



*République Algérienne Démocratique et Populaire Ministère de l'Enseignement Supérieur et de la
Recherche Scientifique*

Université Mohamed Khider – BISKRA

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

Département d'informatique

N°d'ordre : SIOD19/M2/2021

Mémoire

Présenté pour obtenir le diplôme de Master Académique en

Informatique

Parcours: Système d'Informatique Optimisation et Décision (SIOD)

Titre : Extraction des règles d'associations pour l'analyse des paniers d'achats

*Par : **BOUZAHER Fatiha***

Soutenu le : ../...../2021 devant le jury composé de :

- HOADJLI Hadia MAA Rapporteur*
- Examineur*
- Président*

Année Universitaire 2020- 2021

Dédicaces

Pour Ma Mère : Autant de phrases aussi expressives soient-elles ne sauraient montrer le degré d'amour et d'affection que j'éprouve pour toi. Tu m'as comblé avec ta tendresse et affection tout au long de mon parcours. Puisse le tout puissant te donner santé, bonheur et longue vie afin que je puisse te combler à mon tour.

*Autant de phrases et d'expressions aussi éloquentes soit-elles ne sauraient exprimer ma gratitude et ma reconnaissance. Tu as su m'inculquer le sens de la responsabilité, de l'optimisme et de la confiance en soi face aux difficultés de la vie. Tes conseils ont toujours guidé mes pas vers la réussite. Ta patience sans fin, ta compréhension et ton encouragement sont pour moi le soutien indispensable que tu as toujours su m'apporter. Je te dois ce que je suis aujourd'hui et ce que je serai demain et je ferai toujours de mon mieux pour rester ta fierté et ne jamais te décevoir. Merci de me rappeler aussi souvent que nécessaire que rencontrer des difficultés en tant que femme, en tant que fille, en tant que mère, en tant qu'amie et en tant qu'employée est normal et qu'on vit tous des moments plus difficiles à surmonter; mais j'ai toujours la chance de t'avoir à mes côtés quand vient le temps de les affronter. Pour Toi Ma Très Chère Grande Sœur **HASSINA***

Pour mes enfants qui m'ont fait de moi la plus fière des mamans.

Une maman pleine d'amour, une maman épanouie.

Vous avez été ma raison de vie ...

Vous êtes ce que la vie m'a donné de plus joli.

Vous êtes le sens de mon existence.

La raison pour laquelle je suis née,

Ma raison d'être en ce bas-monde.

Mes enfants Ma vie !

A Mes Sœurs et Frères ainsi pour que ses Enfants pour leurs aides et soutient.

FATIHA

Remerciements

*Je remercie tout d'abord **DEU TOUT PUISSANT** de m'avoir donné le courage, la force et la patience pour achever ce modeste travail.*

Je remercie Mme Hadia Hoadjli, ma superviseure de cette thèse, pour ses encouragements et pour son aide qu'elle m'a apporté pendant tout ce travail.

Je tiens ensuite à remercier chacun des membres du jury pour nous avoir fait l'honneur d'accepter d'évaluer mon travail.

Mes remerciements et mes gratitudes vont aux enseignants du département informatique.

Mes remerciements également toute personne ayant contribué à notre éducation et notre formation.

Fatiha BOUZAHER

Liste Des Matières

<i>Liste des Référence</i>	10
<i>Résumé</i>	14
1 CH 01 INTRODUCTION	18
2 <i>Définition de la gestion de relation client</i>	18
3 <i>Historique</i>	20
4 <i>Les Trois Dimensions Du Customer Relationship Management</i>	21
4.1 <i>La GRC analytique</i>	21
4.2 <i>La GRC opérationnelle</i>	21
4.3 <i>La GRC collaborative</i>	22
5 LE DATA WAREHOUSE	23
5.1 <i>Définition</i>	23
6 <i>Fonctions de la GRC</i>	25
6.1 <i>Connaître le client</i>	25
6.2 <i>Choisir son client</i>	25
6.3 <i>Conquérir de nouveaux clients</i>	25
6.4 <i>Fidéliser les meilleurs clients</i>	26
7 <i>L'importance des technologies et les logiciels dans la GRC</i>	27
8 <i>Les approches de la gestion de la relation client :</i>	27
9 <i>Les Enjeux D'une GRC</i>	28
9.1 <i>L'intégration multi canal</i>	28
9.2 <i>L'accroissement de la productivité</i>	29
9.3 <i>L'accroissement de la fidélité des clients</i>	29
10 <i>Le processus de la gestion de la relation client</i>	29
10.1 <i>Identifier</i>	29
10.2 <i>Segmenter</i>	29
10.3 <i>Adapter</i>	30
10.4 <i>Echanger</i>	30
10.5 <i>Evaluer</i>	30
11 <i>Les politiques de la relation client</i>	31
11.1 <i>Politique de (re)conquête</i>	31
11.2 <i>Politique d'abandon</i>	31
11.3 <i>Politique de fidélisation</i>	31
11.4 <i>Politique de rationalisation</i>	31

12	<i>Avantages d'un système CRM au sein d'une entreprise :</i>	31
12.1	<i>Offre une meilleure connaissance des clients</i>	31
12.2	<i>Augmente la satisfaction des clients et leur fidélité</i>	32
12.3	<i>Améliore la gestion des contacts et l'accès à leurs informations</i>	32
12.4	<i>Améliore la communication et la collaboration entre les salariés</i>	33
12.5	<i>Fournit une certaine quantité de rapports et tableaux de bord intéressants</i>	33
12.6	<i>Accroît la productivité et le revenu de l'entreprise</i>	33
12.7	<i>Augmente l'efficacité du marketing</i>	33
12.8	<i>Améliore la qualité des services et produits octroyés aux clients</i>	33
12.9	<i>Améliore la gestion des ventes</i>	34
12.10	<i>Permet des prévisions de ventes précises</i>	34
12.11	<i>Offre une synchronisation avec les autres outils utilisés par l'entreprise</i>	34
13	<i>Inconvénients d'un système CRM au sein d'une entreprise :</i>	34
13.1	<i>Demande de former les employés à l'utilisation du CRM</i>	34
13.2	<i>Entraîne des coûts</i>	35
13.3	<i>Ralentit légèrement le processus</i>	35
13.4	<i>Donne aux salariés le sentiment d'être surveillés</i>	35
13.5	<i>Peut transformer l'entreprise en « usine à gaz » et éliminer l'élément humain</i>	35
13.6	<i>Présente des limites à son utilisation</i>	35
14	<i>Quels sont les bénéfices pour une entreprise d'avoir un CRM ?</i>	36
15	CONCLUSION	37
1	CH 02 <i>Introduction :</i>	39
2	<i>Le Data Mining</i>	40
3	<i>Les tâches du Data Mining :</i>	41
3.1	<i>La classification:</i>	41
3.2	<i>L'estimation :</i>	41
3.3	<i>La prédiction :</i>	41
3.4	<i>La Segmentation :</i>	41
3.5	<i>La description :</i>	41
3.6	<i>L'optimisation :</i>	41
4	<i>Les techniques de Data Mining :</i>	42
4.1	<i>Les arbres de décision :</i>	42
4.2	<i>Les réseaux de neurones :</i>	42
4.3	<i>Les algorithmes génétiques :</i>	42
4.4	<i>Les règles d'association :</i>	42

4.5	<i>Les plus proches voisins (CBR, Case Based Reasoning) :</i>	42
4.6	<i>Le Clustering :</i>	43
5	<i>Les Règles d'Association</i>	43
5.1	<i>Définition des règles d'association</i>	44
5.2	<i>Notions de base sur les règles d'association</i>	45
5.2.1	<i>Représentation des données</i>	45
5.2.2	<i>Item et support</i>	46
5.2.3	<i>Support d'un Itemset</i>	46
5.2.4	<i>Itemset Fréquent</i>	47
5.3	<i>Mesures de qualité de Règles d'association</i>	47
5.3.1	<i>Définition</i>	47
5.3.2	<i>Support et confiance d'une règle d'association</i>	47
5.3.2.1	<i>Support d'une règle d'association</i>	48
5.3.2.2	<i>Confiance d'une règle d'association :</i>	48
5.4	<i>Extraction des règles d'association</i>	49
5.4.1	<i>Préparation des données :</i>	49
5.4.2	<i>Extraction des ensembles fréquents d'attributs :</i>	49
5.4.3	<i>Génération des règles d'association :</i>	49
5.4.4	<i>Visualisation et interprétation :</i>	49
6	<i>Algorithme D'extraction Des Règles D'association</i>	49
6.1	<i>Algorithme Apriori</i>	50
6.2	<i>Exemple de treillis des itemsets</i>	50
6.3	<i>Fonctionnement de l'algorithme Apriori :</i>	51
6.4	<i>Avantages de l'Algorithme Apriori</i>	52
6.5	<i>Inconvénients de l'Algorithme Apriori</i>	52
6.6	<i>OPTIMISATIONS</i>	52
6.6.1	<i>L'algorithme AprioriTID 45</i>	52
6.6.2	<i>L'algorithme apriori partitionné</i>	53
6.6.3	<i>Algorithme de comptage dynamique</i>	53
6.6.4	<i>Algorithme FP-growth</i>	54
6.6.5	<i>Avantages et inconvénients de l'algorithme FP-growth</i>	54
7	<i>Conclusion</i>	55
1	CH 03 <i>Introduction :</i>	58
2	<i>Description de notre travail</i>	58
3	<i>Conception globale du système</i>	59

4	Architecture détaillée	61
4.1	Prétraitement :	61
4.2	Sélectionner l'algorithme.....	62
4.2.1	Etapes Pour Effectuer Algorithme Apriori :	62
4.2.2	Etapes pour effectuer Algorithme Apriori TID :.....	62
5	Spécification fonctionnelle.....	63
5.1.1	Détermination des cas d'utilisation	64
	Description des cas d'utilisation	64
5.1.2	Diagramme de séquence de cas d'utilisation « Extraction ».....	65
5.1.3	Diagramme de séquence de cas d'utilisation « Extraction ».....	66
6	Conclusion	67
1	CH 04 Introduction.....	69
2	Outils utilisés.....	69
2.1	Logiciels utilisés	69
	Ntbeans	69
2.1.1	Scence builder	70
	Java.....	70
	JavaFX.....	71
3	Base de donné utilisé.....	71
4	Description de l'application.....	72
4.1	La première interface.....	72
4.2	L'interface Apriori	73
4.2.1	L'interface d'Apriori-TID.....	74
5	Description de code	75
6	Description de code détaillé	76
6.1	Les interfaces	76
6.1.1	Les bibliothèques.....	76
6.2	Les Fonction d'action des buttons.....	78
7	Conclusion	86

Liste des tableaux

Tableau 1	Base de données binaire [50].....	45
-----------	-----------------------------------	----

<i>Tableau 5 Exemple de treillis d'Itemsets.....</i>	<i>51</i>
<i>Tableau 18 Tableau des classes avec ses Signification.....</i>	<i>76</i>

Liste des Equations

<i>Équation 1 Support d'un d'un Itemset.....</i>	<i>47</i>
<i>Équation 2 Support d'une règle d'association [62].....</i>	<i>48</i>
<i>Équation 3 Confiance d'une règle d'association [64].....</i>	<i>49</i>

Liste des figures

<i>Figure 1 La Gestion Relation Client.....</i>	<i>20</i>
<i>Figure 2 Les Trois Dimensions Du Customer Relationship Management.....</i>	<i>23</i>

Figure 3 LE DATA WAREHOUSE.....	24
Figure 4 Fonctions de la GRC.....	26
Figure 5 Etapes du Processus de Génération des Règles d'Association.....	44
Figure 7 Structure General Du Systeme.....	61
Figure 10 Diagramme des cas d'utilisation.....	64
Figure 11 Diagramme de Cas d'utilisation de système.....	64
Figure 12 Diagramme de séquence de cas d'utilisation « Extraction ».....	65
Figure 13 Diagramme de séquence de cas d'utilisation « Résultats ».....	66
Figure 14 Interface Netbeans.....	69
Figure 16 Interface de SceneBuilder.....	70
Figure 17 Interface de Java.....	71
Figure 18 interface de JavaFx.....	71
Figure 19 Base de Données.csv.....	72
Figure 20 Interface d'exécution.....	73
Figure 21 L'interface d'exécution de l'algorithme Apriori.....	73
Figure 22 L'interface d'exécution de l'algorithme AprioriTID.....	74
Figure 23 Les bibliothèques de javafx utilisé dans l'interface JavaFx.....	77
Figure 24 Interface d'exécution.....	77
Figure 25 Les bibliothèques de javafx et java utilise dans les classes java.....	78
Figure 26 Fonction pour ouvrir Apriori interface et fermé l'ancienne interface.....	78
Figure 27 Fonction d'ouvrir le fichier csv.....	79
Figure 28 Fonction d'exécution l'algorithme Apriori et afficher les résultats.....	79
Figure 29 Fonction d'exécution de l'algorithme AprioriTid et afficher les résultats.....	80
Figure 30 Base de données transactionnelle.....	81
Figure 31 Base de données transactionnelle.....	81
Figure 32 L'algorithme trie la base de donnée.....	82
Figure 33 Calcul le support de chaque Itemset.....	83
Figure 34 Affichage des règles solides.....	84
Figure 35 Génération d'un nouvel ensemble d'itemsets candidats(taille1+tille 2).....	85
Figure 36 Extraction des règles solides.....	86

Liste des Référence

- Achouri Abdelghani,. "Extraction de relations d'association maximales dans les textes : représentation graphique", Université du Québec a Trois-Rivières, 2012.

- Agrawal, T. Imieliński and A. Swami. "Mining association rules between sets of items in large databases". In *Proceeding SIGMOD '93 Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207-216, New York, NY, USA, 1993.
- amarque. Eric, «Management de la banque », Edition, Pearson Education France, paris, 2005, pp138-140.
- Bahri Emna, Lallich Stéphane. "Introduction de l'élagage pour l'extraction de règles d'association de classe sans génération de candidats", Laboratoire ÉRIC, Université de Lyon., QDC 2009.
- Benrhaiem, Morched. "La lecture assistée par ordinateur : une étude exploratoire", Université du Québec à Trois-Rivières, 2015.
- BERRY (M.J.A.), LINOFF (G.). "Data Mining-Techniques appliquées au marketing, à la vente et aux services clients" - 1996 MASSON (Wiley)
- Blanchard Julien, Kuntz Pascale, Guillet Fabrice, Gras Régis. "Mesure de la qualité des règles d'association par l'intensité d'implication entropique", IRIN – École polytechnique de l'université de Nantes, 2002
- Brown, 2000
- Brown, 2000, pp. 253- 255
- Buttle, 2009
- Buttle, 2009
- Buttle, 2009
- Buttle, 2009, p. 366
- Chou et al., 2002, p. 444
- Conservatoire National des Arts et métier de Lille "Data warehouse et Data Mining". version 1.1, le 15 juin 1998, disponible sur : www.imi-utc.com/ptemoign/index.html
- Cours DJEFFAL Abdelhamid 2020-2021
- Cours DJEFFAL Abdelhamid 2020-2021
- DARREL Rigby- REICHFELD Frederick- SCHEFFER Phil. Aviod the Four Perils of CRM, *Harvard Business Review*, février 2002, p.106.
- Demeure. Claude, *Op.cit.* p.363.
- G. Gardarin. *Internet/Intranet et bases de données.* Edition EYROLLES, 1999.
- Grenier, C. et Moine, C., 2003
- <https://docs.oracle.com/javafx/2/overview/jfxpub-overview.htm>

- <https://fr.wikipedia.org/wiki/NetBeans>
- <https://github.com/gluonhq/scenebuilder/wiki>
- <https://www.futurasciences.com/tech/definitions/internet-java-485/>
- <https://www.theserverside.com/definition/NetBeans>
- Inmon, 1996
- Jallat et al., 2018, pp. 121-125
- Jallat et al., 2018, pp. 151-152
- Léfber. René, Venturi. Gilles, *Op.cit.* p.03.
- LEFEBURE - GILLES. *Op.cit.* p. 53
- Lefébure R, venturi G, *Op.cit.* .pp 9-12.
- Lucron, 2018
- Lucron, 2018
- MIRATON Lionel, « introduction à la gestion de la relation client »P.6.
- Mithas, Krishnan & Fornell, 2005
- Nguyen, Sherif & Newby, 2007, p. 103
- Peelen, 2005
- Piatetsky-Shapiro (G.), Frawely (W.J.). "Knowledge Discovery in Databases". AAA Press, 1991.
- PIERRE (A) et PIERRE (A) - Guggémos, « CRM et les clés de réussite », Edition, Organisation, Paris, 2005, P (9-11)
- Plasse Marie, Ndeye Niang, Saporta Gilbert, Villeminot Alexandre, Leblon Laurent. "Méthodes de classification pour l'extraction de règles",CNAM Laboratoire CÉDRIC, 2008.
- R. Agrawal, T. Imielinski, and A. Swami. Mining Association Rules between sets of items in large Databases. *ACM SIGMOD Records*, pages 207—216, 1993.
- R. Agrawal and R. Skirant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Databases*, pages 478—499, June 1994.
- R. Agrawal and R. Skirant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Databases*, pages 478—499, June 1994.

- *R. Agrawal et R. Srikant. "Fast algorithms for mining association rules in large databases". In VLDB'94 : Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487–499, San Francisco, CA, USA, 1994*
- *R. Rakotomalala. "Les règles d'Association, market data analysis ou L'analyse du panier de la ménagère"*
- *Shashikumar G. Totad, Geeta R. B, Prasad Reedy. "Batch Processing for Incremental FP-tree Construction", international Journal of Computer Applications, 2010*
- *Site web de Indice RH, 15.05.2019*
- *Site web de Indice RH, 15.05.2019*
- *Site Web de Noeo, 12.05.2019*
- *Site Web de targetintegration, 22.05.2019*
- *Site Web de Teamleader, 24.05.2019*
- *Tihi, Adil. "Mise en place d'un projet de data mining pour identifier les règles d'association", HEC Montréal, 2007.*
- *www. Djeffal abd alhamid.com*
- *www. Djeffal abd alhamid.com*
- *www.jmdoudoux.fr/java/dej/chap-presentation.htm*

Résumé

L'objectif des techniques des règles d'association consiste à détecter des relations ou des associations entre des modalités spécifiques de variables catégorielles dans des fichiers de données volumineux.

On s'intéresse aux deux algorithmes *Apriori* et *AprioriTid*, en utilisant des "seuils" prédéfinis ce qui nous permet de conclure les règles de corrélation entre eux pour l'aide à la décision.

Mots Clés: règles d'association, algorithme A-priori, A-prioriTid, génération des règles d'association .

ملخص

الهدف من تقنيات قواعد الارتباط هو اكتشاف العلاقات أو الارتباطات بين فئات محددة من المتغيرات الفئوية في ملفات البيانات الكبيرة.

نحن مهتمون بالخوارزميتين *Apriori* و *AprioriTid*، باستخدام "عتبات" محددة مسبقًا والتي تسمح لنا بإبرام قواعد الارتباط بينهما لدعم القرار.

الكلمات المفتاحية: قواعد الارتباط، خوارزمية *A-prefi*، *A-premiTid*، إنشاء قواعد الارتباط

Abstract

The goal of association rule techniques is to detect relationships or associations between specific categories of categorical variables in large data files.

We are interested in the two algorithms *Apriori* and *AprioriTid*, using predefined "thresholds" which allows us to conclude the correlation rules between them for decision support.

Keywords: association rules, A-priori algorithm, A-prioriTid, generation of association rules

Introduction General

Introduction General

Dans un environnement de plus en plus concurrentiel et face à l'individualisation de la Consommation, les entreprises sont obligées, aujourd'hui, non plus de réagir au marché mais bien de l'anticiper. Les entreprises ont pour principal objectif de cibler le mieux possible leurs clients afin de répondre à leurs attentes. Le Customer Relationship Management (CRM) devient alors une stratégie essentielle pour l'entreprise. En effet, face à la volatilité de la clientèle, il n'existe qu'une solution : recentrer l'entreprise pour mieux répondre aux besoins des clients. Ceci est l'objectif du CRM, qui place le client au centre de ses préoccupations, anticipe ses besoins et met en place avec lui un rapport durable.

Le CRM qui, grâce à l'informatique, tire parti de toutes les informations disponibles sur le client pour lui proposer de nouveaux services et produits vraiment adaptés à ses besoins, est une technique qui a connu un développement très important ces dernières années. L'intensification de la concurrence et la volatilité de la clientèle accompagnée du développement des systèmes d'informations ont permis l'émergence du Data Mining dans les entreprises.

