



République Algérienne Démocratique et Populaire
Ministère De L'enseignement Supérieur Et De La recherche
Scientifique
Université de Mohamed Khider - BISKRA
Faculté des sciences exactes, des sciences naturelles et de la vie
D´département d'informatique

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Option : **Système d'information optimisation et
décision(SIOD)**

Extraction de connaissances pour l'analyse des paniers d'achats

Par :

Boukebbal Rayen

Soutenu le 23/09/2020, devant le jury :

Xxxxxx	xxx	Président
HOADJLI	Hadia	Rapporteur
Xxxxxx	xxx	Examineur

Année universitaire :2019/2020

Dédicaces

Nous dédions ce modeste travail :

À ma mère la flamme de ma vie, mon exemple éternel, mon soutien moral et la source de ma joie

Ainsi que de mon bonheur ; celle qui s'est toujours sacrifiée pour me voir réussir.

À mon père celui à qui je dois tous le respect et l'amour, pour son soutien, sa tendresse, sa patience

Et surtout son aide dans les moments les plus difficiles de ma vie.

À toute notre familles À tous nos amis chacun à son nom

Et à tous nos enseignants

Remerciement

Je voudrais d'abord remercier Dieu Tout-Puissant de m'avoir
donné la force

Volonté et patience d'accomplir cet humble travail.

Après cela, je tiens à remercier chaleureusement Madame
Hoadjli Hadia, directrice de recherche, pour ses efforts, ses
conseils et sa disponibilité. Je remercie également sincèrement

les membres du jury qui ont accepté

de juger mon travail de fin d'études et de l'enrichir de leurs
propositions. Nos remerciements et gratitude aux enseignants

du Département d'informatique de l'université de Biskra.

Nous remercions également tous ceux qui ont contribué à notre
éducation et à notre formation.

Table des matières

IntroductionGénéral	2
Chapitre 1 : Entreprise et panier d'achat	
1.Introduction	5
2. L'entreprise	5
2.1 Définition	5
2.2Les partenaires de l'entreprise	6
2.3 Les problématique de gestion de l'entreprise	6
3. Les données volumineuses	7
3.1Définition de base de données volumineuse	7
3.2 Les problématique de données volumineuse	8
4. L'historique de l'informatisation de gestion d'entreprise	8
5.Système d'une entreprise	10
5.1 Composants d'un système d'entreprise	10
5.2 Système d'information	10
5.3 Systèmes experts	11
6. Solution logiciel pour la gestion de l'entreprise	12
6.1 LOGICIEL ERP	12
6.1.1définition	12
6.1.2 Caractéristiques de l'ERP	12
6.2 GRH	13
6.3 SCM	13
6.3.1 domaines d'intervention du SCM	13
6.3.2Les logiciels de Supply Chain Planning (SCP) et Advanced Planning System (APS) ...	14
6.4CRM	14
6.4.1 Définition	14
6.4.2 Le principe	14
6.4.3Importance d'un CRM pour une entreprise	15
6.4.4Les avantage du CRM	15
6.4.5 La CRM pour mettre le client au premier	16

6.5la gestion de CRM	16
6.5.1 La gestion de production	16
6.5.2 Lagement commerciale, La logistique	17
6.5.3 La comptabilité	17
6.5.3.1 Pour l'entreprise	18
6.5.3.2Pour les tiers	18
6.5.3.3 Pour la nation	18
6.5.3.4Pour l'administration fiscale	18
7. les paniers d'achat	18
7.1 Introduction.....	18
7.2 Définition	19
7.3 Comment l'entreprise comprend les besoins du client via le panier d'achat	19
8. Conclusion	20

Chapitre 2 : Data mining

1.Introduction	22
2. Processus d'extraction de connaissances(ECD)	22
2.1 Le principe.....	22
2.2 Les étapes principales d'ECD	23
3.Fouille de données (Data mining)	24
3.1. Introduction	24
3.2. Définition	25
3.3 .Le Data Mining, un avantage compétitif pour toutes les entreprises.....	25
3.4.Domains d'application.....	26
3.4.1.La banque	26
3.4.2.L'assurance.....	26
3.4.3 .Le commerce.....	26
3.4.4 .La détection de fraude	27
3.4.5.La médecine et la pharmacie.....	27
3.5. Processus de fouille de données.....	27
3.6. Principales taches de fouille de données	27
3.6.1 .Classification.....	29
3.6.1.1 Classification supervisée (apprentissage).....	29
3.6.1.2 Classification non supervisée.....	30
3 .6.2. L'estimation	30

