerging electronic partnerships: antecedents and dimensions of EDI use from the supplier’s perspective. *J. Manage. Inform. Syst.* 14 (4), 87–111, 1998.
- [85] De Keyser, A., Lariviere, B., How technical and functional service quality drive consumer happiness: moderating influences of channel usage. *J. Serv. Manag.* 25 (1), 30–48, 2014.
- [86] Park, S.B., Chung, N., Woo, S.C., Do reward programs build loyalty to restaurants? The moderating effect of long-term orientation on the timing and types of rewards. *Manag. Serv. Qual.: Int. J.* 23 (3), 225–244, 2013.
- [87] Sadeghi, M., GhafariCherati, M., Dadkhah, R., YaghoubiBijarboneh, B., Jafari, M., Shahbazi, B., Effect of functional quality, technical quality and mental image on the satisfaction of 3, 4 and 5-star hotels customers in Mashhad City. *J. Ind. Strateg. Manag.* 10 (32), 29–46, 2014.
- [88] Miyazaki, A.D., & Fernandez, A. Consumer perceptions of privacy and security risks for online shopping. *Journal of Consumer Affairs*, 35, 1, 27-44, 2001.

- [89] Kerkhof, P., & Van Noort, G. Third party internet seals: reviewing the effects on online consumer trust. In I. Lee (Ed.), *Encyclopedia of E-Business Development and Management in the Global Economy*, Information Science Reference (pp. 701-708), Hershey PA, (2010).
- [90] Omariba, Z.B., Masese, N.B., & Wanyembi, G. Security and Privacy of Electronic Banking. *International Journal of Computer Science Issues*, 9, 4, 3, 432-446, (2012).
- [91] Berson, S. Smith, and K. Thearling. *Building Data Mining Applications for CRM*. McGraw-Hill Professional, 1999.
- [92] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. "From Data Mining to Knowledge Discovery: An Overview." In *Advances in Knowledge Discovery and Data Mining*, Fayyad et al (Eds.) MIT Press, 1996.
- [93] Y. Chen, D. Pavlov, P. Berkhin, A. Seetharaman, and A. Meltzer. Practical Lessons of Data Mining at Yahoo! In *CIKM '09 : Proceeding of the 18th ACM conference on Information and knowledge management*, pages 1047–1056, New York, NY, USA, 2009.
- [94] Cios, K.J., Pedrycz, W., Swiniarski, R.W., Kurgan, L.A. (2007). "Data Mining: A Knowledge Discovery Approach". *Springer*, 2007.
- [95] El Moukhtar Zemmouri. Représentation et gestion des connaissances dans un processus d'Extraction de Connaissances à partir de Données multi-points de vue. Apprentissage [cs.LG]. Ecole Nationale Supérieure d'Arts et Métiers - Mekkès, 2013. Français. <tel-00940780>
- [96] Shearer, C. "The CRISP-DM model: the new blueprint for data mining". *Journal of Data Warehousing*, Vol. 15(4), pp. 13–19, 2000.
- [97] M.-S. Chen, J. Hun, P. S. Yu, I. T. J, and W. R. Ctr. Data mining : An Overview From a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8 :866–883, 1996.
- [98] M. Lobur, Y. Stekh, A. Kernytssky, and F. Sardieh. Some Trends in Knowledge Discovery and Data Mining. In *Perspective Technologies and Methods in MEMS Design 2008*, number 95-97, 2008.
- [99] Federico Rajola, *Customer Relationship Management: Organizational and Technological Perspectives*, Springer Science & Business Media, 2013.
- [100] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, pages 487-499, Santiago, Chile, 1994.
- [101] Han, J. and Kamber, M. *Data Mining Concepts and Techniques*, 2nd edition, Morgan Kaufmann, 2006.
- [102] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*", Berkeley, University of California Press, 1:281-297, 1967.
- [103] Mitra, S., Pal, S. K., & Mitra, P. Data mining in soft computing framework: A survey. *IEEE Transactions on Neural Networks*, 13, 3–14, 2002.
- [104] Weimin Chen, Guocheng Xiang, Youjin Liu, Kexi Wang, Credit risk Evaluation by hybrid data mining technique, In *Systems Engineering Procedia*, Volume 3, Pages 194-200, 2012.