Le datamining se base sur plusieurs techniques pour l'extraction de connaissances à partir des données, Ces connaissances ont un grand intérêt pour les utilisateurs des bases de données volumineuses, puisqu'elles sont non détectables avec les méthodes classiques d'analyse. Les techniques utilisées se basent en majorité sur des statistiques [¹]; cependant, nous nous intéresserons, spécifiquement dans notre travail, à la technique de l'extraction des règles associatives qui vise à rechercher les éventuelles corrélations entre les données support [²].

L'extraction des règles associatives est une méthode qui a vu le jour avec la recherche en bases de données pour retrouver les relations entre produits dans le panier de la ménagère qui fait son marché. En effet, dans un supermarché, la superposition des expositions ne se fait plus de façon aléatoire, une étude du comportement des clients est devenue indispensable. Par exemple on sera capable de dire que 75% des clients qui achètent du lait achètent en même temps des œufs (beurre → œuf :0.75), une telle constatation est très

¹ G. Gardarin. *Internet/Intranet et bases de données*. Edition EYROLLES, 1999.

² R. Agrawal, T. Imielinski, and A. Swami. *Mining Association Rules between sets of items in large Databases*. *ACM SIGMOD Records*, pages 207–216, 1993.

Introduction General

[Tapez le titre du document]

intéressante puisqu'elle aide le gestionnaire d'un supermarché à ranger ses rayons de telle sorte que le beurre et les œufs soient à proximité.

Plusieurs algorithmes ont été présentés dans la littérature pour l'exploration des règles associatives. On détaillera dans ce thème aux algorithmes les plus connus Apriori et AprioriTid. Ces deux algorithmes ont été présentés par Agrawal et Al en 1994 pour l'extraction des règles associatives entre les articles dans des bases de données volumineuses [3].

Notre mémoire se compose de quatre chapitres:

- 1. **Etat de l'art sur la Gestion de Relation Client:** englobant le principe de la gestion de la relation client, et quels sont ses principaux outils et objectifs.*
- 2. **Processus d'extraction des règles d'association:** Ce chapitre est divisé en deux partis ; la première représente des généralités sur la fouille de données .La deuxième partie de ce chapitre est consacrée a les mesures d'extraction des règles d'association, telle que les mesures basées sur le support et la confiance. Nous passons en revue les différents algorithmes d'extraction, Nous mettons la lumière sur l'algorithme APRIORI et APRIORITid, et nous avons clarifié ses principe de fonction ainsi leur avantages et inconvénients.*
- 3. **Conception du système:** décrivant l'architecture générale et l'architecture détaillée de notre système.*
- 4. **Implémentation et résultats:** présentant la mise en œuvre de notre travail et les résultats obtenus.*

³ R. Agrawal and R. Skirant. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Databases, pages 478—499, June 1994.

CHAPITRE I

CRM

Customer Relationship Management



INTRODUCTION

Les technologies de l'information permettent de fournir aux entreprises des outils pour mieux connaître et fidéliser leur clientèle. Lorsqu'une société se dote d'une solution de CRM c'est toujours dans l'objectif de mieux cerner les besoins de leurs clients et de mieux y répondre.

Les sources d'informations sur vos clients sont bien souvent éparpillées entre différentes personnes et gérées dans des logiciels distincts. C'est la raison pour laquelle de nombreuses entreprises se dote d'un outil de CRM afin de collecter et de structurer les informations.

Le but est d'améliorer la productivité des équipes marketing, commerciales et de tous les autres services impliqués dans le processus de vente.

*La Gestion de la Relation Client (**GRC ou CRM**) est un processus qui vous permet de traiter tout ce qui concerne l'identification de vos clients et de vos prospects. La gestion de la relation client est bien plus qu'une simple application logicielle, c'est une philosophie de travail qui trouve sa source dans l'évolution du consommateur et de son comportement d'achat.*

Définition de la gestion de relation client

Demeure Claude définit la GRC « La Gestion de la Relation Client, plus communément appelée par son acronyme GRC (en anglais CRM, ou Customer Relationship Management) désigne l'ensemble de la démarche qui, à partir d'un entrepôt de données et d'applications logicielles spécifiques, permet de pratiquer un marketing ouvert (multipliant les points de contact) et relationnel avec ses clients dans le but d'augmenter la rentabilité globale de l'entreprise »⁴.

Selon Léfber. René, Venturi. Gilles, la Gestion de la relation client, connue sous l'acronyme de CRM (Customer Relationship Management) en anglais, combine les technologies et les stratégies commerciales pour offrir aux clients les produits et les services qu'ils attendent ou qu'ils sont prêts à payer. La gestion de la relation client est la capacité à identifier, à acquérir et à fidéliser les meilleurs clients dans l'optique d'augmenter le chiffre d'affaires et les bénéfices.

⁴ Demeure. Claude, *Op.cit.* p.363.

Le terme de gestion de la relation client est devenu le fédérateur de nombreux fournisseurs de solutions informatiques. Des logiciels d'automatisation de la force de vente, aux outils de data mining, de centres d'appels ou de géomarketing, tout le monde fait de la gestion de la relation client ou du capital client. Il est évident que le discours marketing arrive à donner un côté neuf à des préoccupations anciennes et constantes des entreprises.⁵

À partir de ces différentes définitions, la gestion de la relation client consiste à savoir cibler, à attirer et à conserver les bons clients et représente un facteur déterminant du succès de l'entreprise. Construire et développer des relations avec ses clients est un défi, particulièrement lorsque l'entreprise possède des milliers (voir des millions) de clients qui communiquent avec celle-ci de multiples manières.

Pour arriver à un résultat satisfaisant, les systèmes de gestion des relations clients (Customer Relationship Management-CRM en anglais) doivent permettre aux responsables d'entreprise de mieux comprendre leurs clients pour adapter et personnaliser leurs produits ou leurs services. Un système de GRC est un Système d'information marketing, orientation client.

EN SOMME : La GRC vise à répondre à deux questions :

- *Qui sont mes clients?*
- *Comment puis-je mieux les servir?*

⁵ Léfber. René, Venturi. Gilles, *Op.cit.* p.03.



Figure 1 La Gestion Relation Client

Historique

L'émergence du concept de gestion de la relation client est le résultat d'une lente évolution de la mentalité des entreprises et des réflexions académiques, et surtout d'une transformation des systèmes marketing.⁶

Dans les années 90, les sociétés sont passées du simple recueil d'informations sur leurs clients, dans l'optique de répondre au mieux à leurs besoins, à la création d'un nouveau type d'échanges qui enrichissait l'acte d'achat et de vente : la fidélisation.

La fidélisation des clients devient l'un des axes majeurs de développement de la relation client et donc de la performance des entreprises.

⁶ Lefébure R, venturi G, Op.cit. .pp 9-12.

Cette stratégie a eu pour résultat d'accroître les revenus et l'améliorera l'appréciation de l'entreprise auprès de ses clients en leur accordant des bons points, bonus et autres cadeaux.

Alors le passage d'une orientation produit à une orientation client est dû évidemment à la volonté des entreprises pour être à l'écoute de leurs clients de façon à anticiper leurs besoins. Ce phénomène est marqué au début des années 90, ainsi le début de l'ère du client avec la naissance du marketing one To one.

Les années 2000 : Inversion des relations clients fournisseurs et le one to one.

Les années 2000 marquent l'intensification de cette tendance client, avec l'émergence du concept du marketing one To one. C'est-à-dire une offre spécifique pour chaque client. Quels que soient les secteurs d'activités, les entreprises concentrent leurs efforts sur le service et la gestion de la relation client. En parallèle, les technologies de l'information et des communications (TIC) orientent également vers une inversion des rôles, où le consommateur joue un rôle de plus en plus actif jusqu'à se substituer au distributeur à s'autoconseiller et assurer lui-même son propre service client.

Les Trois Dimensions Du Customer Relationship Management

Au cours de notre recherche, nous avons fréquemment retrouvé le CRM divisé en trois domaines d'activité interdépendants : le CRM opérationnel, le CRM collaboratif et le CRM analytique. 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 client⁷.

La GRC analytique

Vise à améliorer la connaissance et la compréhension du client. Elle permet également diffusée l'information dans l'ensemble des processus commerciaux (entrepôts des données, gestionnaire de campagne). En général, la GRC analytique analyse des informations collectées au sein de datamining et les statistiques en sont les outils de prédilection.

La GRC opérationnelle

⁷ Lamarque. Eric, «Management de la banque », Edition, Pearson Education France, paris, 2005, pp138-140.

Centré sur la gestion quotidienne de la relation avec le client, à travers l'ensemble des points de contact (centres de contacts à distance par téléphone ou internet, outils de force de vente).

Il coordonne les différents canaux d'interaction entre l'entreprise et ses clients en synchronisant les informations pour le marketing, la vente et l'ensemble des services. Ainsi le CRM opérationnel comprend différents outils: des outils de gestion du client à distance avec les centres de contacts (Call Center et web Center) , des outils d'intégration des systèmes téléphoniques et informatiques, des outils de gestion et de partage des connaissances pour optimiser la qualité des réponses apportées (grâce aux bases de données documentaires, aux FAQ - questions et réponses les plus fréquemment posées), des configurateurs de produits, des outils de pilotage de la force de vente avec la remontée automatique des ventes, de l'activité et le partage des agendas. L'ensemble de ces outils optimise le travail de la force de vente en rendant les contacts avec les clients plus fructueux.

La GRC collaborative

La GRC collaborative se traduit, par la mise en œuvre de techniques collaboratives destinées à faciliter les communications entre l'entreprise et ses clients ainsi que l'intégration avec les autres départements de l'entreprise, d'un un autre manière il regroupe tous les canaux d'échanges avec le client et les partenaires, elle met en œuvre les technologies de travail de groupe et consiste à mettre en place les canaux ou les actions pour dialoguer avec les clients : messagerie électronique, conférences, fax....

Elle a pour objet essentiel d'optimiser les contacts clients et de transmettre le bon message, au bon moment par le bon canal.

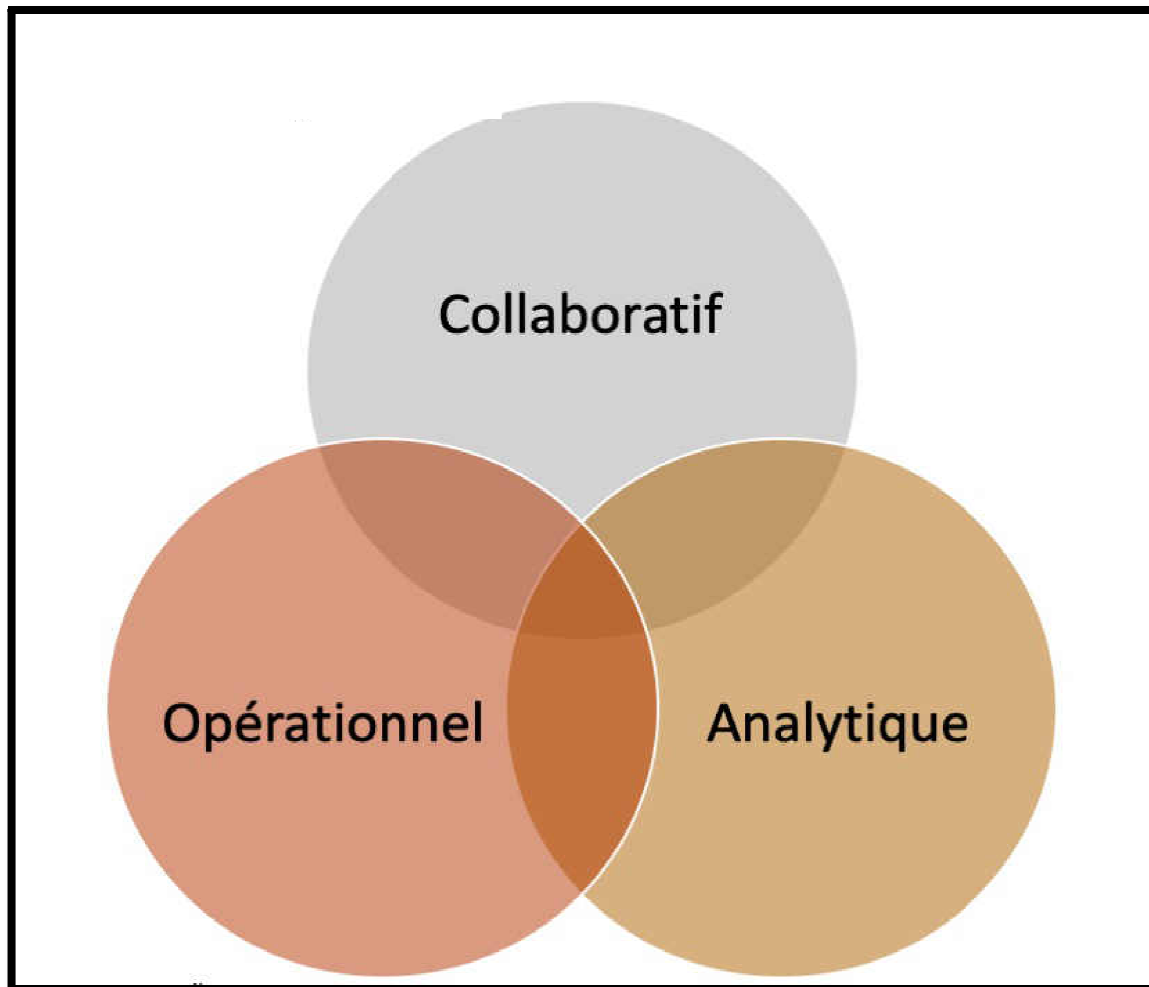


Figure 2 Les Trois Dimensions Du Customer Relationship Management

LE DATA WAREHOUSE

Définition

Lorsque l'on pense CRM, parler de data warehouse est incontournable. Il permet le stockage et l'analyse de données, et est à la base du partage de l'information et de la création des connaissances indispensables à la compréhension des clients et à la prévision de leurs comportements. Avant d'aborder ses activités, donner une définition d'un entrepôt de données nous donnera des éléments de compréhension utiles pour la suite.

« Un entrepôt de données est une collection de données orientée sujet, intégrée, non volatile et variant dans le temps à l'appui des décisions de la direction ». »⁸

⁸ Inmon, 1996

«L'entrepôt de données est « orienté sujet », c'est-à-dire organisé autour des thèmes majeurs qui touchent au métier de l'entreprise »⁹. Les données sont à l'origine regroupées par processus fonctionnels, alors qu'ici elles sont structurées par thème.

Ces données se doivent donc d'être intégrées à l'entrepôt de données. Elles sont de fait mises en forme et réunies à partir de systèmes hétérogènes et de sources externes dans une seule et même base de données. Cette phase du processus est la plus complexes du data warehousing

Troisièmement, l'entrepôt de données stocke des données « non-volatiles », ce qui veut dire qu'elles ne sont pas mises à jour, ni modifiables. Elles sont comme photographiées, seule leur lecture est possible ce qui n'est pas le cas des données contenues dans les systèmes d'information opérationnels.

Enfin, dans la continuité, les données sont simplement ajoutées les unes après les autres au sein de l'entrepôt de manière chronologique. Chaque donnée est ainsi associée à une date à laquelle elle a été créée.

En résumé, le data warehousing est un élément fondamental du CRM analytique puisqu'il permet d'extraire et de stocker les données dans un seul endroit afin d'y être analysées et d'en tirer des connaissances qui serviront de support à la prise de décisions et à l'amélioration des processus au sein de l'entreprise.

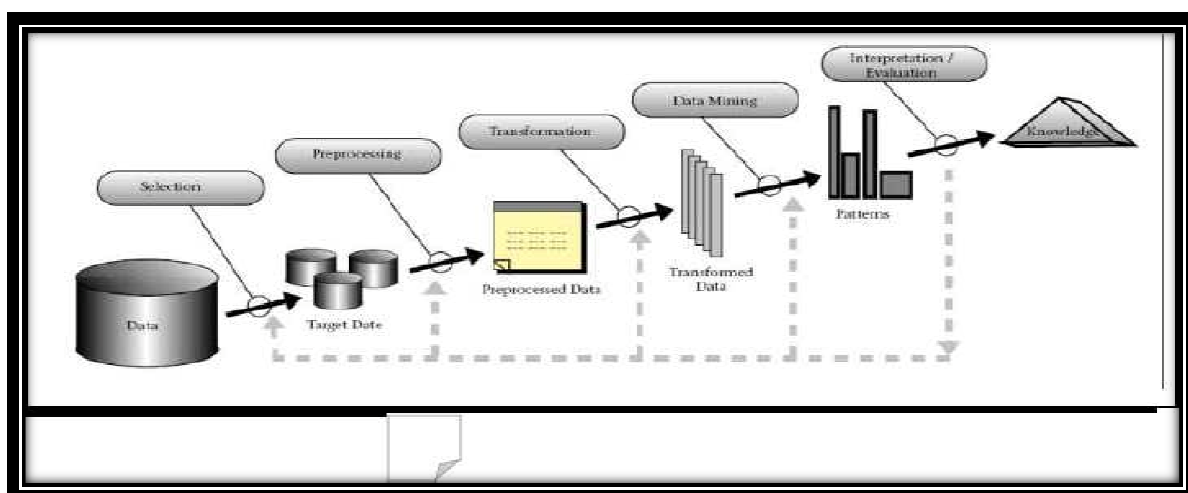


Figure 3 LE DATA WAREHOUSE

⁹ Grenier, C. et Moine, C., 2003

Fonctions de la GRC

Les fonctions d'un CRM peuvent être résumées à : connaître, choisir, conquérir et fidéliser la clientèle.

Connaître le client

L'entreprise doit rassembler les informations lui permettant de décrire et de caractériser sa clientèle, de la positionner sur son marché et de détecter de nouveaux segments. Tous les moyens technologiques existent aujourd'hui pour constituer, gérer et analyser des quantités massives de données. Gérer la relation client consiste à valoriser son capital client. D'un point, de vue technique, le CRM implique de capturer, au niveau de l'entreprise, l'ensemble des données clients, collectées en interne ou auprès d'organisations extérieures, et de les intégrer dans UN entrepôt de données qui orienter client.

Choisir son client

L'étape suivante consiste à analyser ces données avec les techniques les plus évoluées. Le Datamining, analyse statistique à rendre les résultats accessibles à tous les canaux d'interaction avec les clients.

Le Datamining permet d'analyser et d'interpréter un gros volume de données, de différentes sources afin de dégager des tendances, de rassembler les éléments similaires en catégories statistiques et de formuler des hypothèses.

À partir des informations collectées, l'entreprise pourra obtenir des réponses objectives sur lesquelles fonder sa stratégie opérationnelle. La centralisation des données clients doit ainsi faciliter le pilotage de toute l'activité de la société.

En effet, l'informatique décisionnelle (Business Intelligence et Datamining) permet d'élaborer les diverses composantes de la stratégie : (commerciale, marketing, canaux de vente, fidélisation) et fournit tous les tableaux de bord nécessaires.

Ainsi il faut différencier les clients en fonction de leurs besoins et de leurs contributions aux résultats et dialoguer avec eux de manière à diminuer les coûts de la relation commerciale et à en augmenter l'efficacité. Ce dialogue doit permettre de faire remonter l'information.

Conquérir de nouveaux clients

La mise en œuvre d'une stratégie orientée client concerne l'ensemble du processus commercial. Les nouveaux canaux de ventes (télévente, commerce électronique, etc.) créent des opportunités métiers. De nouveaux outils (Sales Forces Automation) permettent aux commerciaux de mieux gérer leur activité et d'augmenter leur efficacité en construisant leurs propositions en interaction directe avec le client

Fidéliser les meilleurs clients

Les programmes de fidélisation bénéficient de nouvelles possibilités technologiques, telles que la carte de fidélité. Le service après-vente devient l'occasion privilégiée de concrétiser une relation personnalisée et durable avec le client, en lui proposant une offre encore mieux adaptée à ses besoins.

Le vecteur idéal de cette relation est le centre d'appel (call center) qui permet d'orchestrer tous les éléments de la stratégie client, depuis la base de connaissances qui fournit la vue unique du client nécessaire à cette relation "one to one", jusqu'au scénario personnalisé qui guide l'entretien pour lui présenter une offre adaptée à ces besoins.