3.6.3. La prédiction.....	30
3.6.4. Les règles d'association	30
3.6.5. La segmentation.....	31
3.7 Apprentissage supervisé Vs Apprentissage non supervisé	31
3.7.1. Les techniques supervisées	31
3.7.2. Les techniques non supervisées	31
3.8. Les méthodes de Data Mining	31
3.8.1. Les méthodes classiques	32
3.8.2. Les méthodes sophistiquées	32
3.8.2.1. Arbre de décision	32
3.8.2.2. Les réseaux de neurones	33
3.8.4. Règles d'association.....	34
3.9 Etapes de création d'un projet data mining	34
4.Conclusion	35

Chapitre3:les règles d'associations pour l'analyse de panier d'achat

1.Introduction	37
2.Les règles d'associations.....	37
3.Concepts et définition de support et de confiance	38
3.1 support et de confiance.....	Error! Bookmark not defined.
4.Algorithmes	39
4.1. Les itemsets.....	40
4.1.1 Des itemsets fréquents maximaux	42
4.1.2 Des itemsets fréquents fermés	42
4.2 L'algorithme	44
4.2.1 Apriori	44
4.2.1.1 Les propriétés d'Apriori.....	44
4.2.1.2 Principe d'algorithme apriori	45
4.2.1.3 Inconvénient de l'algorithme Apriori	46
4.2.2 Améliorations proposées à Apriori	46
4.2.3 Principe de l'Algorithme	49
8. Conclusion	52

chapitre 4:Conception

1.Introduction	54
2.Description de notre travaille	54

3.Conception globale du système	55
4.Conception détaillée	57
4.1L'étape 1	57
4.1.1Prétraitement	57
4.1.2Les données transactionnelles	57
4.1.3 Les données formelles	58
4.1.4 Fonctionnement	59
4.2L'étape 2.....	59
4.2.1 Génération des itemsets	59
4.2.2Découverte d'itemsets fréquents	60
4.2.3Sauvegarde	60
4.2.4 Combinaison	60
4.3L'étape 3.....	61
4.3.1 Extraction des règles	61
4.3.2 Interprétation	63
5.Conclusion.....	63

Chapitre5: Implémentation

1.Introduction	67
2.Outils et langages de développement.....	67
2.1Environnement et outils de programmation	67
2.1.1 Ntbeans	67
2.1.2 Scence builder.....	68
2.2Langages de programmation	68
2.2.1 Java	68
2.2.2 Javafx.....	69
2.2.3 Différents entre Java suing et Java fx et pour quoi le Java fx.....	70
3.Description des différentes classes	71
4.Base de données utilisé.....	72
5.Présentation de l'application	73
5.1. Première interface	73
5.2 Deuxième interface	74
6.Conclusion	79
Conclusion générale:.....	81
BIBLIOGRAPHi.....	83

Tables des Figures

Figure1 : Schéma de fonctionnement d'un système expert.....	11
Figure 2 : processus de l'extraction de connaissance.....	24
Figure 3 : Les taches de la fouille de donnée.....	28
Figure 4 : Les Technique.....	29
Figure 5 : Nœud.....	33
Figure 6 : Treillis d'itemsets.....	41
Figure 7 : Treillis composé des itemsets fréquents.....	42
Figure 8 : Treillis des itemsets.....	43
Figure 9 : Diagramme de séquence de notre système.....	56
Figure 10 : Schéma général du système 57.....	
Figure 11: le prétraitement.....	60
Figure 12 : l'étape 3(Extraction des règles).....	63
Figure 13 : Le système proposé pour analyse les paniers d'achat.....	64
Figure14. Interface Netbeans.....	67
Figure 15 : interface de Scene Builder.....	68
Figure 16 : interface de Java.....	69
Figure 17 : interface de JavaFx.....	69
Figure 18 : partie de base de données groceries.....	72
Figure19 : La forme principale.....	73
Figure 20 : interface de traitement.....	74
Figure21 : base de données transactionnel.....	75
Figure 22 : base de données formel.....	74
Figure 23 : génération 1 et leur itemsets fréquents.....	76
Figure 24 : génération 2 et les itemsets fréquents.....	77
Figure 25 : génération 3 et les itemsets.....	78
Figure 26 : les règles d'associations extrait.....	79

Résumé

L'extraction de connaissances à partir de données (ECD), plus précisément le data mining, fait référence au processus non simple d'extraction d'informations utiles et de connaissances enfouies dans des bases de données, le data mining est l'une des solutions auxquelles de nombreuses entreprises recourent pour filtrer leurs données et en extraire les connaissances et les informations utiles.