- [105] Gusti Ngurah Narindra Mandala, Catharina Badra Nawangpalupi, Fransiscus Rian Praktikto, Assessing Credit Risk: An Application of Data Mining in a Rural Bank, In *Procedia Economics and Finance*, Volume 4, Pages 406-412, 2012.
- [106] Rashidah Abdul Rahman, Irda Syahira Khair Anwar, Effectiveness of Fraud Prevention and Detection Techniques in Malaysian Islamic Banks, In *Procedia - Social and Behavioral Sciences*, Volume 145, Pages 97-102, 2014.
- [107] Masoumeh Zareapoor, Porya Shamsolmoali, Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier, In *Procedia Computer Science*, Volume 48, Pages 679-685, 2015.
- [108] C.-J. Cheng, S.W. Chiu, C.-B. Cheng, J.-Y. Wu, Customer lifetime value prediction by a Markov chain based data mining model: Application to an auto repair and maintenance company in Taiwan, In *Scientia Iranica*, Volume 19, Issue 3, Pages 849-855, 2012.
- [109] Mahboubeh Khajvand, Mohammad Jafar Tarokh, Estimating customer future value of different customer segments based on adapted RFM model in retail banking context, In *Procedia Computer Science*, Volume 3, Pages 1327-1332, 2011.
- [110] Manpreet Kaur, Shivani Kang, Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining, In *Procedia Computer Science*, Volume 85, Pages 78-85, 2016.
- [111] Tao Huang, Robert Fildes, Didier Soopramanien, The value of competitive information in forecasting FMCG retail product sales and the variable selection problem, In *European Journal of Operational Research*, Volume 237, Issue 2, Pages 738-748, 2014.
- [112] Mirko Böttcher, Martin Spott, Detlef Nauck, Rudolf Kruse, Mining changing customer segments in dynamic markets, In *Expert Systems with Applications*, Volume 36, Issue 1, Pages 155-164, 2009.
- [113] Su-Yeon, K., Tae-Soo, J., Eui-Ho, S. and Hyun-Seok, H. ‘Customer segmentation and strategy development based: a case study’, *Expert Systems with Applications*, Vol. 31, No. 1, pp.101–107, 2006.
- [114] Wei, C. P., Chang, H. T., and Lee, Y. H. Telecommunication Data Mining for Target Marketing. *Journal of computer*, 12(4), 2000.
- [115] Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, U. Abbasi, Improved churn prediction in telecommunication industry using data mining techniques, In *Applied Soft Computing*, Volume 24, Pages 994-1012, 2014.
- [116] Kyoungok Kim, Chi-Hyuk Jun, Jaewook Lee, “Improved churn prediction in telecommunication industry by analyzing a large network”, In *Expert Systems with Applications*, Volume 41, Issue 15, Pages 6575-6584, 2014.
- [117] Ying Huang, Tahar Kechadi, An effective hybrid learning system for telecommunication churn prediction, In *Expert Systems with Applications*, Volume 40, Issue 14, Pages 5635-5647, 2013.
- [118] Ayman E.Khedra et al, “Proposed Framework for Implementing Data Mining Techniques to Enhance Decisions in Agriculture Sector Applied Case on Food Security Information Center Ministry of Agriculture, Egypt”, *International Conference on Communication, Management and Information Technology (ICCMIT)* ,Volume 65, 2015, Pages 633-642, 2015.

- [119] Raorane A. and Kulkarni F. Data Mining: An effective tool for yield estimation in the agricultural sector, *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, Volume 1, Issue 2, pp.1, 2, 4, 2012.
- [120] Wang Y., Tang J. and Cao W. Grey prediction model-based food security early warning prediction, School of Business, Jiangnan University, Wuxi, China, *Grey Systems: Theory and Application* Vol. 2 No. 1, Emerald Group Publishing Limited, 2012, pp. 2, 2012.
- [121] Anand Shanker Tewari, Asim Gopal Barman, Sequencing of items in personalised recommendations using multiple recommendation techniques, *Expert Systems with Applications*, 2017.
- [122] Özden Gür Ali, Umut Arıtürk, Dynamic churn prediction framework with more effective use of rare event data: The case of private banking, In *Expert Systems with Applications*, Volume 41, Issue 17, Pages 7889-7903, 2014.