Figure 4 Fonctions de la GRC

L'importance des technologies et les logiciels dans la GRC

Une GRC a pour but de centralisation des informations sur le client. Les logiciels permettent de définir et de gérer la mise en œuvre des tactiques de marketing direct son rôle est de définir des campagnes et des actions, sélection des cibles, extraction des cibles...etc. En général, ces outils permettent de planifier des actions conditionnelles soit en fonction des événements propres au client (par exemple, pour une relance quelques mois avant l'échéance d'une garantie), soit en fonction de la campagne (par exemple, sélection pour un appel de relance en cas de non-réponse dans les quinze jours suivant l'envoi d'un mailing).¹⁰

Les technologies de GRC peuvent aider pour ¹¹ :

- Analyser les données de revenu et de coûts sur les clients de manière à identifier ceux à forte valeur ;
- Mieux cibler les efforts marketing direct ;
- Disposer de données pertinentes sur les comportements face aux produits et services ;
- Créer de nouveaux circuits de distribution ;
- Développer de nouveaux modèles de fixation des prix ;
- Construire des communautés des clients ;
- Gérer les transactions plus rapidement ;
- Fournir de meilleures informations au personnel en contact ;
- Gérer plus efficacement la logistique et la chaîne d'approvisionnement ;
- Aligner les stimulants du personnel sur les indicateurs de qualité de qualité de la relation ;
- Déployer des systèmes de gestion des connaissances ;
- Suivre les taux de défection et de fidélité ;
- Suivre le degré de satisfaction des clients.

Les technologies représentent une potentialité nouvelle de relation directe entre le client et l'entreprise, ces technologies doivent mener des solutions plus efficaces pour que l'entreprise puisse satisfaire et même fidéliser ses clients.

Les approches de la gestion de la relation client :

¹⁰ LEFEBURE - GILLES. *Op.cit*, p. 53

¹¹ DARREL Rigby- REICHFELD Frederick- SCHEFFER Phil. *Aviod the Four Perils of CRM*, Harvard Business Review, février 2002, p.106.

Ce concept peut avoir différentes significations, plus précisément, on peut distinguer cinq approches :

- ❖ **La GRC est un processus** : Les relations clients fournisseurs se développent au fil du temps et doivent évoluer pour perdurer. La réussite de la GRC est liée à la capacité de détecter et de répondre à l'évolution des besoins et préférences des clients.
- ❖ **La GRC est une stratégie** : La valeur à vie du client (CLTV) détermine le montant et les types de ressources que la firme investit dans une relation particulière. La réussite de la GRC nécessite que les entreprises évaluent et hiérarchisent les relations avec les clients en fonction de la rentabilité de leur valeur à vie (CLTV).
- ❖ **La GRC est une philosophie**. La GRC est mieux réalisée par une focalisation sur le renforcement et le maintien des relations avec les clients. La réussite de la GRC exige que la firme soit, d'une part, orientée client et d'autre part, guidée par la compréhension du changement des besoins des clients.
- ❖ **La GRC est une capacité**. Les relations profitables de long terme est le résultat de l'adaptation permanente du comportement de la firme envers les clients. La réussite de la GRC est liée à la possession de la firme d'un ensemble de ressources matérielles et immatérielles qui lui confèrent la flexibilité lui permettant de changer son comportement envers les clients individuels de manière continue.
- ❖ **La GRC est une technologie**. La connaissance et les technologies de gestion des interactions représentent les ressources clés de la firme nécessaires pour l'établissement des relations profitables de long terme avec les clients. La réussite de la GRC résulte principalement de l'utilisation de la technologie et de ses fonctionnalités dans l'objectif d'établir une connaissance des clients et de gérer leurs interactions avec la firme. Les approches de la GRC dans leur globalité visent à mieux comprendre les clients et répondre à leurs besoins et attentes. Elles s'inscrivent dans le cadre du marketing relationnel

Les Enjeux D'une GRC

L'enjeu principal de la gestion de la relation client est d'améliorer la relation et la connaissance des clients pour que l'entreprise puisse garantir des gains mais aussi garantir un client fidèle. La GRC peut être une réponse à trois enjeux qui sont :

L'intégration multi canal

Les entreprises étendent progressivement leurs modes de communication et de distributions auprès de leurs clients (représentants, points de vente, catalogues, mailing, centres d'appel, Internet et autres canaux électroniques...etc.). Où les clients veulent avoir une expérience continue et interagir avec la même entreprise, quelque soit le canal utilisé. Et pour mieux identifier les clients, il faut conserver la mémoire des interactions avec eux, faire partager cette connaissance entre les différents canaux, permettre au client d'interagir de façon cohérente avec l'entreprise à travers multitude de canaux, sont autant d'enjeux qui deviennent ainsi une préoccupation des entreprises.

L'accroissement de la productivité

Dans la gestion de la relation avec le client, de nombreuses tâches qui peuvent être partiellement ou totalement automatisées, aussi bien en matière de communication que de commercialisation. Cette automatisation peut conduire à des réductions de coût, c'est-à-dire à des gains de productivité que recherchent constamment les entreprises. La GRC s'inscrit dans cette perspective.

L'accroissement de la fidélité des clients

Dans les économies développées d'où la croissance est relativement faible, les entreprises fidélisent leurs clients pour maintenir leurs parts du marché et pour répondre à cette attente, les entreprises individualisent leurs relations avec ses clients avec la connaissance de leurs profils, leurs habitudes et avec une communication plus personnalisée.

Le processus de la gestion de la relation client

Le processus de la gestion de la relation client passe par cinq (5) étapes essentiels :¹² identifier, segmenter, adapter, échanger, évaluer.

Identifier

L'identification consiste à collecter des informations sur chaque client, ou du moins ceux qui ont été ciblés dans le plan d'action. Ces informations doivent pouvoir être formalisées et intégrées dans une base de données afin que l'entreprise bénéficie d'un mode de connaissance systématique et automatisé des clients.

Segmenter

¹² PIERRE (A) et PIERRE (A) - Guggémos, « CRM et les clés de réussite », Edition, Organisation, Paris, 2005, P (9-11)

L'identification des clients n'est qu'une première étape ; il faut analyser les données et regrouper les clients. La segmentation sur bases de données consiste à regrouper les clients en fonction de caractéristiques communes susceptibles d'affecter leur comportement. On pourra ainsi segmenter les clients en fonction de l'importance de leurs achats (petits/moyens/gros clients...), du type de produit achetés, de leurs centre d'intérêt exprimé, de leur probabilité calculée de défection, etc. La segmentation par rentabilité est un type de segmentation extrêmement important pour conduire une politique relationnelle.

Adapter

L'identification et la segmentation des clients doivent conduire à adapter le service et la communication auprès des clients. Cette adaptation portée sur les offres, sur les canaux de contact ou encore sur le contenu de la communication. Internet est un canal particulièrement propice à la personnalisation de la communication. Un site web peut en effet adapter son contenu en fonction du profil du visiteur, reconnaître et accueillir un visiteur identifié (qui aura donné son identité lors d'une visite), proposer des services liés au profil des visiteurs tels que des recommandations personnalisées et enrichir progressivement le profil de l'internaute.

Echanger

Les interactions résultent soit de campagnes organisées par l'entreprise, soit d'une réponse aux sollicitations du client. Dans le premier cas, l'entreprise postera un courrier, passera un appel téléphonique, enverra un e-mail, etc. pour faire une offre au client ou entretenir une relation. dans le second cas, c'est le client qui sollicite l'entreprise, en appelant un standard téléphonique, en envoyant un courrier ou un e-mail, ou en se rendant dans point de vente : dans ces derniers cas, l'entreprise doit tirer profit des opportunités de contact en nourrissant la base de données d'informations nouvelles sur le client ainsi qu'en lui proposant en réponse d'offre spécifique. La pertinence de l'interaction est décisive pour la satisfaction du client. Une interaction mal conçue, une offre ou une communication mal adaptée est de faible valeur pour le client : elle peut même détruire de la valeur en irritant inutilement le client.

Evaluer

La relation clients se construit dans le temps de s'enrichit à chaque interaction. De ce fait, l'apprentissage est une démentions essentiel de ce processus. Les objectifs doivent être quantifiés pour pouvoir être évalués. Les indicateurs peuvent prendre plusieurs formes :

indice de satisfaction, taux d'attrition, chiffre d'affaire par client, rentabilité par client, part de client, taux de transformation des courriers ou de bons, taux d'ouverture de newsletters, etc. Au sein de l'entreprise, des résultats partiels et démontrables dans le cadre d'une démarche progressive sont les meilleurs arguments pour faire avancer le dossier de la relation client.

Les politiques de la relation client

On peut proposer au moins 4 types de politiques¹³ :

Politique de (re)conquête

Cette politique vise à transformer des prospects ou des anciens clients de l'entreprise en clients actifs. Elle suppose la mise en œuvre d'arguments de séductions (offres spéciales) nouveaux et puissants.

Politique d'abandon

Cette politique consiste à délaissier une clientèle peu rentable et peu stratégique. Le degré de rentabilité de cette clientèle détermine le degré d'urgence de la cessation des relations commerciales.

Politique de fidélisation

Cette politique ambitionne l'intensification et a pérennisation d'une relation Commerciale. Elle suppose souvent des efforts de la part du vendeur.

Politique de rationalisation

Cette politique cherche à améliorer la rentabilité de clients réguliers mais peu lucratifs. Cette rationalisation passe par la réduction de tous les coûts liés à ces clients (coûts commerciaux, coûts industriels...).

Ces politiques sont choisies en fonction de la rentabilité actuelle et potentielle, degré de priorité stratégique du client et de la notion de RFM (Récence, Fréquence, Montant).

Avantages d'un système CRM au sein d'une entreprise :

Offre une meilleure connaissance des clients

¹³ MIRATON Lionel, « introduction à la gestion de la relation client »P.6.

Un logiciel CRM est une base de données dans laquelle sont centralisées toutes les informations et données que l'entreprise possède sur ses clients¹⁴. Les données stockées par cet outil sont multiples¹⁵. On y retrouve, premièrement, toutes les informations ayant trait à l'identité du client que ce soit son nom, son adresse, son sexe, son âge, etc. Toutes les informations relatives aux personnes de contact sont également disponibles dans le système, ses données de contact (adresse email, numéro de téléphone, de téléphone portable, ...),

ces informations offrent à l'entreprise et à tous ses collaborateurs, une meilleure connaissance de leurs clients¹⁶. et en contrepartie l'entreprise peut offrir une expérience individualisée à ses clients et leur proposer des biens et services personnalisés par rapport à leur profil.

Augmente la satisfaction des clients et leur fidélité

Conséquence directe du points précédent, les logiciels CRM augmente la satisfaction des clients et par conséquent leur fidélité vis à vis de l'entreprise. En effet, l'entreprise bénéficiant de beaucoup d'informations sur ses clients, elle peut répondre à leurs besoins et envies personnelles¹⁷. Le fait de connaître personnellement le client permet de lui offrir des services et biens personnalisés¹⁸. Ce service sur mesure est un réel avantage pour montrer au client qu'il a de la valeur et compte pour l'entreprise.

Améliore la gestion des contacts et l'accès à leurs informations

Grâce au système CRM, les données des clients et toutes informations relatives aux interactions entre les clients et l'entreprise sont centralisées au même endroit¹⁹. L'accès à toutes ces données est donc facilité car elles sont accessibles partout en ligne via un appareil connecté et la personne qui a besoin d'une certaine information est sûre de la trouver dans le CRM (si l'information est connue de l'entreprise) alors que par le passé certaines informations sur un client était dans le fichier Excel d'un commercial, d'autres dans l'ordinateur du service comptabilité, etc. De plus, l'accès à ces données est disponible pour tous les employés de l'entreprise ayant un accès au CRM²⁰.

¹⁴ Buttle, 2009, p. 366

¹⁵ Jallat et al., 2018, pp. 151-152

¹⁶ Site web de Indice RH, 15.05.2019

¹⁷ Chou et al., 2002, p. 444

¹⁸ Jallat et al., 2018, pp. 121-125

¹⁹ Peelen, 2005

²⁰ Site web de Indice RH, 15.05.2019

Améliore la communication et la collaboration entre les salariés

L'implémentation d'un système CRM dans une organisation améliore la collaboration entre les employés de l'entreprise et entre les différentes équipes composant l'entreprise. Il améliore le partage des connaissances²¹.

La cohésion d'une équipe et sa réussite passent par une bonne collaboration et communication entre ses membres.

Fournit une certaine quantité de rapports et tableaux de bord intéressants

Les logiciels CRM compilent des données de nature variée et en grande quantité ce qui leur permettent de proposer des rapports et tableaux de bord intéressants pour les différents salariés de l'entreprise. Ces rapports aident l'entreprise à adopter un style de prise de décision basé sur les données plutôt que basé sur l'intuition

Il est également possible de comparer les performances de l'entreprise avec des analyses prévisionnelles pour repérer les anomalies.

Accroît la productivité et le revenu de l'entreprise

Comme discuté dans les points précédents, un outil CRM permet de mieux connaître ses clients et donc par conséquent de mieux les satisfaire et les fidéliser à l'entreprise. Un client qui est satisfait et qui est fidèle à l'entreprise renouvellera son achat chez son fournisseur actuel plutôt que d'aller voir chez la concurrence²².

Augmente l'efficacité du marketing

L'implémentation d'un logiciel CRM permet à l'entreprise de segmenter ses clients et prospects, de générer des listes de clients ou prospects selon certains critères, de lancer et évaluer des campagnes marketing et de gérer les leads.²³

Améliore la qualité des services et produits octroyés aux clients

En permettant à l'entreprise de mieux connaître ses clients, le logiciel CRM lui offre la possibilité d'offrir une meilleure qualité des services et des produits à ses clients. En effet, en connaissant la personnalité ainsi que les besoins ou les préférences de ses clients, l'entreprise peut personnaliser les biens ou services qu'elle offre à tel client. Un produit

²¹ Nguyen, Sherif & Newby, 2007, p. 103

²² Buttle, 2009

²³ Buttle, 2009

adapté à un client est toujours un produit de meilleure qualité qu'un produit qui ne correspond pas au client²⁴.

Améliore la gestion des ventes

Plusieurs fonctionnalités des logiciels CRM améliorent la gestion des ventes au sein de l'entreprise²⁵. Premièrement, il y a un système de gestion des transactions pour surveiller leur état d'avancement. Ensuite, il y a l'automatisation du processus de vente²⁶. Il permet à l'équipe commerciale de passer moins de temps à réaliser les tâches administratives.

De la même manière, tout le processus lié aux leads peut être automatisé.

Permet des prévisions de ventes précises.

Grâce au logiciel CRM, les commerciaux disposent d'une base de données reprenant toutes les informations et étant fiable. Ils peuvent donc prévoir plus efficacement les ventes grâce par exemple au système de ventes en cours ou en attente.

Offre une synchronisation avec les autres outils utilisés par l'entreprise

Les logiciels CRM proposent une synchronisation aisée et intuitive avec les autres outils qui sont utilisés par l'entreprise²⁷.

Inconvénients d'un système CRM au sein d'une entreprise :

Demande de former les employés à l'utilisation du CRM

Lors de l'implémentation d'un système CRM dans une entreprise, l'organisation doit former l'ensemble des personnes qui seront en contact avec le logiciel²⁸. L'entreprise doit donc prévoir des horaires de formation pour tous ses employés et doit peut-être prévoir d'engager des professionnels (des consultants en CRM par exemple) pour former les employés par groupe.

Former les employés coûtent donc de l'argent mais aussi du temps. La quantité de temps et d'argent à investir dans la formation dépend de beaucoup de variables dont la taille de l'organisation

²⁴ Mithas, Krishnan & Fornell, 2005

²⁵ Buttle, 2009

²⁶ Buttle, 2009

²⁷ Site Web de Teamleader, 24.05.2019

²⁸ Brown, 2000

Entraîne des coûts

L'un des plus grands désavantages et l'un des plus grands défis que doit surmonter une entreprise lors de l'implémentation d'un système CRM, est son coût²⁹.

Ralentit légèrement le processus

En effet, implémenter un outil de ce style ralentit légèrement les processus pour certains postes de l'entreprise. Après avoir eu un client au téléphone, après avoir rendu visite à un client ou après avoir reçu une plainte du client, le commercial doit encoder toutes ces informations dans le CRM. Par conséquent, tout ceci allonge un peu le temps nécessaire pour chaque tâche³⁰. Surtout au début de l'implémentation, les employés ne sont pas encore trop à l'aise avec l'outil et encoder toutes les données peut sembler long et fastidieux mais avec le temps et l'expérience, l'encodage se fera beaucoup plus rapidement et les bénéfices à long-terme surpassent largement le faible temps utilisé à court terme. Les commerciaux par exemple, se plaignent de devoir passer des heures à renseigner le CRM plutôt que de dédier leur temps à leur objectif principal : la vente³¹.

Donne aux salariés le sentiment d'être surveillés

L'un des principaux désavantages de l'utilisation d'un système CRM au sein d'une organisation est que certains collaborateurs ont le sentiment que l'outil a été implémenté pour les surveiller et les contrôler. Ils ne voient pas l'outil comme un moyen d'optimiser leur travail et leur planning mais comme un moyen pour l'entreprise de surveiller ce qu'ils font au jour le jour³². En effet, loin des rapports mensuels, le CRM permet de voir les tâches encodées quotidiennement par chaque employé.

Peut transformer l'entreprise en « usine à gaz » et éliminer l'élément humain.

Dans le cas de certains CRM, en particulier les CRM produits par des grosses entreprises, le CRM peut devenir tellement imposant et contenir tellement de données que l'entreprise devient ce que l'on peut appeler une usine à gaz ou des milliers de données entrent et ressortent analysées

Présente des limites à son utilisation

²⁹ Brown, 2000, pp. 253- 255

³⁰ Site Web de targetintegration, 22.05.2019

³¹ Lucron, 2018

³² Lucron, 2018

on citait dans les avantages des logiciels CRM qu'ils sont combinables avec d'autres outils, c'est avant tout parce que les fonctionnalités proposées par les logiciels CRM restent en général assez limitées quand ils sont utilisés seuls et ils doivent être combinés à d'autres outils pour être totalement efficaces³³.

Un des problèmes des logiciels est que c'est une bonne base de données qui tente d'aider dans plusieurs domaines de l'entreprise mais vous êtes rarement à 100% efficace sans les combiner à d'autres outils. Par exemple, pour les campagnes marketing, les logiciels CRM proposent une aide pour les entreprises mais ils doivent être combinés à des outils d'emailing ou de marketing automation pour être réellement efficaces.

Quels sont les bénéfices pour une entreprise d'avoir un CRM ?

Un logiciel de gestion de la relation avec les clients permet d'aider une entreprise à mieux définir sa stratégie d'affaires pour éviter qu'un client insatisfait passe à la concurrence. Il est donc essentiel pour une entreprise de comprendre les attentes et les besoins de ses clients. Ainsi, la mise en place d'un logiciel de CRM est devenue indispensable. L'entreprise pourra facilement centraliser les demandes de ses clients et fournir les ressources nécessaires en fonction des priorités.

Une entreprise qui a installé un logiciel de CRM pourra bénéficier :

- d'une augmentation du pourcentage de clients fidèles.
- **d'un gain de temps** : les employés de l'entreprise seront libérés des tâches répétitives grâce à l'automatisation des tâches par le logiciel. Ils seront en charge de missions davantage stratégiques et productives.
- **d'une collaboration optimale entre les services** : l'ensemble des étapes du cycle de vente sont impactées grâce au logiciel qui diffuse les données recueillies à chaque département.
- **de réactivité** : les informations récoltées par le logiciel permettent de réagir aux problèmes réguliers, d'avoir une vision globale des difficultés et d'évaluer les besoins des clients.
- **d'un accroissement du chiffre d'affaires** : la valeur du portefeuille client sera développée et des marges plus élevées seront dégagées.

³³ Site Web de Noeo, 12.05.2019

CONCLUSION

La gestion de la relation client consiste à savoir cibler, à attirer et conserver les bons clients et représente un facteur déterminant du succès de l'entreprise. Construire et développer des relations avec ses clients est un défi, particulièrement lorsque l'entreprise possède des millions de clients qui communiquent avec celle-ci de multiples manières.

Pour arriver à un résultat satisfaisant, les systèmes de gestion des relations clients (GRC) doivent permettre aux responsables d'entreprise de mieux comprendre leurs clients pour adapter et personnaliser leurs produits ou leurs services. Il doit aussi permettre aussi différents départements de l'entreprise de collecter à travers le partage d'information concernant toute interaction avec le client.

Le choix d'une stratégie de relation client a des implications profondes sur toute l'entreprise. Le fait de mettre le client au cœur de l'activité suppose l'existence d'un système d'information très développé (CRM), d'une véritable stratégie de connaissance et d'un alignement de l'entreprise pour mettre en cohérence ses actes vis-à-vis du client.

La fidélisation de la clientèle est déterminée par la nature des services présentés par l'entreprise. Cette dernière doit avoir des mécanismes et un arsenal de moyens pour réaliser une étude sur le comportement de sa clientèle prenant compte l'évolution de l'environnement et de l'introduction des technologies d'information et de communication (TIC).

Avec une exploitation des bases de données plus affinées pour une meilleure détermination des besoins et une gestion de la relation client plus efficace.