L'une des techniques utilisées dans le data mining est les règles d'associations qui nous fournissent et nous expliquent les liens et les relations qui sont importants dans les grandes bases de données dans notre système nous avons adopté l'algorithme Apriori TID qui nous permettra la divulgation des listes d'articles les plus vendus et des règles d'association entre eux, à travers lesquelles les entreprises et les institutions peuvent bénéficier de leurs énormes données et comprendre les habitudes de comportement de leurs clients et ainsi étendre leur influence sur le marché et augmenter leurs profits.

Mots Clés : CRM, ECD, data mining, règles d'association.

ملخص

استخراج المعرفة من البيانات (ECD) ، وبشكل أكثر تحديداً استخراج البيانات ، يشير إلى العملية غير البسيطة لاستخراج المعلومات والمعرفة المفيدة المدفونة في قواعد البيانات ، واستخراج البيانات هو واحد من الحلول التي تستخدمها العديد من الشركات لتصفية بياناتها واستخراج المعارف والمعلومات المفيدة.

إحدى التقنيات المستخدمة في التنقيب عن البيانات هي قواعد الارتباط التي تزودنا وتشرح لنا الروابط والعلاقات المهمة في قواعد البيانات الكبيرة في نظامنا ، وقد اعتمدنا خوارزمية Apriori TID التي ستسمح لنا الكشف عن قوائم الأصناف الأكثر مبيعاً وقواعد الارتباط بينها ، والتي من خلالها يمكن للشركات والمؤسسات الاستفادة من بياناتها الضخمة وفهم العادات السلوكية لعملائها وبالتالي توسيع نفوذهم في السوق وزيادة أرباحهم.

الكلمات المفتاحية: CRM, استخراج المعرفة من قواعد البيانات (ECD)، التنقيب عن البيانات، قواعد الارتباط.

ABSTRACT

Knowledge Extraction from Data (ECD), more specifically data mining, refers to the non-simple process of extracting useful information and knowledge buried in databases, data mining is one of the solutions that many companies use to filter their data and extract useful knowledge and information.

One of the techniques used in the data mining is the association rules that provide us and explain to us the links and relationships that are important in large databases in our system we have adopted the Apriori TID algorithm which will allow us the disclosure of the lists of best-selling items and the rules of association between them, through which companies and institutions can benefit from their huge data and understand the behavioral habits of their customers and thus extend their influence in the market and increase their profits.

Key words: CRM, Knowledge Extraction from Databases (ECD), Data Mining, Correlation Rules.



Introduction

Générale

Introduction générale

Avec l'expansion du monde et le développement de la technologie, tout est exposé à l'inflation des entreprises et des institutions en raison des besoins humains croissants, ce qui a conduit à une augmentation des services pour répondre aux besoins des citoyens. L'entreprise en général est une unité économique indépendante avec des ressources humaines et matérielles qu'elles collectent pour produire des biens et services préparés pour la vente [63], son objectif premier est de satisfaire le client en répondant à ses différents besoins. Tout cela a entraîné une augmentation du nombre de grandes bases de données qui contiennent souvent beaucoup d'informations et il est donc difficile d'interpréter et d'extraire des informations utiles.

Malgré toutes les difficultés que les institutions rencontrent en raison de la mauvaise gestion, de l'inflation des données et de la difficulté à comprendre le comportement des consommateurs, elles s'efforcent de développer leurs compétences de base représentées dans la facilité de gestion, de compréhension et de traitement des données, ce qui représente un avantage concurrentiel et un point fort pour la simplification de la position sur le marché et la continuité de son travail, ils ont donc recours à des techniques et des méthodes pour résoudre leurs problèmes.