- [123] Niels Holtrop, Jaap E. Wieringa, Maarten J. Gijzenberg, Peter C. Verhoef, No future without the past? Predicting churn in the face of customer privacy, In *International Journal of Research in Marketing*, Volume 34, Issue 1, Pages 154-172, 2017.
- [124] Bingquan, et al., “Customer churn prediction in telecommunications”, *Expert Systems with Applications*, Volume 39, Issue 1, Pages 1414-1425, 2012.
- [125] Faris, H., “Neighborhood cleaning rules and particle swarm optimization for predicting customer churn behavior in telecom industry”. *Int. J. Adv. Sci. Technol.* 68, 11–22, 2014.
- [126] Burez, J., and Van den Poel, D., “Crm at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services”. *Expert Systems with Applications*, 32, 277–288, 2007.
- [127] Coussement, K., & Van den Poel, D. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327, 2008.
- [128] V.L. Miguéis, et al., “Modeling partial customer churn: On the value of first product-category purchase sequences”, *Expert Systems with Applications*, Volume 39, Issue 12, 15, Pages 11250-11256, 2012.
- [129] Ssu-Han Chen, “The gamma CUSUM chart method for online customer churn prediction”, *Electronic Commerce Research and Applications*, Volume 17, Pages 99-111, 2016.
- [130] Effendy V, Baizal ZA. Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest. In: 2014 2nd International Conference on Information and Communication Technology (ICoICT), IEEE; .p. 325–30, 2014.
- [131] M. Clemente-Císcar, et al., “A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings”, *European Journal of Operational Research*, Volume 239, Issue 1, 2014.
- [132] Shin-Yuan Hung, et al., “Applying data mining to telecom churn management”, *Expert Systems with Applications*, 31, pp515–524, 2006.
- [133] Jae-Hyeon Ahna, Sang-Pil Hana, Yung-Seop Lee., “Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry”, *Telecommunications Policy*, Volume 30, Issues 10–11, Pages 552–568, 2006

- [134] Xie , et al., “Customer churn prediction using improved balanced random forests”, *Expert Systems with Applications*, 36(3), pp.5445-5449, 2009.
- [135] K. Coussement and K. W. De Bock, “Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning”, *Journal of Business Research*, 66, 1629–1636, 2013.
- [136] M.A.H. Farquad, Vadlamani Ravi, S. Bapi Raju, “Churn prediction using comprehensible support vector machine: An analytical CRM application”, *Applied Soft Computing*, Volume 19, Pages 31-40, 2014
- [137] R. Ait daoud, et al., “Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco”, *World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering*, Vol:9, No:8, 2015.
- [138] Xiaobing Yu, et al., “An extended support vector machine forecasting framework for customer churn in e-commerce”, *Expert Systems with Applications*, vol.38, pp.1425–1430, 2011.
- [139] Reinartz W, Kumar V. The mismanagement of customer loyalty. *Harvard business review*. 80(7):86–95, 2002.
- [140] Yu-Xiang Yen and Der Juinn Horng, “Effects of satisfaction, trust and alternative attractiveness on switching intentions in industrial customers”, *International Journal of Management and Enterprise Development*, 2010.
- [141] Wathne, K.H., Biong, H. and Heide, J.B., "Choice of Supplier in Embedded Markets: Relationship and Marketing Program Effects", *Journal of Marketing*, Vol.65, pp.54-66, 2001.
- [142] Anderson, et al., “Customer satisfaction, market share, and profitability: Findings from Sweden”. *Journal of Marketing*, 58 (3), 53–66, 1994.
- [143] Morgan, et al., “The commitment-trust theory of relationship marketing”. *Journal of Marketing*, 58 (3), 20–38, 1994.
- [144] Peelen, et al., “Direct marketing for establishing the relationships between buyers and sellers”, *Journal of Direct Marketing*, 3 (1), 7–14, 1989.
- [145] Alshurideh, M., Masa'deh, R., & Alkurdi, B. The Effect of Customer Satisfaction Upon Customer Retention in the Jordanian Mobile Market: An Empirical Investigation. *European Journal of Economics, Finance and Administrative Sciences*, 47, 69-78, 2012.