*Enfin nous concluons que **La relation client se divise en plusieurs stades** ; le premier stade c'est **L'avant vente** qui consiste à étudier le marché, c'est-à-dire les besoins des clients et ensuite à les démarcher. Puis on analyse les réponses obtenues afin de répondre plus précisément aux attentes de ces clients.*

*Et le deuxième stade c'est **La gestion de la relation client valorise le client**, qui aime se sentir connu et reconnu de l'entreprise.*

*Et le dernier stade **L'après-vente** comprend à fournir une assistance au client après l'achat.*

Ces différents stades sont des clés de réussite de la relation client. L'entreprise doit rester constamment en veille vis-à-vis de ses clients.

CHAPITRE II

PROCESSUS D'EXTRACTION DES

REGLES D'ASSOCIATION



1 Introduction :

Les évolutions technologiques réalisées ces dernières années ont permis de diminuer de façon considérable les coûts de collecte et de stockage de données. De nombreuses entreprises ont saisi l'opportunité d'archiver de nombreuses informations (informations clients, informations produits, ...) dans l'espoir de pouvoir les exploiter. Ces sources d'informations potentielles jouent un rôle important dans le domaine de la concurrence ; il reste à les utiliser au mieux. Les techniques traditionnelles de statistiques ne sont pas utilisables sur d'aussi grandes quantités de données sans l'utilisation de techniques d'échantillonnage. Le Data Mining apparaît alors comme la solution pour extirper des connaissances de cet amas de données disponibles.

Le Data Mining répond à un ensemble de tâches du domaine scientifique en fournissant un ensemble de technique qui répond à ces tâches spécifiques. extraction de règle d'association est devenue une des tâches fondamentales du Data Mining, Les règles d'association sont les méthodes les plus répandues dans le domaine de marketing et de la distribution. Leur principe consiste à trouver les groupes d'articles qui apparaissent le plus fréquemment ensembles, et de générer des règles d'association. Ces règles sont simples, faciles à comprendre et assorties d'une probabilité, ce qui en fait un outil agréable et directement exploitable par l'utilisateur métier

Bien qu'elle soit traditionnellement liée au secteur de la distribution, l'extraction des règles d'association peut s'appliquer à d'autres domaines :

- L'utilisation d'une carte de crédit (location de voiture, réservation de chambre d'hôtel) peut renseigner sur les habitudes d'un client.
- Les options achetées par les clients des télécommunications (attente d'appel, renvoi d'appel, appel rapide, etc.) aident à savoir comment regrouper ces services pour maximiser le revenu.
- Les carnets de santé des patients peuvent donner des indications sur les complications induites par certaines combinaisons de traitements.
- Les services bancaires (comptes numéraires, bons de caisse, services d'investissement, locations de voitures) permettent d'identifier les clients susceptibles de désirer d'autres services.

- *La combinaison dans les demandes d'assurances inhabituelles peut signaler une fraude et déclencher des suppléments d'enquête.*

La plupart des algorithmes d'extraction des règles d'association mettent en œuvre deux propriétés le support et la confiance. On parlera de ces critères et des mesures, mais aussi de l'algorithme d'exploration des données permettant d'extraire les règles d'association ultérieurement .

Dans ce chapitre, nous présentée le processus complet d'extraction de connaissances à partir de données ainsi que les domaines d'applications, les tâches et les techniques du Data Mining. Ensuite on va focaliser sur une technique très répandue qui est les règles d'association. Un plus détail est donné sur cette technique et leurs différents algorithmes .

Le Data Mining

Le terme de Data Mining est souvent employé pour désigner l'ensemble des outils permettant à l'utilisateur d'accéder aux données de l'entreprise et de les analyser. On distingue le terme « Data Mining » (fouille de données) et le terme « Extraction de connaissance»³⁴, même si par abus de langage ces termes sont utilisés pour définir la découverte (ou extraction) de connaissance : KDD (Knowledge Discovery in Data bases). L'extraction de connaissances est réalisée grâce au processus particulier qu'est la fouille de données (Data Mining) qui apparaît donc comme une méthode d'extraction parmi d'autres. Nous restreindrons ici le terme de Data Mining aux outils ayant pour objet de générer des informations riches à partir des données de l'entreprise, notamment des données historiques, de découvrir des modèles implicites dans les données. On pourrait définir le Data Mining comme une démarche d'exploration de données ayant pour objet de découvrir des relations et des faits (des connaissances), à la fois nouveaux et significatifs sur de grands ensembles de données. Ces concepts s'appuient sur le constat qu'il existe au sein de chaque entreprise des informations cachées dans le gisement de données. Ils permettent, grâce à un certain nombre de techniques spécifiques, de faire apparaître des connaissances.³⁵

³⁴ Piatetsky-Shapiro (G.), Frawely (W.J.). "Knowledge Discovery in Databases". AAA Press, 1991.

³⁵ Conservatoire National des Arts et métier de Lille "Data warehouse et Data Mining". version 1.1, le 15 juin 1998, disponible sur : www.imi-utc.com/ptemoign/index.html

Les tâches du Data Mining :

Une multitude de problèmes d'ordre intellectuel, économique ou commercial peuvent être regroupés, dans leur formalisation, dans l'une des tâches suivantes : ³⁶

La classification:

La classification consiste à examiner des caractéristiques d'un élément nouvellement présenté afin de l'affecter à une classe d'un ensemble prédéfini.

L'estimation :

Contrairement à la classification, le résultat d'une estimation permet d'obtenir une variable continue. Celle-ci est obtenue par une ou plusieurs fonctions combinant les données en entrée. Le résultat d'une estimation permet de procéder aux classifications grâce à un barème.

La prédiction :

La prédiction ressemble à la classification et à l'estimation mais dans une échelle temporelle différente. Tout comme les tâches précédentes, elle s'appuie sur le passé et le présent mais son résultat se situe dans un futur généralement précisé.

La Segmentation :

La segmentation consiste à segmenter une population hétérogène en sous-populations homogènes. Contrairement à la classification, les sous-populations ne sont pas préétablies.

La description :

C'est souvent l'une des premières tâches demandées à un outil de Data Mining. On lui demande de décrire les données d'une base complexe. Cela engendre souvent une exploitation supplémentaire en vue de fournir des explications.

L'optimisation :

Pour résoudre de nombreux problèmes, il est courant pour chaque solution potentielle d'y associer une fonction d'évaluation. Le but de l'optimisation est de maximiser ou

³⁶ BERRY (M.J.A.), LINOFF (G.). "Data Mining-Techniques appliquées au marketing, à la vente et aux services clients" - 1996 MASSON (Wiley)

minimiser cette fonction. Quelques spécialistes considèrent que ce type de problème ne relève pas du Data Mining.

Les techniques de Data Mining :

Plusieurs techniques peuvent être inscrites dans le contexte du Data Mining, on en cite :

Les arbres de décision :

Ce sont des outils très puissants principalement utilisés pour la classification, la description ou l'estimation. Il s'agit d'une représentation graphique sous forme d'arbre de décision représentant un enchaînement hiérarchique de règles logiques qui permet de diviser la base d'exemples en sous-groupes.

Les réseaux de neurones :

Inspirés de la biologie, les réseaux de neurones représentent une transposition simplifiée des neurones du cerveau humain. Ils sont utilisés dans la prédiction et la classification.

Les algorithmes génétiques :

Ce sont des méthodes d'optimisation de fonctions basée sur les mécanismes génétiques pour élaborer des paramètres de prévision des plus optimaux. Ils permettent de résoudre des problèmes divers, notamment d'optimisation, d'affectation ou de prédiction.

Les règles d'association :

Ce sont les méthodes les plus répandues dans le domaine de marketing et de la distribution. Elles peuvent être appliquées dans différents secteurs d'activité pour lesquels, il est intéressant de trouver des groupements d'articles qui apparaissent le plus fréquemment ensembles, et de générer des règles d'associations. Le but principal de cette technique est descriptif ou prédictif.

Les plus proches voisins (CBR, Case Based Reasoning) :

C'est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimation. De façon similaire, le « raisonnement basé sur la mémoire », technique du DataMining dirigé, permet de classer ou de prédire des données inconnues à partir d'instances connues.

Le Clustering :

C'est une classification non supervisée, les classes possibles et leur nombre ne sont pas connus à l'avance et les exemples disponibles sont non étiquetés. Le but est donc de découvrir des relations intéressantes qui peuvent exister implicitement entre les données et qui permettront de regrouper dans un même groupe ou cluster les objets considérés comme similaires.

*Nous nous intéresserons dans nos travaux à la technique des **règles d'associations**. Nous détaillerons plus loin dans ce chapitre les grandes lignes de cette technique.*

Les Règles d'Association

Le concept de règles d'association a été introduit en 1993 par Agrawal [3].

Les règles d'association sont traditionnellement liées au secteur de la distribution. Leur principale application est "l'analyse du panier de la ménagère" ³⁷ L'analyse des tickets de caisse des clients permet de comprendre leurs habitudes de consommation, d'agencer les rayons d'une surface commerciale, d'organiser les promotions, de gérer les stocks, etc. Dans les bases de données de vente, un tuple est une transaction composée d'un ensemble d'articles achetés ou items. La base de données est donc un ensemble de transactions appelé base transactionnelle ou base de transactions. Une règle d'association décrit une corrélation entre des ensembles d'items, ou itemsets, dans une base de transactions. Autrement dit, étant donné deux itemsets I_1 et I_2 , l'objectif est de découvrir une transaction contenant I_1 et I_2 ³⁸ .

Exemple: "80% des clients qui achètent un ordinateur achètent aussi une imprimante et un abonnement à Internet" est une règle d'association associant l'itemset {ordinateur} à l'itemset {imprimante, abonnement à Internet}.

La technique qui génère des règles d'association peut être appliquée à tout secteur d'activité pour lequel il est intéressant de rechercher des groupements potentiels de produits ou de services: services bancaires, services de télécommunications, par exemple. Elle peut être également utilisée dans le secteur médical pour la recherche de complications dues à des

³⁷ R. Rakotomalala. "Les règles d'Association, market data analysis ou L'analyse du panier de la ménagère"

³⁸ R. Agrawal et R. Srikant. "Fast algorithms for mining association rules in large databases". In VLDB'94 : Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487-499, San Francisco, CA, USA, 1994

associations de médicaments ou à la recherche de fraudes en recherchant des associations inhabituelles.³⁹

Ces règles montrent comment des produits ou des services se situent les uns par rapport aux autres. Elles peuvent ainsi être facilement utilisées dans le système d'information de l'entreprise. Cependant, il faut noter que la méthode, si elle peut produire des règles intéressantes, peut aussi produire des règles triviales ou inutiles. La Figure 6 illustre les différentes phases de traitements des données pour arriver à l'extraction de la connaissance.

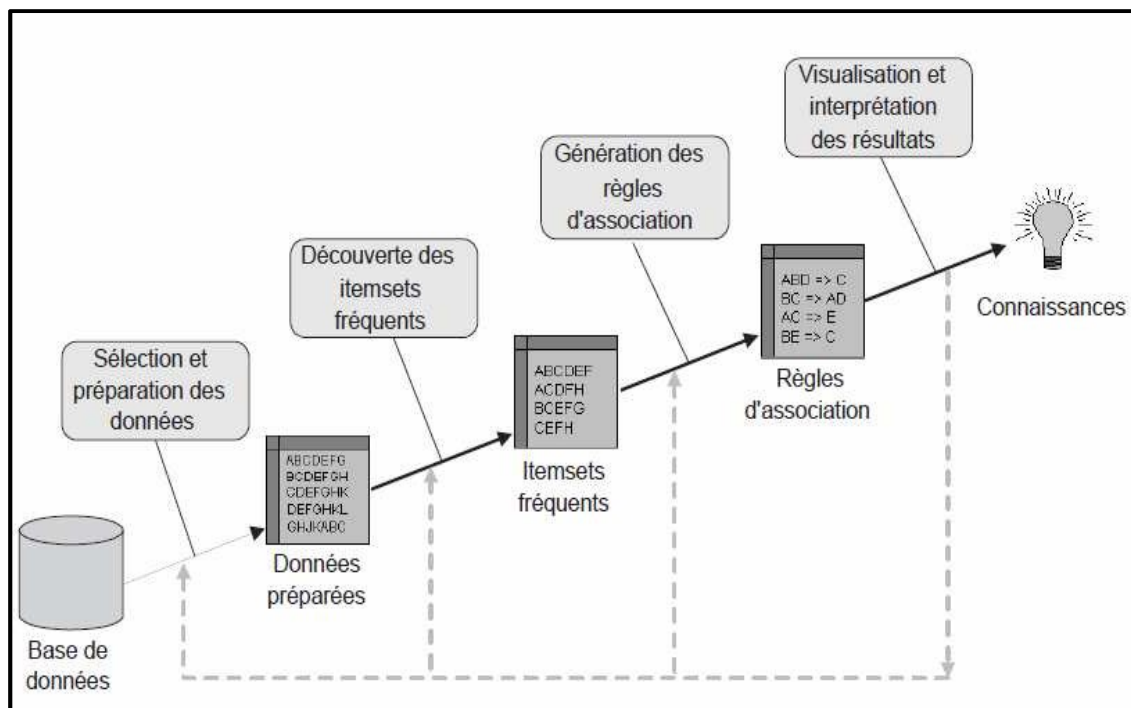


Figure 5 Etapes du Processus de Génération des Règles d'Association

Définition des règles d'association

La recherche des règles d'association est un procédé important dans le Data Mining. Plusieurs algorithmes de recherche des règles d'association existent et permettent de découvrir des relations d'intérêt entre deux ou plusieurs variables stockées dans de très grandes bases de

³⁹ Agrawal, T. Imieliński and A. Swami. "Mining association rules between sets of items in large databases". In *Proceeding SIGMOD '93 Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207-216, New York, NY, USA, 1993.

données. Nous avons par exemple les algorithmes Apriori, FP-growth, et AprioriTID [48].

La plupart des algorithmes d'extraction des règles d'association mettent en œuvre deux propriétés le support et la confiance. On parlera de ces critères et des mesures, mais aussi de l'algorithme d'exploration des données permettant d'extraire les règles d'association ultérieurement [49].

Dans ce chapitre on présentera un état de l'art sur les règles d'association, ainsi que les techniques d'extraction, les différents algorithmes et leurs variantes.

Notions de base sur les règles d'association

Représentation des données

Les données issues des différents documents et bases de données transactionnelles peuvent être représentées sous la forme d'une matrice booléenne à deux dimensions. Dans une telle base, chaque tuple représente une transaction tandis que les différents champs correspondent aux objets inclus dans la transaction. On note par N_e le nombre de transactions, par p le nombre d'articles, par 0 l'évènement d'absence de chaque article et par 1 sa présence dans la transaction. De ce fait on construira une matrice binaire de la base de données [50].

Transactions	Article1	Article2	Article3	Article4	Article5	Article6	Article7
T1	0	1	1	1	0	1	1
T2	0	0	0	1	1	1	0
T3	1	0	1	0	0	0	1
T4	0	0	0	0	1	1	1
T5	0	1	1	1	0	0	0

Tableau 1 Base de données binaire [50].

Dans le tableau 1, on retrouve la représentation binaire d'une base de données. Par exemple

la transaction T_1 , contient les articles 2, 3, 4, 6, 7.

Ce tableau représente une matrice creuse a deux dimensions $N_e * P$. Avec $N_e = 5$ (Le nombre de transactions), et $P = 7$. (Le nombre d'items).

Une transaction T représente un sous ensemble E .

Notre exemple (**Tableau 1**) donne ceci :

$T_1 = \{\text{Article2}, \text{Article3}, \text{Article4}, \text{Article6}, \text{Article7}\}$

Item et support

Un **Item** est un objet, élément ou un article d'une base de données.

Exemple 2 : Article3 représente un item.

Un **Itemset** est un ensemble d'items, d'objets ou d'articles d'une base de données.

Exemple : { item2, item3, item4, item6 }

Un **K-Itemset** est un ensemble de k éléments, ou k -Items, il est aussi un Itemset.

Exemple 1: {item2, item3, item4, item6} représente un 4-Itemset.

Support d'un Itemset

Le **support d'un Itemset** représente le nombre total des transactions d'une base de données comportant cet Itemset divisé par le nombre total des observations de cette base de données. Par exemple, soit une base de données D et soit X un Itemset de n éléments. Dans une base de données transactionnelle D , le support de l'itemset X est le nombre de transactions dans D incluant X , divisé par le nombre total des transactions de D .

$$\text{Support}(X) = \frac{\text{card}X}{\text{card}D}$$

Équation 1 Support d'un d'un Itemset

Exemple 1 : Soit X un Itemset, avec $X = \{\text{Article2}, \text{Article3}\}$.

Soit D la base des transactions présentées précédemment dans le tableau 1.

$\text{Card}(X)$ est le nombre de transactions dans D , de tel que les Items Article2 et Article3 apparaissent simultanément dans chacune de ces transactions de D . Il est égal à 2 $\text{Card}(D)$ est le nombre total des transactions. Il est égal à 5.

Alors $\text{Support}(X) = 2/5$

Itemset Fréquent

On dit qu'un Itemset X est un Itemset fréquent si et seulement si le support associé à cet Itemset est supérieur à un support minimum défini par l'utilisateur [56].

Mesures de qualité de Règles d'association**Définition**

Une règle d'association est une application de la forme $X \Rightarrow Y$, qui exprime une corrélation de cooccurrence⁴⁰.

Il existe deux mesures importantes, le support et la confiance, la robustesse d'une règle d'association est déterminée grâce à ces deux métriques. Une règle d'association qui a un support faible va être observée rarement. La confiance mesure la pertinence de l'inférence dans une règle, par exemple plus grande est la mesure de confiance de la règle $X \Rightarrow Y$, plus cette règle sera pertinente⁴¹.

Support et confiance d'une règle d'association

Les notions de support et de confiance ont été identifiées lors des premières études de recherche des règles d'association menées par Hajek, Havel et Chytil 1966 en

⁴⁰ Plasse Marie, Ndeye Niang, Saporta Gilbert, Villeminot Alexandre, Leblon Laurent. "Méthodes de classification pour l'extraction de règles", CNAM Laboratoire CÉDRIC, 2008.

⁴¹ Blanchard Julien, Kuntz Pascale, Guillet Fabrice, Gras Régis. "Mesure de la qualité des règles d'association par l'intensité d'implication entropique", IRIN – École polytechnique de l'université de Nantes, 2002

l'occurrence la méthode GUHA ⁴²

Support d'une règle d'association

Le support d'une règle d'association s'exprime par le nombre de transactions qui contiennent les éléments de X et les éléments de Y divisé par le nombre total de transactions de la base des transactions.

Dans une base de données D, le support d'une règle d'association transactions qui contiennent X et Y divisé par le nombre total des transactions.

$$\text{Support } (X \Rightarrow Y) = \frac{\text{Card } X \cup Y}{\text{Card } D}$$

Équation 2 Support d'une règle d'association [62].

Exemple la règle d'association "Lait \Rightarrow Pain" littéralement, **Si Lait Alors Pain**.

Le support représente le nombre de transactions dans lesquelles on trouve les Items Lait et Pain, divisé par le nombre total des transactions.

Confiance d'une règle d'association :

La confiance d'une règle d'association s'exprime par le nombre de transactions qui contiennent la relation d'union entre la transaction X et la transaction Y divisé par le nombre des transactions qui contiennent la transaction X. ⁴³

La confiance d'une règle d'association est définie comme suit :

$$\text{Confiance } (X \cup Y) = \frac{\text{support } (XUY)}{\text{Support } (X)}$$

⁴² Tihi, Adil. "Mise à jour de la méthode GUHA", HEC Montréal, 2007.

⁴³ Benrhaim, Morched. "La lecture assistée par ordinateur : une étude exploratoire", Université du Québec à Trois-Rivières, 2015.

Équation 3 *Confiance d'une règle d'association [64]*

Exemple, la confiance d'une règle d'association " Lait \Rightarrow Pain ", est égale au support de la règle " Lait \Rightarrow Pain " divisé par le support de l'Item " Lait ".

Extraction des règles d'association

L'extraction des règles d'association consiste à extraire les règles dont le support et la confiance sont au moins égaux respectivement, à des seuils minimaux de support et de confiance définis par l'utilisateur. La plupart des approches proposées pour l'extraction des règles d'association reposent sur les quatre phases suivantes :

Préparation des données :

Cette phase consiste à sélectionner les données (attributs et objets), de la base de données, utiles à l'extraction des règles d'association et transformer ces données en un contexte d'extraction.

Extraction des ensembles fréquents d'attributs :

Cette phase consiste à extraire du contexte tous les itemsets fréquents, un itemset est fréquent si son support est supérieur ou égal au seuil minimal de support défini par l'utilisateur.

Génération des règles d'association :

Cette phase consiste à utiliser les itemset fréquents extraits de la phase précédente pour générer les règles d'association ayant la confiance supérieure ou égale à la confiance minimale.