Et pour atteindre son objectif, parmi ces techniques, il y a ce que l'on appelle l'extraction de connaissances à partir des données, qui a été créée au milieu des années 1990 pour tenter de faciliter la tâche aux chercheurs et aux industriels et leur apporter des solutions, elle permet par le biais de ses outils à trier les grandes bases de données et à en extraire des informations utiles, comme analyser les achats des clients afin de comprendre leurs habitudes et leurs comportements. L'une des techniques de data mining les plus utilisées sont les règles d'association, qui permettent de découvrir les relations et les liens cachés qui existent dans d'énormes bases de données. Si bien que les règles d'association sont extraites par plusieurs algorithmes, parmi les plus connus l'algorithme Apriori, mais cela dépend du passage répétitif de la base de données pour en extraire les règles d'association, ce qui laisse plusieurs inconvénients, dont l'exploitation de beaucoup de temps et d'espace.

Nous avons donc eu recours à l'algorithme Apriori TID, qui est considéré comme une amélioration de l'algorithme Apriori, puisqu'il s'appuie sur un passage unique sur la base de

Introduction générale

données pour extraire les règles de lien Ce mémoire est composé de quatre chapitres en plus de cette introduction et une conclusion générale, il est organisé de la manière suivante :

Dans le premier chapitre, nous avons présenté les entreprises, les difficultés auxquelles elles sont confrontés et de leur recours à la programmation pour trouver des solutions à leurs problèmes et augmenter leurs profits. Quant au deuxième chapitre, nous présentons l'extraction de connaissances et de données en tant que solution pour résoudre la problématique de données au sein de l'entreprise, ou nous avons cité les caractéristiques de ses différents outils, puis nous avons consacré le troisième chapitre pour les règles d'associations qui représentent la solution qu'on a choisi, avec les algorithmes d'extraction de règles d'associations parmi lesquelles L'algorithme Apriori TID que nous avons utilisé dans notre système. Dans le quatrième chapitre, nous avons expliqué le fonctionnement de notre système ainsi que l'architecture que nous avons proposé en détaillant chaque composant. Enfin, le cinquième chapitre est consacré à la mise en œuvre du logiciel qui matérialise notre étude.



Chapitre 1 :

Entreprise et panier d'achat

1.Introduction

Depuis le début de l'implantation d'institutions et d'entreprises dans leurs différentes spécialités, leur objectif principale était d'obtenir le plus grand nombre de profits possible en attirant l'attention du client en essayant de comprendre ses besoins.

Avec l'élargissement de son périmètre et l'augmentation de sa taille, les envies de l'entreprise se sont accrues et ses intérêts diversifiés, ce qui a conduit à une augmentation du nombre de ses clients, mais cet aspect avait des points négatifs, tels que :

- La surpopulation des données.
- Difficulté à contrôler si entre clients ou dans les entrées et sorties.
- Difficulté à négocier entre le travailleur et le client.
- La façon dont les choses vont.

Mais l'émergence de la technologie et ses progrès, a considérablement réduit les pressions et les problèmes évoqués précédemment, en particulier l'informatique avec ses programmes et applications.

Dans cette partie nous allons définir l'entreprise, quels sont les problèmes les plus importants auxquels elle est confrontée et comment les résoudre.

2. L'entreprise

2.1 Définition

Dans l'histoire économique on trouve de nombreuses définitions de l'entreprise, notamment :

Une entreprise est une organisation humaine dont le but est de produire des biens ou des services en vue de les délivrer à des clients, cela dans un but d'équilibre ou d'excédent financier.

« L'entreprise est une unité économique autonome disposant de moyens humains et matériels qu'elle combine en vue de produire des biens et services destinés à la vente.[1]

L'entreprise est une unité économique et juridique qui a pour principale fonction la production de biens et services destinés à être vendus sur un marché.

L'activité d'une entreprise peut être décomposée en deux phases distinctes :

- l'activité productive, c'est à dire la création de biens ou services.
- l'activité de répartition des richesses en contrepartie des biens ou services .[2]

2.2 Les partenaires de l'entreprise :

Pour que l'entreprise réussisse, de nombreuses parties interfèrent avec l'objectif d'atteindre une efficacité élevée