- [146] Danesh, S., Nasab, S., & Ling, K. The Study of Customer Satisfaction, Customer Trust and Switching Barriers on Customer Retention in Malaysia Hypermarkets. *International Journal of Business and Management*, 7(7), 141-150, 2012.
- [147] Rao, S., Goldsby, T., Griffis, S., & Iyengar, D. Electronic Logistics Service Quality (e-LSQ): Its Impact on the Customer’s Purchase Satisfaction and Retention. *Journal of Business Logistics*, 32(2), 167-179, 2011.
- [148] Omotayo, O., & Joachim, A. A. Customer Service in the Retention of Mobile Phone Users in Nigeria. *African Journal of Business Management*, 2(2), 26-31, (2008).
- [149] Newell, F., “Loyalty. Com: Customer relationship management in the new era of internet marketing”, McGraw- Hill, New York, NK, 2000.
- [150] Reichheld, F.F., Learning from customer defections. *Harvard Business Review* 74 (2), 56–69, 1996.

- [151] Rust, R.T. and Zahorik, A.J., “Customer satisfaction, customer retention, and market share”, *Journal of Retailing*, 69 (2), 193–215, 1993.
- [152] Mozer, et al., “Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry”, *IEEE Transactions on Neural Networks*, 11 (3), 690–696, 2000.
- [153] Reichheld FF & Sasser WE, *Zero Defections: Quality Comes to Services*, Harvard Business Review, September-October, 105-111, 1990.
- [154] Srinivasan, S. S., Anderson, R., and Ponnalu, K. “Customer Loyalty in E-Commerce: An Exploration of its Antecedents and Consequences,” *Journal of Retailing* (78:1), pp. 41-50, 2002.
- [155] Venkatesh, et al., “Customer satisfaction and loyalty in online and offline environments”, *Intern. J. of Research in Marketing*, 20, 153 –175, 2003.
- [156] Jonathan Burez, Dirk Van den Poel, CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services, In *Expert Systems with Applications*, Volume 32, Issue 2, Pages 277-288, 2007.
- [157] Buckinx, W. and Van den Poel, D., “Customer base analysis: partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting”, *European Journal of Operational Research*, 164(1), 252–268, 2005.
- [158] R. Ait daoud, et al., “Combining RFM model and clustering techniques for customer value analysis of a company selling online”. *IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, 1-6, 2015.
- [159] H. H. Wu, et al., “Analyzing Patients’ Values by Applying Cluster Analysis and LRFM Model in a Pediatric Dental Clinic in Taiwan”, Hindawi Publishing Corporation *The Scientific World Journal*, Vol 2014, Article ID 685495, 7 pages, 2014.
- [160] M. Hughes, “Strategic database marketing”, Probus Publishing, 1994
- [161] Jo-Ting Wei, et al., “Customer relationship management in the hairdressing industry: An application of data mining techniques”, *Expert Systems with Applications*, Volume 40, Issue 18, Pages 7513-7518, 2013.
- [162] Der-Chiang Li, et al., “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business”, *Expert Systems with Applications*, Volume 38, Issue 6, Pages 7186-7191, 2011.
- [163] M. Hughes, “Boosting response with RFM”. *Marketing Tools*, Vol.3, No.3, pp. 4-7, 1996.
- [164] X. Yang, “How to develop new approaches to RFM segmentation”, *Journal of Targeting, Measurement and Analysis for Marketing*, Vol.13, No.1, pp. 50-60, 2004.
- [165] S. M. S. Hosseini, et al., “Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty”, *Journal of Expert Systems with Applications*, Vol.37, No.7, pp. 5259–5264, 2010.
- [166] C. Yeh, et al., “Knowledge discovery on RFM model using Bernoulli sequence”, *Expert Systems with Applications*, Vol.36, No.3, pp. 5866–5871, 2009.
- [167] H. C. Chang and H. P. Tsai, “Group RFM analysis as a novel framework to discover better customer consumption behavior”, *Expert Systems with Applications*, Vol.38, No.12, pp.14499–14513, 2011.

- [168] H. H. Chang and S. F. Tsay., “Integrating of SOM and K-mean in data mining clustering: an empirical study of CRM and profitability evaluation”, *Journal of Information Management*, vol. 11, no. 4, pp. 161–203, 2004.