Visualisation et interprétation :

Cette phase consiste à la visualisation des règles d'association extraites afin d'en déduire des connaissances utiles pour l'amélioration de l'activité concernée.

Algorithme D'extraction Des Règles D'association

Une approche naïve pour l'extraction des motifs fréquents consiste à parcourir

l'ensemble de tous les motifs, à calculer leurs nombres d'occurrences (support) et à ne garder que les plus fréquents. Malheureusement, cette approche est trop consommatrice en temps et en ressources. En effet, le nombre de motifs est 2^p (p est le nombre de propriétés), et en pratique, on veut manipuler des bases ayant un grand nombre d'attributs.⁴⁴

Algorithme Apriori⁴⁵

L'algorithme Apriori proposé par Agrawal et ses co-auteurs en 1994 est un algorithme de base qui permet d'extraire des motifs fréquents dans une base ayant plusieurs milliers d'attributs et plusieurs millions d'enregistrements. L'idée est d'effectuer une extraction par niveaux selon le principe suivant :

- *On commence par chercher les motifs fréquents de longueur 1 ;*
- *On combine ces motifs pour obtenir des motifs de longueur 2 et on ne garde que les fréquents parmi eux ;*
- *On combine ces motifs pour obtenir des motifs de longueur 3 et on ne garde que les fréquents parmi eux ;*
- *continuer jusqu'à la longueur maximale.*

Cette approche s'appuie sur les deux principes fondamentaux suivants (qui reposent sur la décroissance du support) :

- 1) *Tout sous-motif d'un motif fréquent est fréquent.*
- 2) *Tout sur-motif d'un motif non fréquent est non fréquent*

Exemple de treillis des itemsets

Le nombre d'itemsets fréquents qui peuvent être générés de n items est de 2^n , la génération des Itemsets fréquents est de complexité exponentielle, il est alors essentiel de trouver la méthode de recherche la plus optimale. Ces Itemsets représentent un treillis d'Itemsets représenté sous la forme d'un diagramme de Hasse présenté par la suite :

⁴⁴ www.Djeffalabdahamid.com

⁴⁵ www.Djeffalabdahamid.com

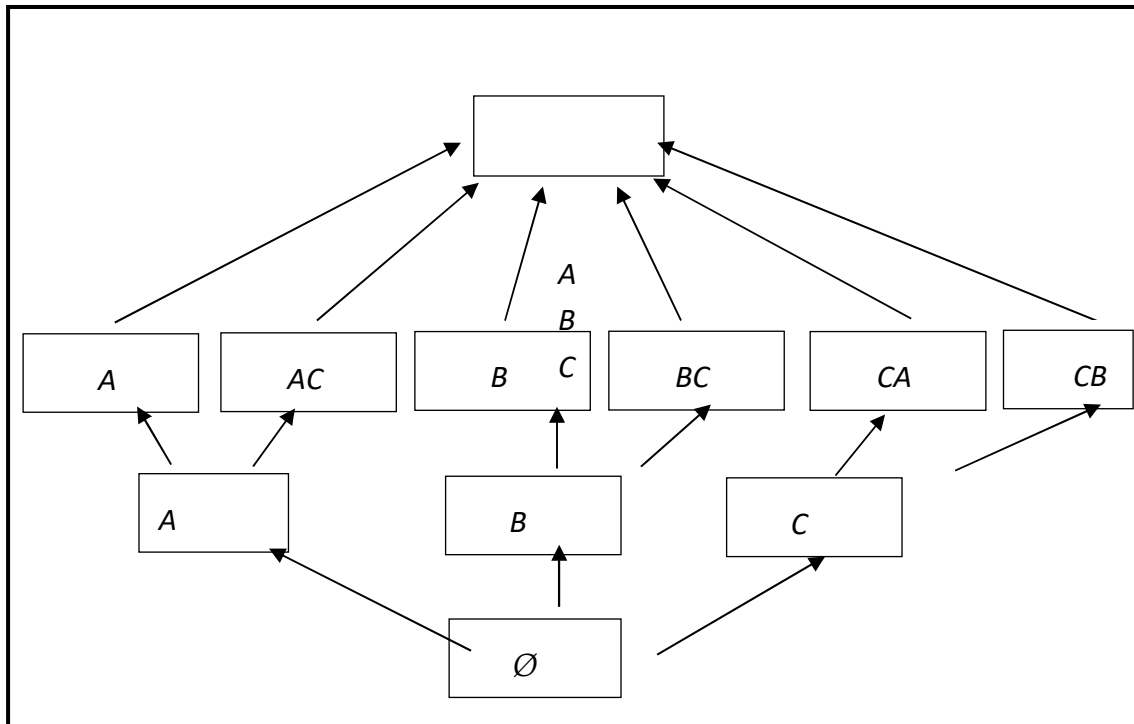


Tableau 2 Exemple de treillis d'Itemsets

Fonctionnement de l'algorithme Apriori :

D'une manière plus concise, le déploiement de l'algorithme Apriori se fait commesuit :

1. Générer les Règles candidates.
2. Calculer le support pour chaque règle candidate.
3. Apparier les règles dont on a calculé le support avec le support choisi.
4. On rejette les candidats dont le support est inférieur au $supp_{min}$.
5. On termine en sortie avec toutes les règles dont le support est supérieur au $supp_{min}$.

Entrées: base de données de transactions D , Seuil de support minimum σ

Sorties: Ensemble des items fréquents

$i \leftarrow 1$

$C_1 \leftarrow$ ensemble des motifs de taille 1 (un seul item)

tantque $C_i \neq \emptyset$ **faire**

Calculer le Support de chaque motif $m \in C_i$ dans la base

$F_i \leftarrow \{m \in C_i \mid support(m) \geq \sigma\}$

$C_{i+1} \leftarrow$ toutes les combinaisons possibles des motifs de F_i de taille $i + 1$

$i \leftarrow i + 1$

fin tantque

retourner $\bigcup_{(i>1)} F_i$

Avantages de l'Algorithme Apriori

Il existe une multitude d'avantages dans l'utilisation de l'algorithme Apriori. On en énumère quelques-uns⁴⁶:

- La découverte rapide de règles d'association pertinentes entre objets.
- La facilité d'interprétation des résultats lors de l'extraction des règles d'association, malgré le nombre important de ces dernières.

Inconvénients de l'Algorithme Apriori

Les inconvénients auxquels on fait face lors d'une utilisation de l'algorithme Apriori sont les suivants⁴⁷:

- Les algorithmes d'extraction liés à l'approche support / confiance génèrent un grand nombre de règles d'association.
- Un nombre important de configurations d'items ne peuvent pas engendrer de règles d'association.
- La recherche de règles d'association impose un temps considérable qui peut s'avérer désavantageux si l'on fait face à une énorme base de données.

OPTIMISATIONS**L'algorithme AprioriTID 45**

L'algorithme Apriori nécessite N passes sur la base, N étant la taille du plus grand ensemble susceptible d'être fréquent. Une optimisation possible consiste à générer en mémoire, pendant la première passe, les identifiants (TID) des transactions pour chaque 1-itemset (ensemble de motifs de taille 1) fréquent. Dans la suite, les listes de TID correspondant à chaque k -itemset sont gardées. Le calcul d'un k -itemset se fait toujours à partir des deux $(k-1)$ -itemsets contenant un élément de moins, mais le comptage se fait simplement par intersection des deux listes de TID des deux $(k-1)$ -itemsets source. Nous construisons la liste des TID après avoir déterminé les 1-itemsets fréquents, ce qui est plus efficace en taille

⁴⁶ Achouri Abdelghani, . "Extraction de relations d'association maximales dans les textes : représentation graphique", Université du Québec a Trois-Rivières, 2012.

⁴⁷ Bahri Emna, Lallich Stéphane. "Introduction de l'élagage pour l'extraction de règles d'association de classe sans génération de candidats", Laboratoire ÉRIC, Université de Lyon., QDC 2009.

mémoire mais nécessite deux passes. La première passe permet d'éliminer les produits non fréquents et réduit donc les listes de TID en mémoire construites dans une deuxième passe. Générer les listes de TID en mémoire en parallèle dès la première passe est plus efficace. Cependant, il n'est possible d'éliminer des listes qu'après le comptage total de la base, lorsqu'on sait qu'un 1-itemset ne sera pas fréquent. L'algorithme nécessite alors une taille mémoire de $N \cdot P$ TID, N étant le nombre de transactions et P le nombre de produits. Il devient très inefficace si les listes ne tiennent pas en mémoire.

L'algorithme apriori partitionné

L'algorithme proposé par Savasere permet de résoudre le problème de place mémoire de l'algorithme précédent. L'idée est simplement de diviser la base en Q partitions, de sorte que chaque partition tienne en mémoire. Chaque partition est traitée indépendamment et les ensembles fréquents sont découverts pour chaque partition. La validité de l'algorithme est basée sur le lemme suivant : pour être globalement fréquent, un ensemble doit être fréquent dans au moins une partition. Ayant obtenu les ensembles fréquents sur chaque partition, il suffit dans une passe finale d'explorer l'union des ensembles fréquents sur la base. L'opération se fait par un comptage simple sur la base. Les ensembles non globalement fréquents, mais localement fréquents dans une partition, sont ainsi éliminés. L'avantage de cet algorithme est qu'il nécessite deux passes au plus. Il est aussi facilement parallélisable, les partitions pouvant être traitées indépendamment. Dans ce cas, il faut cependant beaucoup de mémoire pour tenir les partitions et listes de TID en mémoire.

Algorithme de comptage dynamique⁴⁸

L'algorithme DIC a été proposé par Brin et al. Pour réduire le nombre de parcours de la base de données. DIC partitionne la base de données en blocs de M transactions. Durant le calcul des supports des k -itemsets, après le parcours d'une partition de taille M de D , on vérifie les k -itemsets candidats qui ont déjà atteint le support minimum, DIC les utilise alors pour générer des candidats de taille $(k+1)$, et

⁴⁸ Cours DJEFFAL Abdelhamid 2020-2021

commence à compter leurs supports. Ainsi les supports de candidats de tailles différentes sont calculés durant les mêmes parcours de D . En contrepartie de la diminution du nombre de balayages de la base de données, DIC considère des itemsets candidats de tailles différentes simultanément. Ceci pose le problème de stockage des itemsets candidats traités simultanément et du coût de calcul des supports des candidats qui est plus important que pour Apriori.

Algorithme FP-growth⁴⁹

A l'opposé d'Apriori qui génère des itemsets candidats et qui les teste pour ne conserver que les itemsets fréquents, FP-Growth (Han et al. (2000)) construit les itemsets fréquents sans génération de candidats. En fait, il utilise la stratégie "diviser et dominer" (divide- and-conquer). Tout d'abord, il compresse les itemsets fréquents représentés dans la base de données à l'aide d'un FP-Tree (frequent-pattern tree) dont les branches contiennent les associations possibles des items.

Chaque association peut être divisée en fragments (pattern fragment) qui constituent les itemsets fréquents. La méthode FP-Growth transforme le problème de la recherche de l'itemset fréquent le plus long par la recherche du plus petit et sa concaténation avec le suffixe correspondant (le dernier itemset fréquent de la branche aboutissant à l'item considéré). Ceci permet de réduire le coût de la recherche.

Avantages et inconvénients de l'algorithme FP-growth

Avantages

- La principale force de l'algorithme FP-Growth est le fait que l'algorithme ne fait que deux balayages de la base de transactions. Le premier balayage pour trouver les k -itemsets et pour construire la liste des items fréquents et le second pour construire le squelette de l'arbre FP-tree.
- L'algorithme est considéré comme étant complet, la structure étant constituée de toutes les informations sur les éléments fréquents

⁴⁹ Cours DJEFFAL Abdelhamid 2020-2021

- *La structure contient uniquement les objets fréquents classés par ordre de fréquence décroissante.*

Inconvénients

- *Dans beaucoup de cas d'utilisation, on rencontre des bases de transactions beaucoup trop volumineuses, ce qui bloque le processus de recherche puisque les ressources en mémoire sont insuffisantes pour accueillir toute la structure.*

La construction de l'arbre FP-Tree peut être très longue du fait de l'utilisation de beaucoup de ressources de calcul⁵⁰.

Conclusion

L'extraction des règles d'association est une méthode descriptive de fouille de donnée qui a reçu beaucoup d'intérêt de la part des chercheurs. On peut la définir comme étant la recherche de relation entre des Items dans un ensemble de données.

Elles peuvent être appliquées à divers secteurs d'activité pour lesquels il est intéressant de rechercher des corrélations potentielles entre des objets de diverses catégories. Prenons par exemple le domaine médical, on pourrait chercher des complications dues à des associations de produits pharmaceutiques, ou dans le domaine de Marketing dans le cas de l'analyse du panier de ménagère, l'extraction des règles d'association permet d'analyser les tickets de caisse des clients particuliers afin de comprendre leurs habitudes de consommation .

Ce présent chapitre nous a permis de découvrir le concept des règles d'association. Nous avons vu que les premiers algorithmes traitant la découverte des règles d'association étaient développés, en premier lieu, dans un objectif de marketing. Actuellement, les règles d'associations trouvent tout leur intérêt dans tout processus de prise de décision.

Plusieurs algorithmes existent pour l'extraction des règles d'association, Ils se distinguent par leurs stratégies d'énumération de l'espace de recherche des Itemsets ainsi que par la méthode de calcul du Support.

⁵⁰ *Shashikumar G. Totad, Geeta R. B, Prasad Reedy. "Batch Processing for Incremental FP-tree Construction", international Journal of Computer Applications, 2010*

Nous avons pu constater que ces algorithmes diffèrent également par les structures de données qu'ils utilisent, que ce soit pour la représentation des itemsets candidats ou pour la représentation de la base de transactions

Dans le chapitre 3, nous présentons notre conception dans laquelle nous utiliserons les notions introduites dans ce chapitre.

CHAPITRE III

CONCEPTION

1 Introduction :

Dans ce chapitre nous aborderons une description générale de notre système, en mettant en évidence son côté conceptuel du prétraitement qui constitue une étape fondamentale avant la mise en œuvre de notre système, ensuite nous détaillerons chaque phase en citant les principaux algorithmes et techniques utilisées dans chacune des dites phases.

Description de notre travail

Notre système se base sur le concept des règles d'association, L'une des applications typiques de l'extraction des règles d'association est l'analyse du panier de la ménagère. Elle consiste à rechercher des corrélations entre les produits à travers les achats des clients afin de comprendre leurs habitudes de consommation, organiser les promotions, gérer les stocks, agencer les rayons du magasin, etc. dernièrement, cette technique est utilisée dans les domaines cherchant à regrouper des objets entre eux par le degré de similarité, tels que le domaine médical, CRM, accidentologie, industrie, les réseaux sociaux, texte mining etc .

Dans ce but nous avons choisi la technique d'extraction de règle d'associations qui se base sur l'algorithme de basa APRIORI et l'algorithme optimisé Apriori TID .

L'idée générale de ces deux algorithmes est la même, en supposant que les attributs sont triés dans un ordre lexicographique. En effet, au début les supports des 1-itemsets sont calculés, en effectuant une première passe sur la base de données. Les itemsets, dont le support n'a pas atteint le seuil minsupp, sont écartés, i.e., ils sont considérés comme non fréquents. Par la suite, un nouvel ensemble des itemsets, dits ensemble des itemsets candidats, est généré. Après une autre passe à travers la base, les supports des itemsets sont calculés.

Les itemsets non fréquents sont écartés et le processus décrit précédemment est relancé, jusqu'au moment où l'on ne peut plus générer d'itemsets fréquents.

Ces algorithmes ont été qualifiés d'algorithmes rapides [⁵¹], puisque leur méthode de génération des ensembles candidats n'est pas de complexité combinatoire. En effet avec les

⁵¹ R. Agrawal and R. Skirant. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Databases, pages 478—499, June 1994.

algorithmes Apriori et AprioriTid, les itemsets candidats sont calculés en utilisant seulement les itemsets fréquents trouvés lors la passe précédente et non pas tous les itemsets de la base transactionnelle.

L'algorithme AprioriTid possède une propriété qui le distingue de l'algorithme Apriori. En effet, dans l'AprioriTid, dès la deuxième passe, la base de données n'est plus utilisée pour calculer les supports des ensembles candidats. On utilise plutôt un ensemble intermédiaire déduit à partir de l'ensemble des itemsets candidats, générés dans la passe précédente qu'on note C_k , donc pour $k = 1$, C_1 correspond à toute la base de données D . Dans les dernières passes, la taille de C_k peut devenir plus réduite que celle de la base de données, ainsi on économisera au niveau de la lecture de la base D .

C'est ce que nous expliquerons en détail dans les éléments suivants.

Conception globale du système

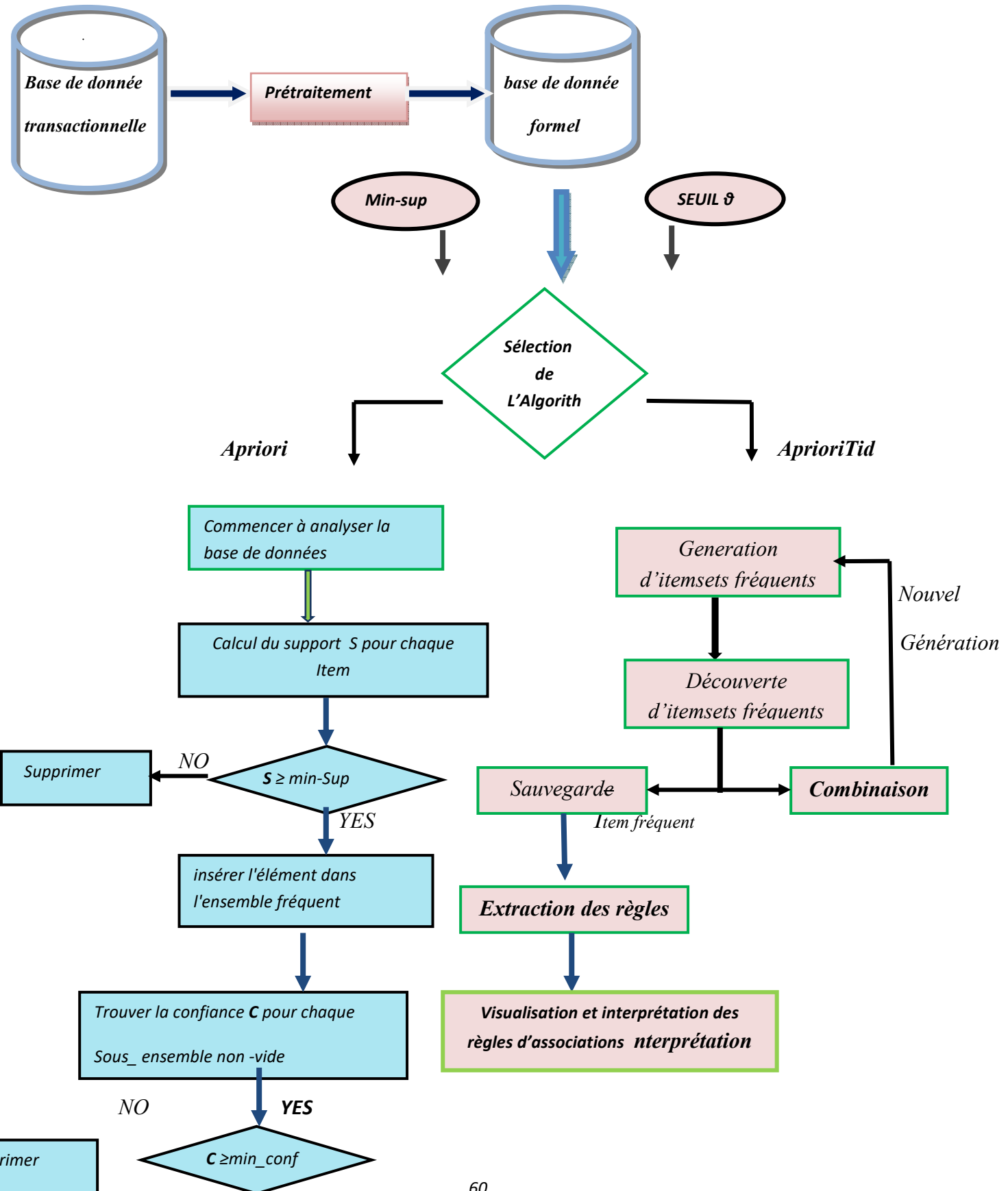
Notre système est un système d'analyse de paniers d'achats qui se base sur l'extraction des règles associatives à partir d'une base de données transactionnelles afin de trouver les motifs fréquents dans les bases de données massives en se basant sur l'algorithme Apriori et Apriori TID .

Le système prend en entrée une base de donnée transactionnelle et la transforme en une base de donnée formelle utilisable par la phase d'apprentissage. Cette transformation est appelée prétraitement.

L'étape suivante est l'extraction des Itemset fréquents et en les conserver après chaque génération des Itemsets.

Finalemnt en termine par la transformation de ces Itemsets fréquents en un ensembles de règle d'association.

La structure de notre système est représentée comme suite :



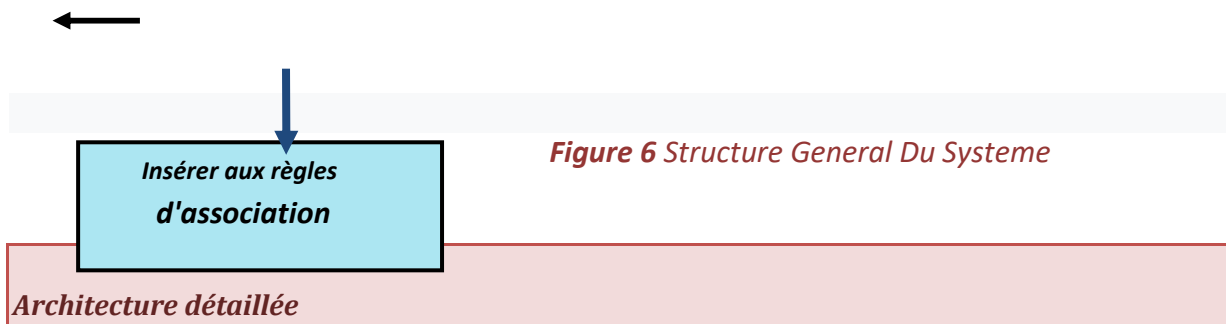


Figure 6 Structure General Du Systeme

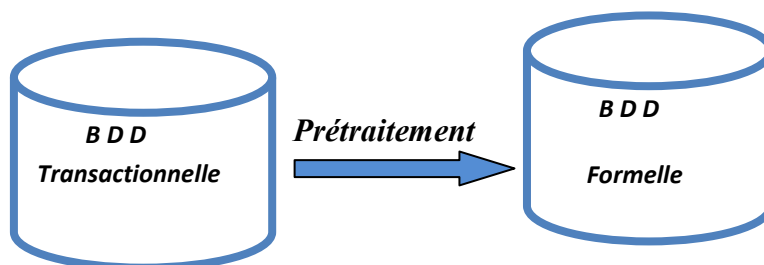
Ci-dessous, nous détaillons chacune des étapes que nous avons traversées dans notre système.

Prétraitement :

Ce module est consacré à la collecte et au prétraitement des données qui seront ensuite manipulées par les modules d'analyse et de représentation.

Notre objectif est d'extraire les règles d'association pour identifier les éléments fréquents dans un ensemble de données quels produits sont commandés plus fréquemment que d'autres et quels produits indépendants sont commandés suite à la commande d'un autre produit. L'algorithme Apriori aide à identifier ces règles d'association.

Pour les besoins de cette procédure nous utilisons la base "Groceries". Elle représente une collection d'items achetés dans une épicerie regroupés par panier (transaction). Chaque ligne représente un panier (une transaction) et chaque colonne représente un item acheté. Nous souhaitons analyser cet ensemble en vue d'extraire les règles solides pertinentes.



Sélectionner l'algorithme**Etapes Pour Effectuer Algorithme Apriori :**

Etape 1 : Scannez la base de données de transaction pour obtenir le support S de chaque itemset, ensuite comparez S avec min_sup , et obtenez un ensemble de k -itemsets fréquents .

Etape 2 : Analysez la base de donnée et calculer le support de chaque Itemset .

Si le support $S \geq \text{min-Sup}$ ALORS insérer l'élément dans l'ensemble fréquent /

Si non supprimer .

Etape 3 : trouvez la confiance C pour tous les sous-ensembles non_vides pour chaque ensemble d'éléments .

SI $C \geq \text{min_conf}$ Alors découvrir des règles d'association intéressantes les plus utiles entre les articles achetés.

Sinon rejeter (supprimer) les ensembles confiance $\leq \text{conf_min}$.

Etape 5 : Insérer aux règles d'association

Etapes pour effectuer Algorithme Apriori TID :

AprioriTID est une variante de l'algorithme **Apriori**. Il a été proposé dans le même article qu'Apriori comme implémentation alternative d'Apriori. Il produit la même sortie qu'Apriori. Mais il utilise un mécanisme différent pour compter la prise en charge des ensembles d'éléments.

L'entrée est une base de données de transaction (base formelle) et un seuil nommé minsup .

L'algorithme **AprioriTID** possède une propriété qui le distingue de l'algorithme **Apriori**. En effet, dans l'AprioriTid, dès la deuxième passe, la base de données n'est plus utilisée pour calculer les supports des ensembles candidats. On utilise plutôt un ensemble

intermédiaire déduit à partir de l'ensemble des itemsets candidats, générés dans la passe précédente qu'on note C_k , donc pour $k = 1$, C_1 correspond à toute la base de données D . Dans les dernières passes, la taille de C_k peut devenir plus réduite que celle de la base de données, ainsi on économisera au niveau de la lecture de la base D .

Génération des itemsets : Dans cette phase, tous les itemsets présentés dans la base de données formelle sont extraits du premier passage, de sorte que chaque itemset est représenté par une liste TID dont les cases contiennent les numéros ou les ID des transactions qui contiennent ce itemset.

Découverte d'itemsets fréquents : Après avoir extrait les itemsets, nous nous débarrassons tous les itemsets non fréquent, à partir de support donné au début, et on ne laissant que les itemsets fréquents pour cette génération

Sauvegarde : Nous conservons tous les itemsets fréquents pour chaque génération

Combinaison : Après avoir supprimé les listes dont la fréquence est inférieure ou égale au seuil, l'intersection entre les listes qui ont les mêmes numéros de client c'est-à-dire le ID d'un client commun se produit de sorte qu'au début une intersection se produit pour chacune avec chaque existant et le nom de la nouvelle listes (nouvel itemsets) résultant de l'intersection de deux séries devient la fusion des noms des deux listes précédentes.

Spécification fonctionnelle

La conception sera modélisée à l'aide de langage UML (Unified Modeling Language) qui est un langage de modélisation graphique à base de pictogrammes.

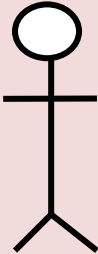
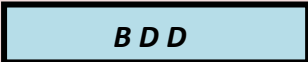
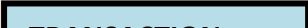
ACTEUR	DESCRIPTION
 <p>UTILISATEUR</p>	<p>L'administrateur a pour mission :</p> <ul style="list-style-type: none"> • Lancer l'extraction • Établir les statistiques
 <p>B D D</p>	<p>Contient un ensemble des transactions</p>
 <p>TRANSACTIONS</p>	

Figure 7 Diagramme des cas d'utilisation

Détermination des cas d'utilisation

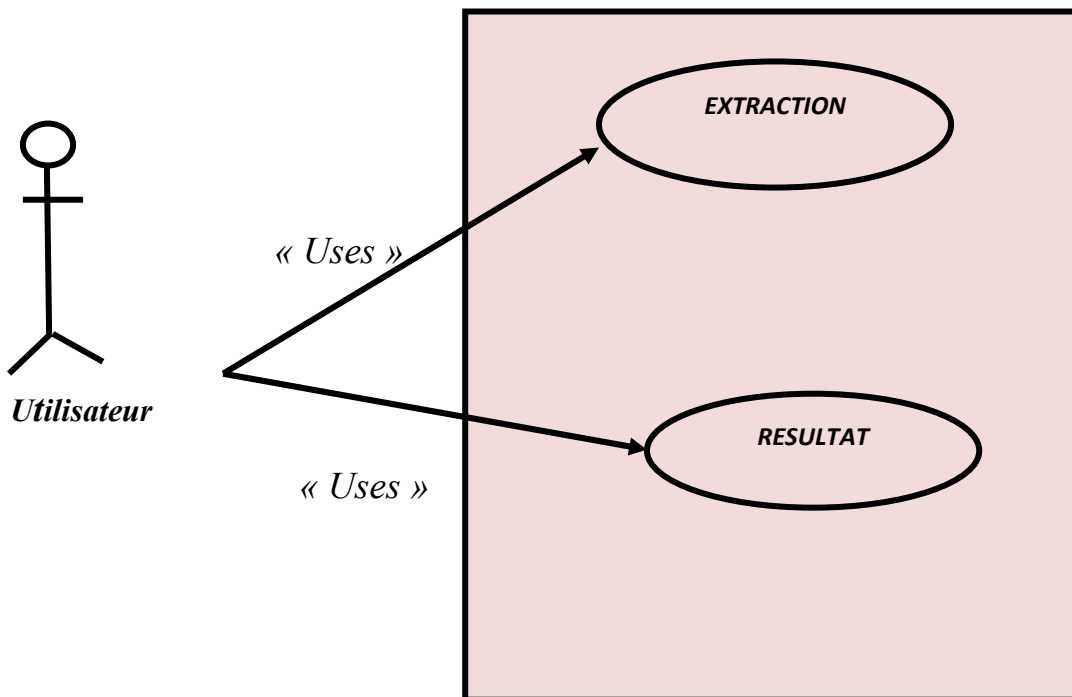


Figure 8 Diagramme de Cas d'utilisation de système

Description des cas d'utilisation

➤ Description de cas d'utilisation « Extraction » :

Titre : extraction des connaissances à partir d'une base de données.

Description : ce cas d'utilisation permet à l'administrateur de faire l'extraction des règles d'association. Lors de cette phase le système extrait les ensembles fréquents

Acteurs : administrateur ; transaction.

➤ *Description de cas d'utilisation « Résultat » :*

Titre : Résultat.

Description : Ce cas d'utilisation permet à l'utilisateur d'établir des résultats sur le système ; en effet cette phase nous permet d'afficher les différents résultats de l'extraction de programme.

Acteurs : administrateur ; base de transaction.

Diagramme de séquence de cas d'utilisation « Extraction »

Dans ce cas, on présente la séquence des scénarios du cas d'utilisation «Extraction ».

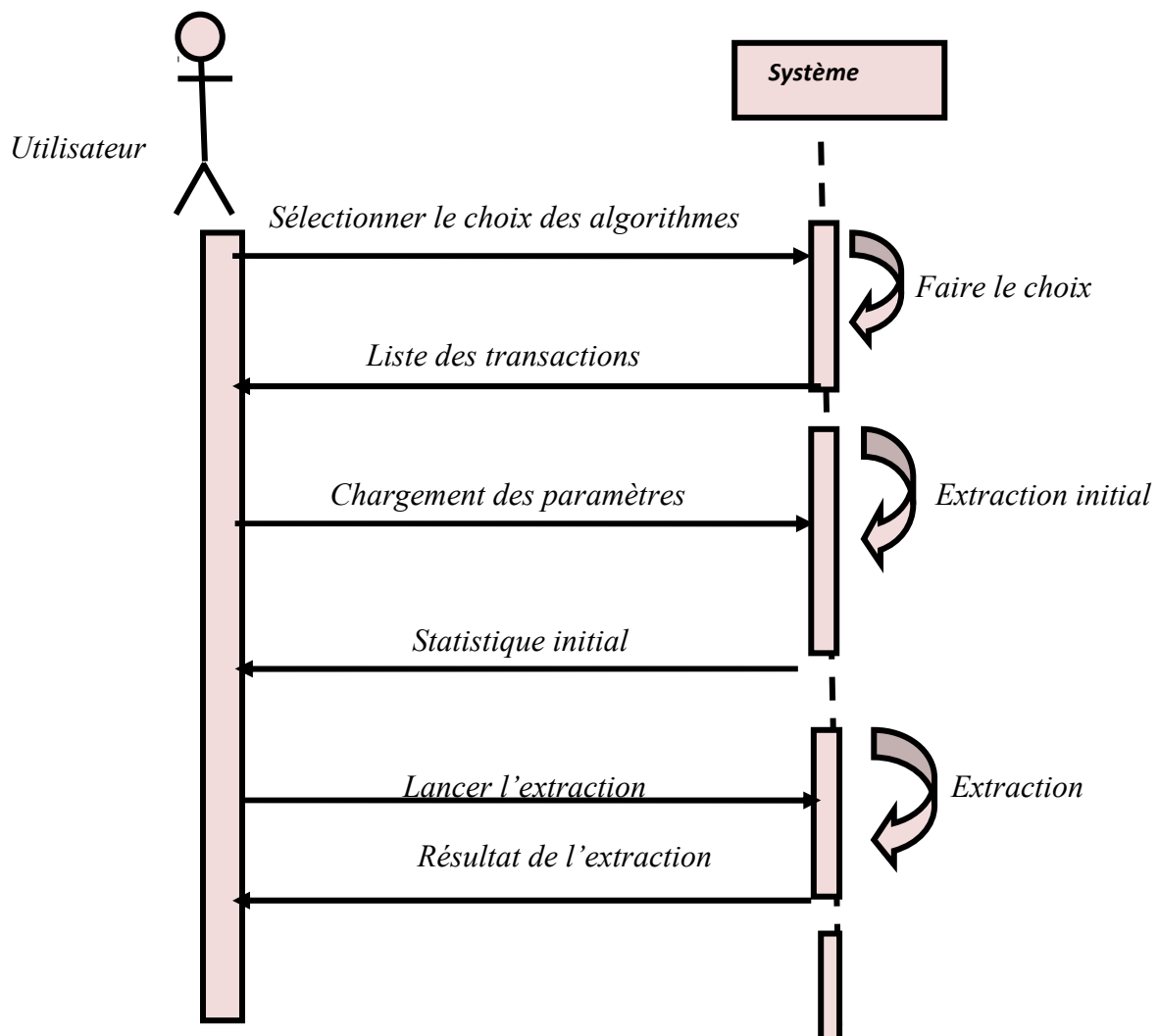


Figure 9 Diagramme de séquence de cas d'utilisation « Extraction ».

Description des scénarios :

➤ *Description de cas d'utilisation « Extraction » :*

- L'utilisateur sélectionne le choix de l'algorithme à effectuer.
- Le système envoie la liste de transaction.
- L'utilisateur charge les paramètres.
- Le système fait une extraction rend des statistiques initial.
- L'utilisateur lance l'extraction .
- Le système rend le résultat final de l'extraction.

Diagramme de séquence de cas d'utilisation « Extraction »

Dans ce cas, on présente la séquence des scénarios du cas d'utilisation «Extraction ».

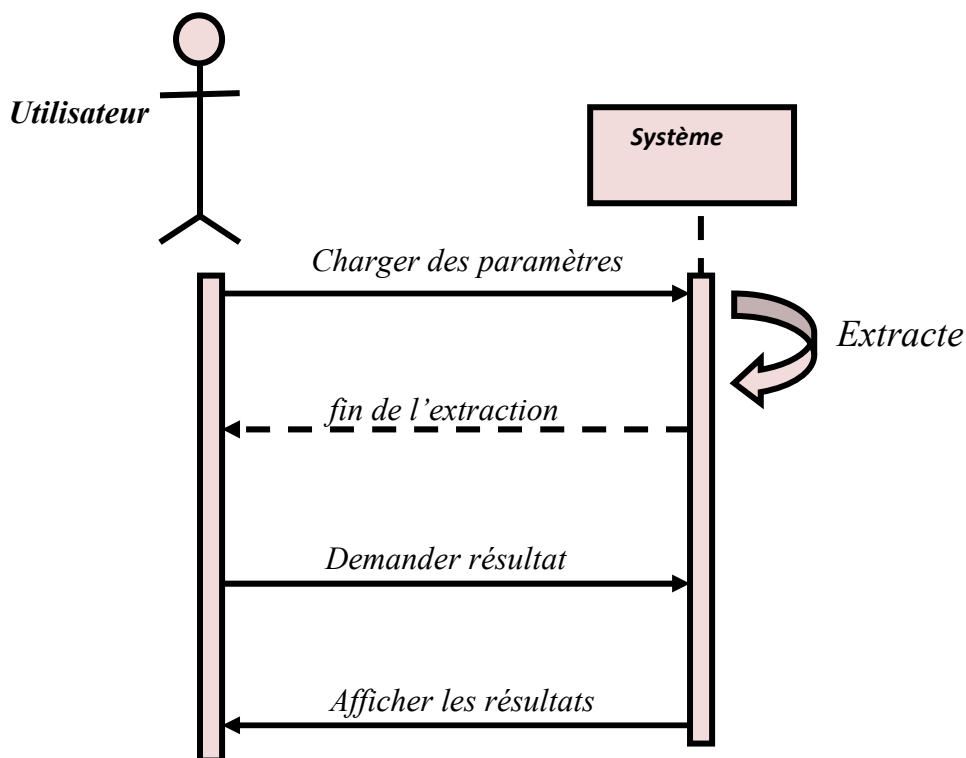


Figure 10 Diagramme de séquence de cas d'utilisation « Résultats ».

Description des scénarios :

➤ Description de cas d'utilisation « Résultats » :

L'utilisateur charge les paramètres.

Le système extrait des résultats.

L'utilisateur demande les résultats.

Le système affiche les résultats.

Conclusion

Dans ce chapitre, nous avons essayé de présenter toute les étapes que nous avons prises pour concevoir l'extraction des règles d'association pour l'analyse les différentes transactions.

Nous avons, également, présenté les modèles correspondants sous forme des diagrammes UML.

Et dans le chapitre suivant, nous expliquerons le processus de mise en œuvre de notre application.

CHAPITRE IV
IMPLEMENTATION

1 Introduction

L'implémentation d'un système informatique, désigne sa réalisation et sa mise en œuvre, pour passer à l'interaction avec les utilisateurs. L'objectif de ce chapitre est de présenter les outils (les logiciels, les langages, les bibliothèques et les données utilisés dans notre système, ensuite nous aborderons l'application.

Outils utilisés

Logiciels utilisés

Dans notre environnement de travail nous avons utilisé **Netbeans** pour lancer les deux Algorithmes Apriori et AprioriTid pour le traitement, l'extraction des règles d'association et la validation des résultats que nous avons obtenus.

Netbeans



Figure 11 Interface Netbeans

NetBeans⁵² est un environnement de développement intégré (EDI) un environnement libre et facile à utiliser. Il supporte plusieurs langages de programmation à savoir JAVA, C++, PHP, et bien d'autres. Il offre toutes les facilités d'un IDE moderne. Il fournit plusieurs outils tels qu'un éditeur de texte doté d'un pré-compilateur avancé (éditeur en couleurs, projets multilingage, refactoring, éditeur graphique d'interfaces et de pages Web), un gestionnaire de projets. Ainsi que des outils de débogage et de test des programmes. C'est un outil qui facilite énormément la phase de développement et des tests.

1. ⁵² <https://www.theserverside.com/definition/NetBeans>

Compilé en Java, NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X ou sous une version indépendante des systèmes d'exploitation (requérant une machine virtuelle Java). Un environnement Java Développement Kit JDK est requis pour les développements en Java.

Scence builder

Scene Builder ^[53] est un outil interactif de conception d'interface graphique pour JavaFX. Créé par Oracle, il permet de construire rapidement des interfaces utilisateurs sans avoir besoin de coder.

Les utilisateurs peuvent faire glisser et déposer des composants d'interface utilisateur dans une zone de travail, modifier leurs propriétés, appliquer des feuilles de style et le code FXML de la mise en page qu'ils créent est automatiquement généré en arrière-plan. Le résultat est un fichier FXML qui peut ensuite être combiné à un projet Java en liant l'interface utilisateur à la logique de l'application.



Figure 12 Interface de SceneBuilder

Java

Java ⁵⁴est un langage de programmation à usage général, évolué et orienté objet dont la syntaxe est proche du C. Ses caractéristiques ainsi que la richesse de son écosystème et de sa

^[53] <https://github.com/gluonhq/scenebuilder/wiki>

⁵⁴ <https://www.futurasciences.com/tech/definitions/internet-java-485/>

communauté lui ont permis d'être très largement utilisé pour le développement d'applications de types très disparates. Java est notamment largement utilisée pour le développement d'applications d'entreprises et mobiles.



Figure 13 Interface de Java

JavaFX

JavaFX est un ensemble de packages graphiques et multimédias qui permet aux développeurs de concevoir, créer, tester, déboguer et déployer des applications clientes riches qui fonctionnent de manière cohérente sur diverses plates-formes^[55].

JavaFX est une famille de produits et de technologies de Sun Microsystems qui appartient à Oracle. Les produits JavaFX ont pour but de créer des applications internet riches (RIA). Actuellement JavaFX est constitué de JavaFX Script et de JavaFX Mobile, bien que d'autres produits soient prévus.