- [169] S. Chow. and R. Holden., “Toward an understanding of loyalty: The moderating role of trust”, *Journal of Management issues*, Vol.9, No.3, pp. 275–298, 1997
- [170] D. Birant, “Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining”, ISBN: 978-953-307-154-1, 2011.
- [171] R. J. Kuo, et al., “Integration of self-organizing feature map and K-means algorithm for market segmentation”, *Computers & Operations Research*, Vol.29, No.11, pp. 1475-1493, 2002.
- [172] Tan P.N., Steinbach M., Kumar V. “Introduction to Data Mining”, Pearson Addison Wesley; Boston, MA, pp. 487–556 ,USA: 2006.
- [173] Jiawei Han and Micheline Kamber, “Data Mining: Concepts and Techniques, 2nd ed”, Morgan Kaufmann Publishers, ISBN 1-55860-901-6, 2006.
- [174] Peter J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, *Journal of Computational and Applied Mathematics*, Volume 20, Pages 53-65, 1987.
- [175] Jo-Ting Wei, Shih-Yen Lin, Chih-Chien Weng, Hsin-Hung Wu, A case study of applying LRFM model in market segmentation of a children’s dental clinic, *Expert Systems with Applications*, Volume 39, Issue 5, Pages 5529-5533, 2012.
- [176] Kristof Coussement, Filip A.M. Van den Bossche, Koen W. De Bock, “Data accuracy’s impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees”, *Journal of Business Research*, Volume 67, Issue 1, Pages 2751-2758, 2014.
- [177] Sunil S. Morade, Suprava Patnaik, Comparison of classifiers for lip reading with CUAVE and TULIPS database, *Optik - International Journal for Light and Electron Optics*, Volume 126, Issue 24, Pages 5753-5761, 2015.
- [178] Zhang, R. “A Support Vector Machine (SVM) Approach to Netlift Modeling”. *Joint Statistical Meetings proceedings, Section on Statistics in Marketing*, pp. 3336 - 3340. American Statistical Association, 2012.
- [179] Cheng-Hsiung Weng, Revenue prediction by mining frequent itemsets with customer analysis, *Engineering Applications of Artificial Intelligence*, Volume 63, Pages 85-97, 2017.
- [180] S. H. Ha and S. C. Park., “Application of data mining tools to hotel data mart on the Intranet for database marketing”, *Expert Systems with Applications*, vol. 15, no. 1, pp. 1–31, 1998.
- [181] Y. L.Chen, et al., “Constructing a multivalued and multi-labeled decision tree”, *Expert Systems with Applications*, 25(2), 199–209, 2003.
- [182] Wei, C. -P., & Chiu, I. -T., “Turning telecommunications call details to churn prediction: A data mining approach”. *Expert Systems with Applications*, 23(2), 103–112, 2002.
- [183] Rokach, L., Maimon, O., *Data Mining with Decision Trees: Theory and Applications*, World Scientific Publishing, 2008.
- [184] Dietterich, T. G. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000a.
- [185] L. Breiman, “Random forests”, *Machine Learning*, 45(1), 5–32, 2001.

- [186] P. M. Guadagni, and J.D.C. Little., “A Logit Model of Brand Choice Calibrated on Scanner Data”, *Marketing Science*, 2(3), pp 203-238, 1983.
- [187] P, Rossi et al., On the value of household purchase history information in target marketing, *Marketing Science*, Vol. 15, No. 4, pp. 321-340, 1996.
- [188] T. Jahromi, et al., “Managing B2B customer churn, retention and profitability”, *Industrial Marketing Management*, Volume 43, Issue 7, Pages 1258-1268, 2014.
- [189] M. Sokolova and G. Lapalme., “A systematic analysis of performance measures for classification tasks”, *Information Processing & Management*, Volume 45, Issue 4, Pages 427-437, 2009.
- [190] Ligang Zhou, et al., “One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies”, *Information Fusion*, Volume 36, Pages 80-89, 2017.
- [191] A Sturm and Bob L, “Classification accuracy is not enough”, *Journal of Intelligent Information Systems*, Volume 41, Issue 3, December 2013, pp 371–406,2013.
- [192] Keaveney, S. M., & Parthasarathy, M. Customer switching behavior in online services: An exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, 29(4), 374–390, 2001.