Figure 14 interface de JavaFx

Base de données utilisé

^[55] <https://docs.oracle.com/javafx/2/overview/jfxpub-overview.htm>

Puisque nous sommes dans un système qui utilise l'extraction des règles d'association, nous avons besoin d'une base de données pour l'entraîner sur les données en utilisant des algorithmes pour prédire des liens entre les différentes transactions.

	A	B	C	D	E	F	G	H	I	J	K	L
1	citrus fruit,semi-finished bread,margarine,ready soups											
2	tropical fruit,yogurt,coffee											
3	whole milk											
4	pip fruit,yogurt,cream cheese ,meat spreads											
5	other vegetables,whole milk,condensed milk,Long life bakery product											
6	whole milk,butter,yogurt,rice,abrasive cleaner											
7	rolls buns											
8	other vegetables,UHT-milk,rolls buns,bottled beer,liquor (appetizer)											
9	pot plants											
10	whole milk,cereals											
11	tropical fruit,other vegetables,white bread,bottled water,chocolate											
12	citrus fruit,tropical fruit,whole milk,butter,curd,yogurt,flour,bottled water,dishes											
13	beef											
14	frankfurter,rolls buns,soda											
15	chicken,tropical fruit											
16	butter,sugar,fruit/vegetable juice,newspapers											
17	fruit/vegetable juice											
18	packaged fruit/vegetables											
19	chocolate											
20	specialty bar											
21	other vegetables											
22	butter milk,pastry											
23	whole milk											
24	tropical fruit,cream cheese ,processed cheese,detergent,newspapers											
25	tropical fruit,root vegetables,other vegetables,frozen dessert,rolls buns,flour,sweet spreads,salty snack,waffles,candy,bathroom cleaner											
26	bottled water,canned beer											
27	yogurt											
28	sausage,rolls buns,soda,chocolate											
29	other vegetables											
30	brown bread,soda,fruit/vegetable juice,canned beer,newspapers,shopping bags											
31	yogurt,beverages,bottled water,specialty bar											
32	hamburger meat,other vegetables,rolls buns,spices,bottled water,hygiene articles,napkins											
33	root vegetables,other vegetables,whole milk,beverages,sugar											
34	pork,berries,other vegetables,whole milk,whipped/sour cream,artif. sweetener,soda,abrasive cleaner											

Figure 15 Base de Données.csv

Pour les besoins de cette procédure nous utilisons la base "Groceries" . Elle représente une collection d'items achetés dans une 'épicerie regroupés par panier (transaction). Chaque ligne représente un panier (une transaction) et chaque colonne représente un item acheté. Nous souhaitons analyser cet ensemble en vue d'extraire les règles solides pertinentes.

Description de l'application

Dans cette partie nous présentons les interfaces de l'application « Apriori et Apriori-TID » ainsi les scénarios d'exécution pour montrer quelques résultats de tests de notre application.

La première interface

Cette interface est primordiale, c'est la phase de choisir l'algorithme pour l'utilisation.



Figure 16 Interface d'exécution

Vous pouvez choisir l'algorithme en cliquant simplement sur le bouton qui contient le nom de l'algorithme.

L'interface Apriori

Interface d'apprentissage avec la méthode apriori.

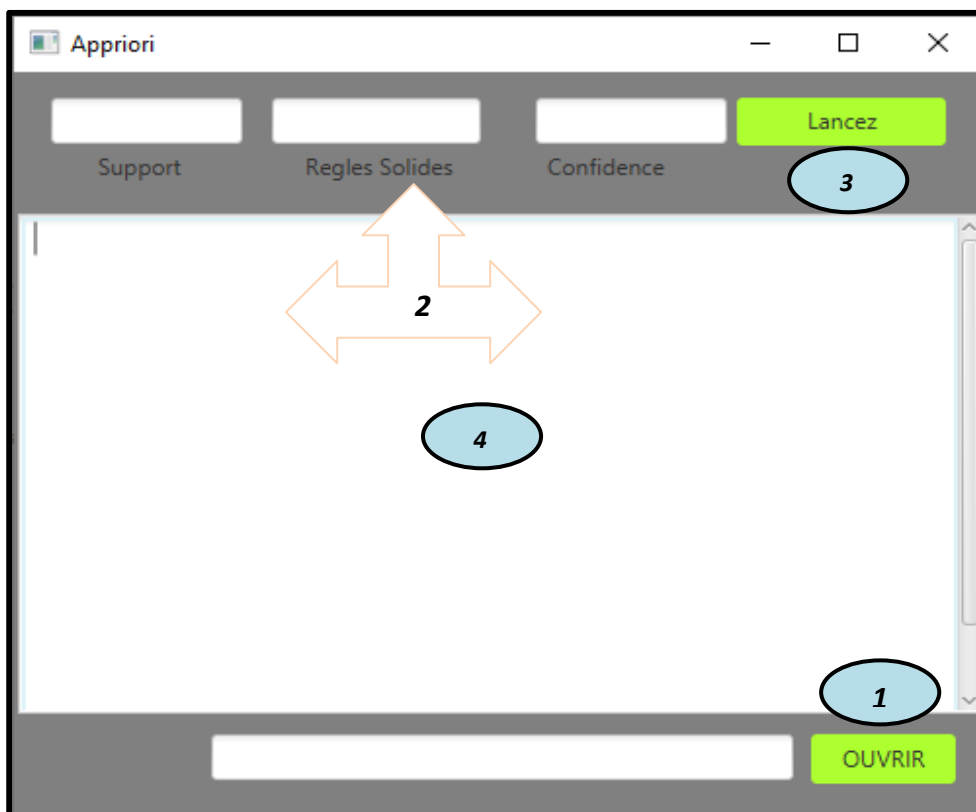


Figure 17 L'interface d'exécution de l'algorithme Apriori

1. Le bouton OUVRIR : qui nous permet de choisir la base de données et à travers elle nous allons extraire les règles d'association.
2. Les 3 espaces sont réservés pour mentionner le support minimum, la confiance minimale et le nombre des règles d'association.
3. Bouton lancez pour Exécution de l'apprentissage.
4. La grande zone dédiée à montrer les résultats après le traitement de la base de données.

L'interface d'Apriori-TID

Interface d'apprentissage avec la méthode AprioriTid

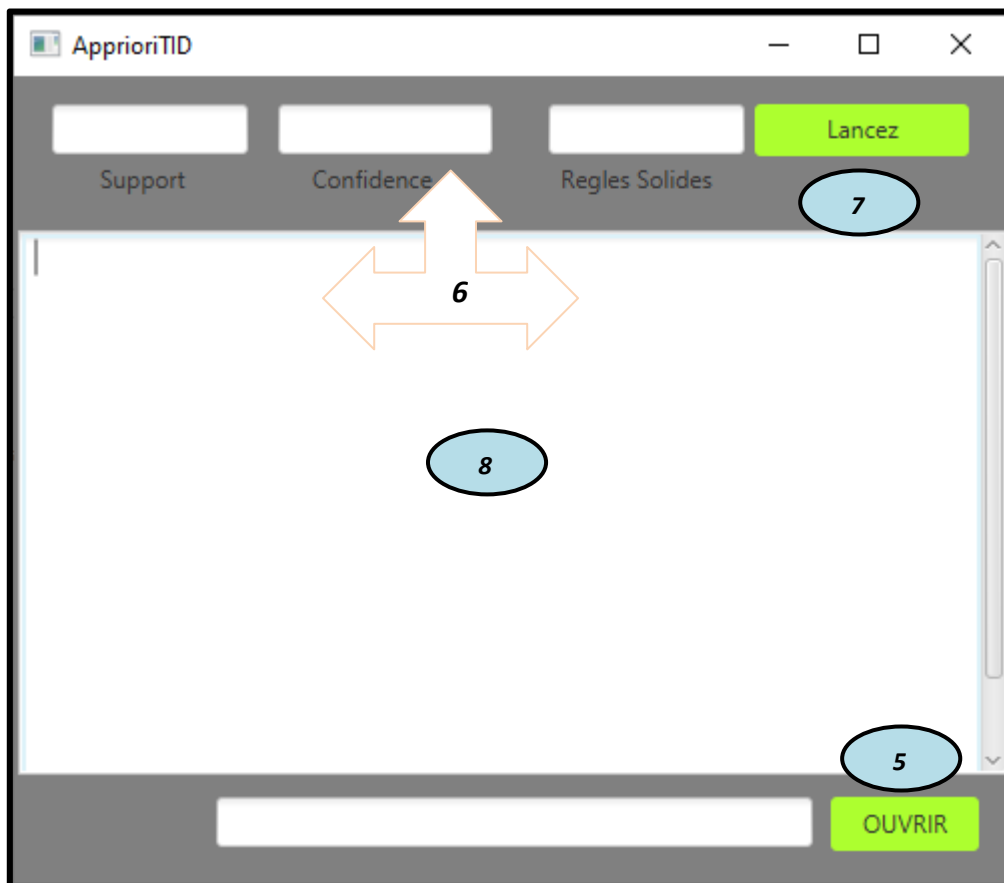


Figure 18 L'interface d'exécution de l'algorithme AprioriTID

5. Le bouton OUVRIR : qui nous permet de choisir la base de données à partir de laquelle nous allons extraire les règles d'association.
6. Les 3 espaces sont réservés pour mentionner le support minimum, la confiance minimale et le nombre des règles d'association.

7. Bouton lancez pour Exécution de l'apprentissage.
8. La grande zone dédiée à montrer les résultats après le traitement de la base de données.

Description de code

Le programme est divisé en plusieurs classes, et chaque classe a un rôle spécifique dans la mise en œuvre du processus d'apprentissage, et tout cela est expliqué dans le tableau suivant :

Classe	Signification
<i>Application.fxml</i>	<i>code XML contenant l'interface de contrôle et les formalités de saisie des informations de l'algorithme apriori-TID (zone pour saisie le lien de base de données, support minimal, confiance minimal de la règle et nombre des règles solide afficher sur l'écran)</i>
<i>ApplicationController.java</i>	<p><i>Le code Java contient le code de base (Controller) à partir de la lecture et de la collecte des itemsets de la base de données, puis de l'extraction de celles fréquentes et enfin de l'extraction des règles d'association qui est relient avec l'interface <i>Application.fxml</i>.</i></p> <p><i>Il contient deux classe importantes, la première classe est appelée Concatinate, de sorte que chaque itemset fréquente devient un objet et est liée à son propre support.</i></p> <p><i>La deuxième classe est appelée regle, qui est chargé d'extraire les règles d'association résultant des itemsets fréquentes par l'assistance de la première classe concatinate, calculez ensuite la confiance de chaque règle et organisez son affichage à l'écran par ordre décroissant, qui va du confiance la plus fort au plus faible, lié par la confiance minimum saisi et le nombre de règles d'association qui doivent être affichées à l'écran.</i></p> <p><i>Ce code est principalement basé sur une liste imbriquée (nestedliste), c'est-à-dire une liste qui contient un groupe de listes.</i></p>
<i>AprioriTidApplication.java</i>	<i>Main de la largeur de l'interface d'entrée les informations initial.</i>

<i>FXMLDocument.fxml</i>	<i>code XML qui contient les formalités de la première interface.</i>
<i>FXMLDocumentController.java</i>	<i>Le code permettant de relier l'interface de saisie au code principal (Controller), afin de respecter les données saisies au début.</i>
<i>Main.css</i>	<i>Spécial pour l'esthétique de nos interfaces d'application.</i>
<i>Transf.java</i>	<i>Le code responsable de la conversion de la base de données d'une base de données transactionnelle en une base de données formel.</i>
<i>Apriori.java</i>	<i>Le code responsable de l'algorithme apriori .</i>
<i>AprioriFXML.fxml</i>	<i>code XML contenant l'interface de contrôle et les formalités de saisie des informations de l'algorithme apriori-TID(zone pour saisie le lien de base de données,support minimal, confiance minimal de la règle et nombre des règles solide afficher sur l'écran)</i>
<i>AprioriFXMLController.java</i>	<i>Le code Java contient le code de base (Controller) à partir de la lecture et de la collecte des itemsets de la base de données, puis de l'extraction de celles fréquentes et enfin de l'extraction des règles d'association qui est relient avec l'interfaceAprioriFXML.fxml. Cette base sur la classe apriori.java qui est chargé d'extraire les règles d'association résultant des itemsets fréquentes, calculez ensuite la confiance de chaque règle et organisez son affichage à l'écran par ordre décroissant, qui va du confiance la plus fort au plus faible, lié par la confiance minimum saisi et le nombre de règles d'association qui doivent être affichées à l'écran.</i>

Tableau 3 Tableau des classes avec ses Signification

Description de code détaillé

Dans cette partie, nous parlerons du code en détail avec une explication des outils utilisateur et des bibliothèques qui ont aidé à coordonner et à configurer ce programme.

Les interfaces

Les bibliothèques

```

<?xml version="1.0" encoding="UTF-8"?>
<?import javafx.scene.control.Button?>
<?import javafx.scene.control.Label?>
<?import javafx.scene.layout.AnchorPane?>
<?import javafx.scene.paint.Color?>
<?import javafx.scene.paint.RadialGradient?>
<?import javafx.scene.paint.Stop?>
<?import javafx.scene.text.Font?>

```

Figure 19 Les bibliothèques de javafx utilisé dans l'interface JavaFx

Description du paquet `javafx.scene.control`

Elle représente la classe de base des contrôles de l'interface utilisateur tels que : `Accordion`, `ButtonBar`, `ChoiceBox`, `ComboBoxBase`, `HTML editor`, etc.

Description du Paquet `javafx.scene.paint`

Elle Fournit l'ensemble de classes pour les couleurs et les dégradés utilisés dans notre application pour remplir les formes et les arrière-plans .

Description du Paquet `javafx.scene.text`

Cette Interface représentant le contenu d'une entrée de texte.

```

<AnchorPane id="AnchorPane" prefHeight="200" prefWidth="387.0" styleClass="main"
  stylesheets="@main.css" xmlns="http://javafx.com/javafx/16" xmlns:fx="http://javafx.com/fxml/1"
  fx:controller="apprioritidapplication.FXMLDocumentController">
  <children>
    <Label fx:id="label" layoutX="10.0" layoutY="33.0" minHeight="16" minWidth="69" styleClass="title"
      text="L'analyse des paniers d'achat">
      <font>
        <Font name="System Bold" size="27.0" />
      </font>
      <textFill>
        <RadialGradient centerX="0.5" centerY="0.5" radius="0.5">
          <stops><Stop><color> <Color /></color> </Stop>
            <Stop offset="1.0"> <color> <Color red="1.0" green="1.0" blue="1.0" /> </color>
          </Stop> </stops> </RadialGradient> </textFill></Label>
    <Button fx:id="button2" layoutX="245.0" layoutY="120.0" onAction="#handleButtonAction2" prefHeight="25.0"
      prefWidth="82.0" styleClass="launcg_btn" text="Apriori" />
    <Button fx:id="button1" layoutX="54.0" layoutY="120.0" onAction="#handleButtonAction" prefHeight="25.0"
      prefWidth="82.0" styleClass="launcg_btn" text="Apriori-TID" />
  </children>
</AnchorPane>

```

Figure 20 Interface d'exécution

```

import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.net.URL;
import java.util.ArrayList;
import java.util.Arrays;
import java.util.Collections;
import java.util.Comparator;
import java.util.LinkedHashSet;
import java.util.List;
import java.util.Map;
import java.util.ResourceBundle;
import java.util.Scanner;
import java.util.Set;
import java.util.stream.Collectors;
import java.util.stream.Stream;
import javafx.event.ActionEvent;
import javafx.fxml.FXML;
import javafx.fxml.Initializable;
import javafx.scene.control.Alert;
import javafx.scene.control.Alert.AlertType;
import javafx.scene.control.TextArea;
import javafx.scene.control.TextField;
import javafx.stage.FileChooser;

```

Figure 21 Les bibliothèques de javafx et java utilise dans les classes java

Description du paquet JAVA.UUTIL

Il Contient les classes de collection, les fonctions de date et d'heure et diverses classes utilitaires. .Les fonctions principales dans le code

Les Fonction d'action des buttons

```

public void handleButtonAction2(ActionEvent event) throws IOException { //apriori ouvrir
    closeStage(); //fermer l'ancien onglet

    Parent parent = FXMLLoader.load(getClass().getResource("AprioriFXML.fxml"));

    Stage stage = new Stage(StageStyle.DECORATED);
    stage.setTitle("Library Assistant");
    stage.setScene(new Scene(parent));
    stage.setTitle("Apriori");
    stage.show();
}

```

Figure 22 Fonction pour ouvrir Apriori interface et fermé l'ancienne interface.

```

private void openFile(ActionEvent event) throws Exception {
    FileChooser fc = new FileChooser();
    fc.setTitle("Ficher");

    csvFile = fc.showOpenDialog(ApprioriTidApplication.scène.getWindow());
    csvPathTE.setText(csvFile.getAbsolutePath());
}

```

Figure 23 Fonction d'ouvrir le fichier csv.

```

private void handleRun(ActionEvent event) throws Exception {
    Apriori apriori2 = new Apriori(); String suppS = support.getText().trim();    confS = confidence.getText().trim();
    srS = solidRule.getText().trim();

    if (suppS.isEmpty()) { Alert alert = new Alert(AlertType.ERROR);    alert.setTitle(" Dialogue d'erreur");
        alert.setHeaderText("Regardez, Dialogue d'erreur");    alert.setContentText("Mmmm, Il y'a un erreur!");
        alert.showAndWait(); } else {    apriori2.CONFIDENCE=Double.valueOf(confS);
        apriori2.SUPPORT=Double.valueOf(suppS);    ArrayList<String> dataList = new ArrayList<>();
        String file="C:\\Users\\admin\\Desktop\\APPLICATION FINAL\\groceries (2).csv";    String delimiter = ","; String line;
    if (csvFile != null) {

        String file_location = csvFile.getAbsolutePath();
        file=file_location;
    }
    outputText.clear();
    outputText.appendText("Taille de fichier : " + csvFile.length() + "\n");

    try (BufferedReader br =
        new BufferedReader(new FileReader(file))) {
        while((line = br.readLine()) != null){
            List values = Arrays.asList(line.split(delimiter));
            String vl=String.valueOf(values);
            dataList.add(vl);
        }    } catch (Exception e){ }
    outputText.appendText("=====BASE DE DONNEES=====\n");    for(String string:dataList){
        outputText.appendText(string+"\n");    }    outputText.appendText("\n\n\n\n===== Les Motifs =====\n\n\n");

    Map<String, Integer> frequentSetMap = apriori2.apriori(dataList);
    Set<String> keySet = frequentSetMap.keySet();
    for (String key:keySet){    outputText.appendText(key+ " : "+frequentSetMap.get(key)+"\n"); }
    outputText.appendText("\n\n\n\n=====Les règles d'association =====\n\n\n");
    Map<String, Double> relationRulesMap = apriori2.getRelationRules(frequentSetMap);
    Set<String> rrKeySet = relationRulesMap.keySet();
    for (String rrKey : rrKeySet){    outputText.appendText(rrKey + " : " + relationRulesMap.get(rrKey)+"\n");    }
    outputText.appendText("\n\n\n\n=====Les règles solides =====\n\n\n");
    Map<String, Double> relationSolide = apriori2.getRelationRules(frequentSetMap);
    ArrayList<Double> valueList = new ArrayList<Double>(relationSolide.values());    double coutt=1;
    List<Double> rrKeySet1=valueList.stream().sorted(Comparator.reverseOrder()).collect(Collectors.toList());
    for (Double rrKey1 : rrKeySet1){outputText.appendText( getKey1(relationSolide, rrKey1)+" : " + rrKey1+"\n");
        relationSolide.remove(getKey1(relationSolide, rrKey1), rrKey1);
        coutt++;    if(coutt>Double.valueOf(srS))    break;    } }
}

```

Figure 24 Fonction d'exécution l'algorithme Apriori et afficher les résultats

```

private void handleRun(ActionEvent event) throws Exception {
    String suppS = support.getText().trim(); confS = confidence.getText().trim().isEmpty() ? "0" : confidence.getText().trim();
    srS = solidRule.getText().trim().isEmpty() ? "0" : solidRule.getText().trim(); if (suppS.isEmpty()) {
        Alert alert = new Alert(AlertType.ERROR); alert.setTitle(" Dialogue d'erreur"); alert.setHeaderText("Regardez, Dialogue d'erreur");
        alert.setContentText("Mmmm, Il y'a un erreur!"); alert.showAndWait(); } else {
String file_location = csvFile.getAbsolutePath(); transf.Lire(file_location); transf.saisieTableau(file_location);
    csvFile = transf.file2; System.out.println( transf.file2.getAbsolutePath()); System.out.println(csvFile.getAbsolutePath());
    float supp = Float.parseFloat(suppS); if (csvFile != null) { Scanner input = new Scanner(csvFile);
ArrayList<List<String>> transactions = new ArrayList<>(); ArrayList<List<String>> TIDList = new ArrayList<>(); List<String> C1;
if (csvFile.exists()) {while (input.hasNext()) { String newLine = input.nextLine(); String[] str = newLine.split(",");
List<String> elementsList = Arrays.asList(str); transactions.add(elementsList); outputText.clear();
    outputText.appendText("Taille de fichier : " + csvFile.length() + "\n"); for (List<String> l : transactions) {
for (String e : l) {System.out.print(e + " "); System.out.println(""); } C1 = transactions.get(0); System.out.println(C1);
for (int j = 0; j < C1.size(); j++) { TIDList.add(new ArrayList<>()); for (int i = 1; i < transactions.size(); i++) {
if (transactions.get(i).get(j).equals("1")) { TIDList.get(j).add(i + ""); } } }
    System.out.println("-----"); for (List<String> l : TIDList) {
for (String e : l) { System.out.print(e + " "); System.out.println(""); } System.out.println("-----");
float mySup = supp * (transactions.size() - 1); System.out.println("====support====>" + mySup);
int mySupInt = Math.round(mySup); System.out.println("====SupportInt====>" + mySupInt);
outputText.appendText("minimum support (" + supp + ") (" + mySupInt + " incetence )\n"); outputText.appendText("Les Motifs C(1) + Liste TID :\n");
int m = 0; for (String i : C1) { outputText.appendText(i + "\t\t\t\t" + TIDList.get(m) + "\n"); m += 1; }
outputText.appendText("\n"); FreqentItems(TIDList, C1, mySupInt); List<List<String>> L = new ArrayList<>();
    for (String i : C1) {
        if (i != null) {
            ArrayList<String> el = new ArrayList<>();
            el.add(i);
            L.add(el);
        }
    }
    System.out.println("L>" + L);

    //TIDList = [i for i in TIDList if i]
    ArrayList<List<String>> TIDList2 = new ArrayList<>();
    for (List<String> i : TIDList) {
        if (i != null) {
            TIDList2.add(i);
        }
    }
    //printing frequent itemset of L(1)
    outputText.appendText("La taille de Motifs de L(1):" + L.size() + "\n");
    outputText.appendText("\n");
    outputText.appendText("Les Motifs fréquents de Taille 1 \n");
    System.out.println(TIDList2);
    List<Integer> list = new ArrayList<>();
    int k = 0;
    for (List<String> q : L) {

```

Figure 25 Fonction d'exécution de l'algorithme AprioriTid et afficher les résultats

Et pour clarifier ce qui précède ; nous donnons les exemples suivants sur notre application depuis le prétraitement jusqu'à l'extraction des règles d'association:

Le système prend en entrée une base brute de transaction et la transforme en une base formel .

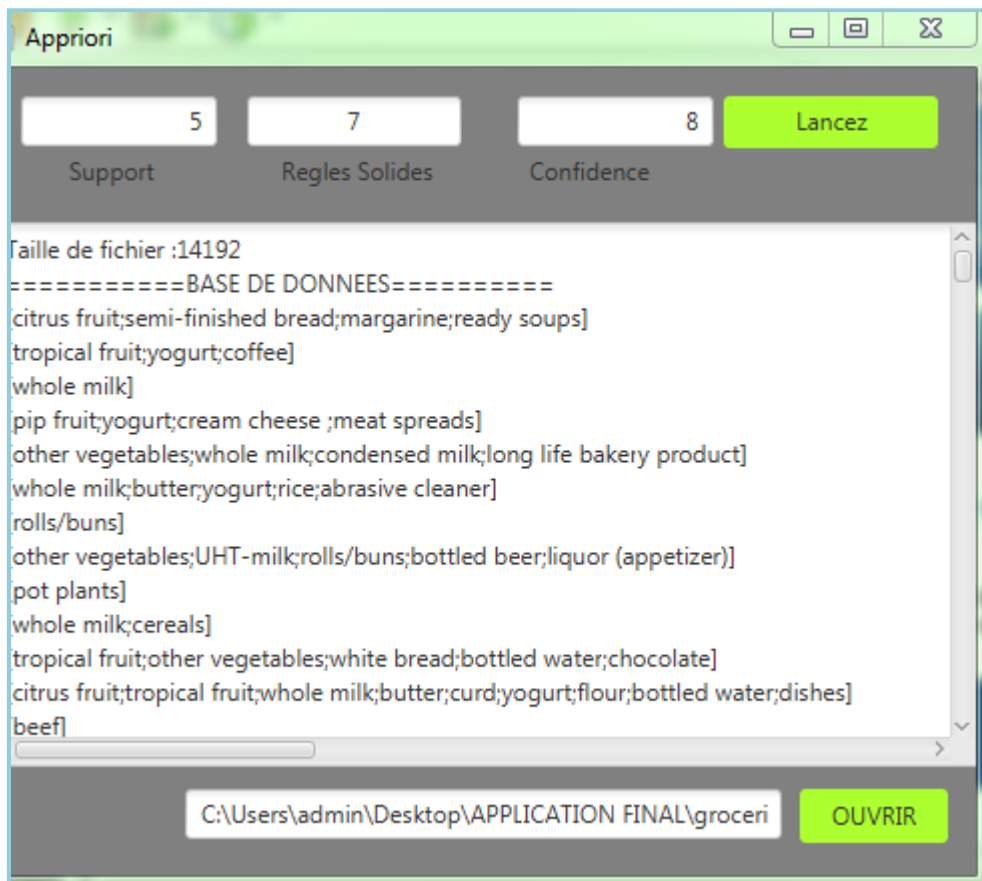


Figure 28 L'algorithmme trie la base de donnée

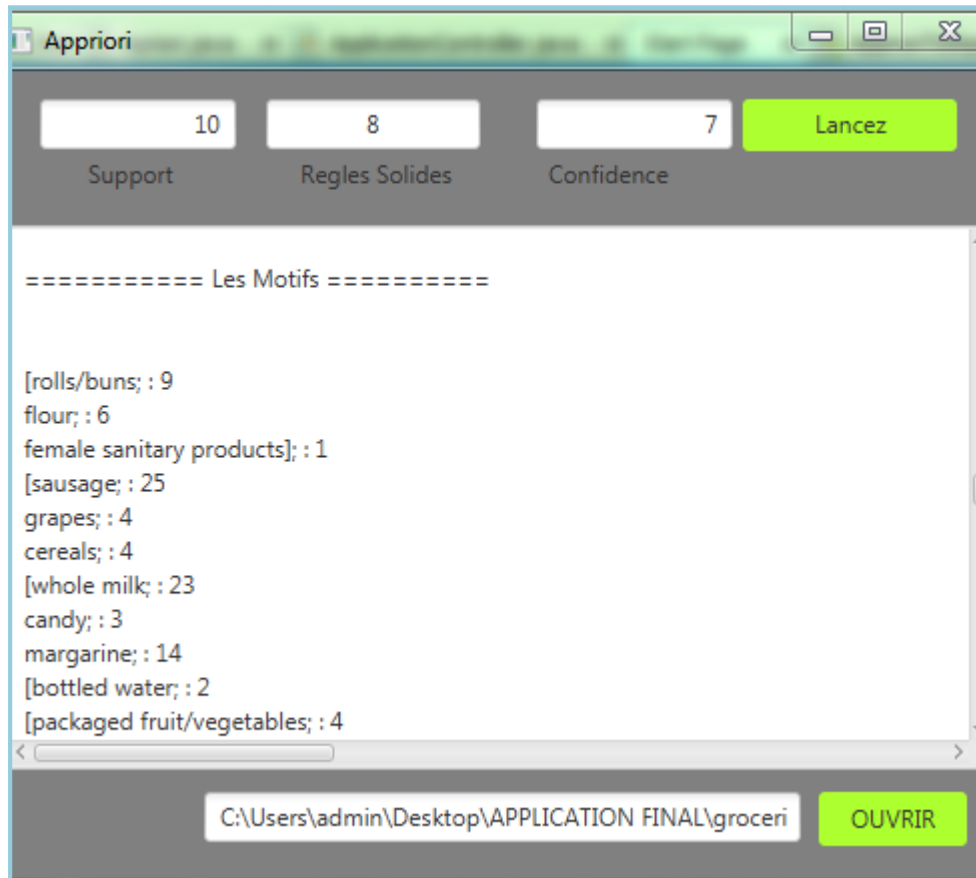


Figure 29 Calcul le support de chaque Itemset

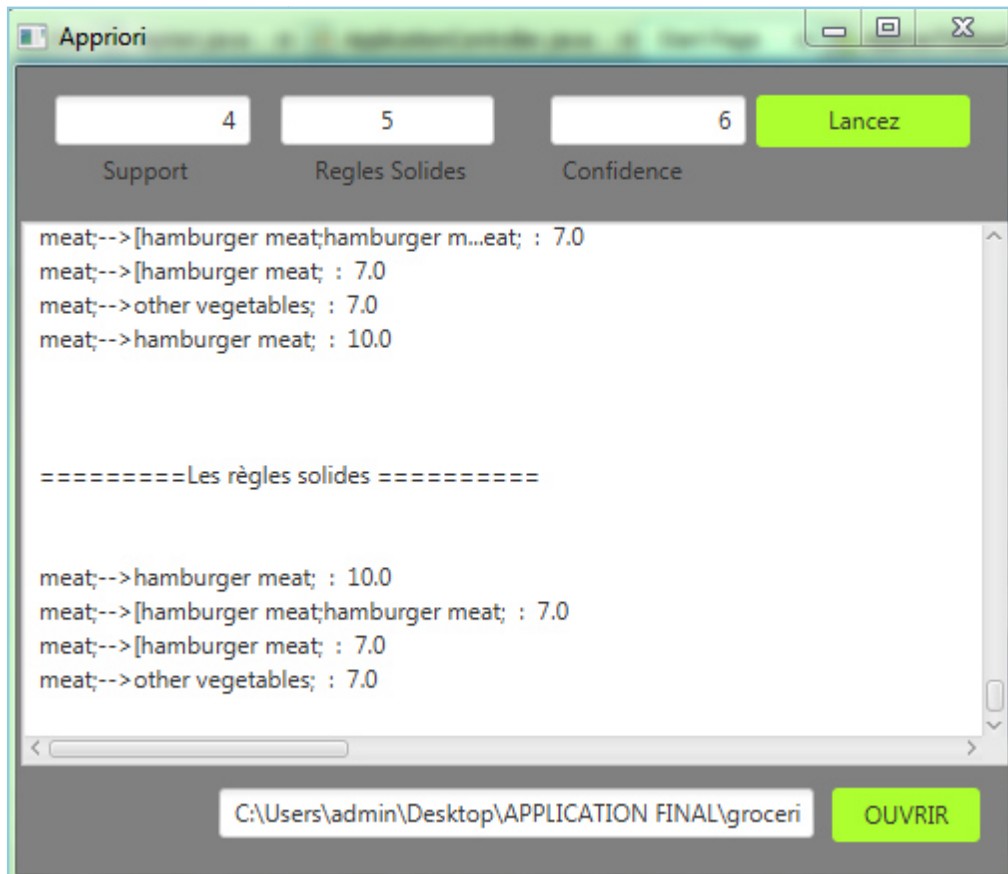


Figure 30 Affichage des règles solides

Et pour la sélection de l'algorithme Apriori Tid ; nous suivons les mêmes étapes que ci-dessus :

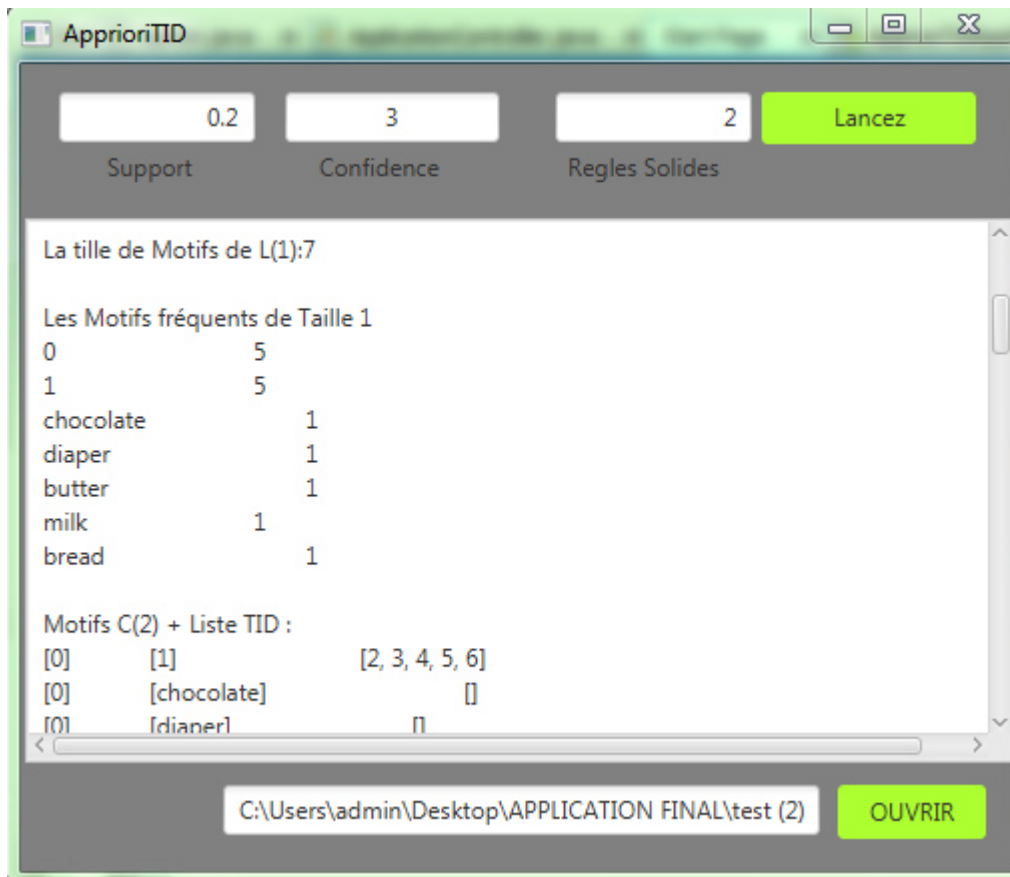


Figure 31 Génération d'un nouvel ensemble d'itemsets candidats(taille1+tille 2)

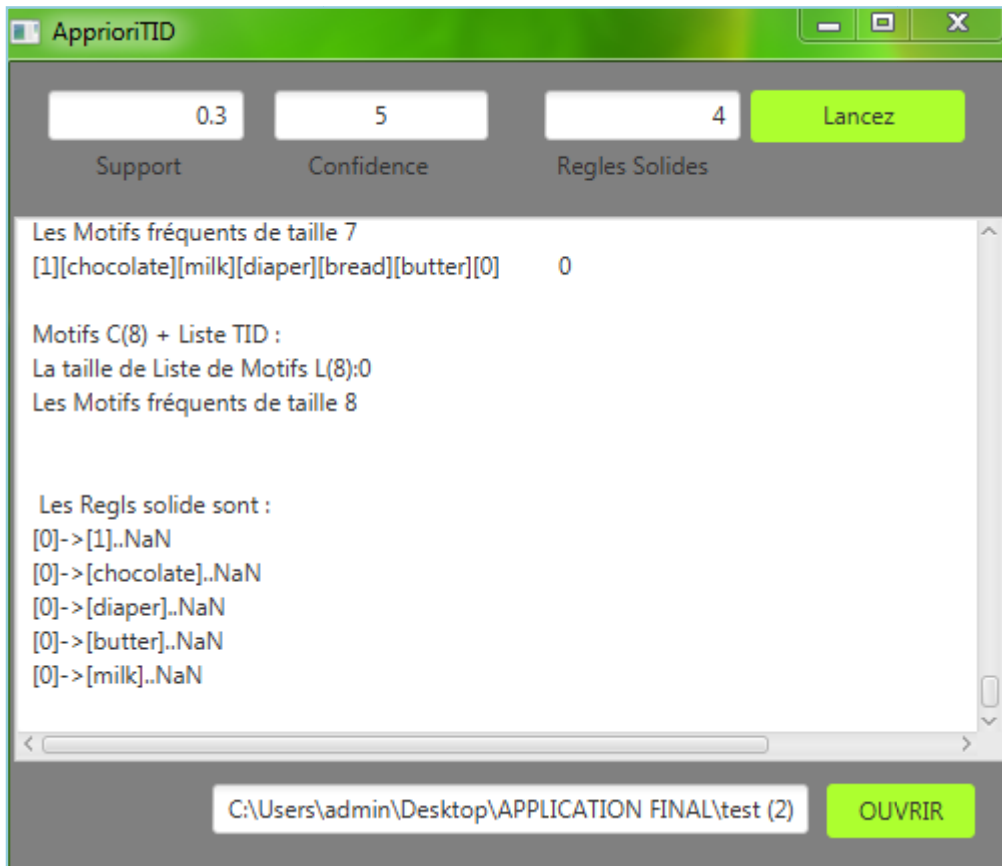


Figure 32 Extraction des règles solides

Conclusion

Nous avons détaillé dans ce chapitre les différents outils utilisés pour la mise en œuvre de notre proposition ainsi que les résultats obtenus et leur discussions ; et nous avons montrés que l'extraction des règles d'association d'une basa de donnée suivant l'algorithme Apriori et AprioriTid permet de nous aidez à découvrir des liens entre les transactions.

Conclusion General

Conclusion General

L'extraction des règles d'association est une méthode descriptive de fouille de donnée qui a reçu beaucoup d'intérêt de la part des chercheurs. On peut la définir comme étant la recherche de relation entre des Items dans un ensemble de données.

Cette technique est très utilisée pour l'analyse des paniers. Il s'agit d'analyser l'ensemble des achats effectués par les clients d'une entreprise commerciale. Chaque achat contient un ensemble d'Items (article), il correspond à un panier. Pour cela, l'entreprise doit archiver les achats (transaction) effectués dans une base de donnée binaire (transaction, Items) ou l'attribut transaction est une clé et Items est l'ensemble de n article article1, article2, ..., Article n.

Plusieurs algorithmes d'extraction de règles d'association à partir de bases de données ont été proposés (Agrawal et al, 1993), (Agrawal et Strikant, 1994)

La génération des règles est donc une opération de transformation des ensembles d'Items en règle de manière efficace sous la forme Si X alors Y.

Ces règles sont intuitivement faciles à interpréter car elles montrent comment des produits ou des services se situent les uns par rapport aux autres. Elles sont particulièrement utiles en marketing et peuvent être facilement utilisées dans le système d'information de l'entreprise.

L'extraction de règles d'association est décomposée en deux sous-problèmes :

- La recherche des ensembles fréquents d'items (les itemsets fréquents) : Cette étape est coûteuse en termes de temps d'exécution car le nombre d'itemsets fréquents dépend exponentiellement du nombre d'items manipulés

- La génération des règles d'association à partir de ces itemsets fréquents : Cette étape est plus simple et beaucoup moins coûteuse que la génération des itemsets fréquents car il n'est plus nécessaire de faire des parcours coûteux de la base de transactions.

Plusieurs travaux ont porté sur la recherche des itemsets fréquents dont l'objectif est d'une part la réduction du temps d'extraction des règles d'association, et d'autre part la pertinence et l'utilité des règles extraites.

Dans notre travail, nous mettons en évidence deux algorithmes :

Comme c'est déjà vu, l'algorithme Apriori est le pionnier de recherche de motifs fréquents. Cependant, il présente deux inconvénients majeurs: le nombre élevé de candidats générés durant chaque itération et le problème du parcours coûteux de la base de données pour chaque itération, ce qui augmente la complexité.

Apriori a été proposé pour optimisation. Son idée principale consiste à réduire le nombre de passages dans la base de données et le nombre de candidats ainsi générés.

Grâce aux recherches qui ont été menées au cours de cette étude, nous avons conclu que le seuil a un grand effet dans l'extraction des règles, en effet si le seuil choisi par l'utilisateur est trop petit, la génération des itemsets fréquents sera trop lourde à cause de la base de données volumineuse puisque il parcourt la base de données plusieurs fois afin de générer les itemsets fréquents et accepter les Itemset dont le support est supérieur au seuil min comme nous l'avons dit plus tôt.

Aussi nous avons constaté que l'application des deux algorithmes Apriori et Apriori Tid sur une base identique (Grocerie) avec le même seuil et support donne les mêmes résultats (que ce soit de temps d'exécution ou à travers les règles d'association).

Du côté positif, on peut dire qu'on peut prédire les articles les moins vendus ou les plus demandés par les clients, Aussi la génération des règles d'association nous permet de trouver des associations entre les différents éléments qui ont été achetés.