- Clients : le marché de l'entreprise dépend des ventes qu'elle réalise donc des achats de ses clients
P Drucker qualifie le client de « pilier de l'affaire dans la mesure où il est le seul à donner du travail à l'entreprise ». Ils considèrent que le seul but de l'entreprise est de satisfaire les besoins du client et à travers lui les besoins de la société toute entière.
- Les autres entreprises : peuvent être fournisseurs des matières premières, installations, services dont leurs prix conditionnent le prix de vente des produits de l'entreprise, ou bien des entreprises concurrents que l'entreprise prend en compte(prix appliqués par les concurrents, caractéristiques de leurs produits).
- Les salariés : la charge salariale représente un élément important du prix de revient du produit de l'entreprise mais aussi la qualité du travail et la productivité de ses salariés.
- L'entreprise doit concilier exigences des salariés et coût qui permet d'être compétitive.
- Banques et institutions de crédit : l'entreprise obtiennent des crédits auprès des banques et verse des intérêts qui représentent un élément de coût.
- Administrations : Etat et les collectivités locales qui lui réclament des impôts et des taxes ainsi que les organismes sociaux à qui elle doit verser des cotisations sociales. Mais elle s'adresse à ces administrations pour obtenir des aides ou utiliser des biens et services collectifs que ces administrations créent et entretiennent [41].

Autres acteurs : investisseurs, public, associations (protection de l'environnement, associations culturelles ...)

2.3 Les problématique de gestion de l'entreprise :

L'entreprise reçoit plusieurs problèmes au niveau de sa gestion. Dans cette partie de notre travaille, nous allons décrire quelque problème

- Problème de communication au sein de l'équipe.

- Perception erronée du cahier des charges.
- Manque d'adaptation au changement.
- Gestion du temps et de la capacité.
- Manque de visibilité sur le reste des mandats[3] .
- Défaut de développement de la compétence et des comportements clés.
- Et surtout la gestion des taux volumineux de données

3. Les données volumineuses :

Comme nous l'avons appris précédemment, toutes les entreprises passent par des étapes et font face à plusieurs difficultés, et elles diffèrent selon l'entreprise, de sorte que chacune a des intérêts particuliers, mais elles partagent un objectif, qui est répondre aux demandes de ses clients en collectant le plus grand nombre possible d'informations qui les aident à répondre à leurs besoins, mais cet aspect peut conduire à l'émergence d'un aspect négatif représenté par ce qu'on appelle des données de ballonnement ou bien les données volumineuse. Que nous allons définir et identifier un certain nombre de ses inconvénients ou sa problématique.



3.1 Définition de base de données volumineuse :

La grande et énorme quantité de données qui sont collectées et non abandonnées même si nous n'en avons pas besoin et même malgré leur existence pendant une longue période de temps, cela conduit à ce qu'on appelle les données volumineuse

Les données sont considérées comme des faits et des chiffres bruts.

La base de données est une collection de données.

Les Base de données volumineuse sont généralement considéré comme une très grande quantité de données pour le stockage et le traitement ou lorsque les données elles-mêmes sont.

Les données en grand volume et différentes variétés peuvent être considérées comme des Big Data.

On les décrit souvent en recourant aux

Trois V : Volume, Variété et Vélocité. Douglas Laney fut l'un des premiers auteurs à employer le concept trois V en 2001 pour décrire ce type de données.

Le volume dans cette description caractérise le fait que ces données sont une avalanche d'informations qui nous tombe dessus (de l'ordre du zettaoctet : 10²¹ octets).

Puisque les données analysées ne sont plus forcément structurées comme dans les analyses statistiques traditionnelles, mais peuvent être

Du texte, des images, du contenu multimédia, des traces numériques, des objets connectés, etc., on parle de données variées[4] .

3.2 Les problématique de données volumineuse :

Le stockage massif des données noie l'information pertinente et engendre des problèmes théoriques liés à la volumétrie des données disponibles.

Il y'a plusieurs problèmes concernons la base de données volumineuse telle que :

- Les énormes coûts de stockage de la grande quantité de données
- La difficulté de traiter toutes ces données
- Problèmes de cohérence
- Manque de qualité
- Répétition fréquente
- la vitesse de réponse
- Trouver le signal dans le bruit

Parmi les difficultés rencontrées par les entreprises à ses débuts figurait la manière de stocker et de gérer ses informations

4. L'historique de l'informatisation de gestion d'entreprise :

Malgré le développement que connaissent actuellement les entreprises, elles se heurtaient par le passé à des obstacles de gestion.

Dans le paragraphe suivant, nous verrons le rôle de l'informatique dans
L'amélioration de ces difficultés :

- les années 60-70 et le mainframe (informatique d'entreprise) : il ne s'agit à cette époque que de gérer des volumes gigantesques de données ; les applications de gestion

automatisent les processus opérationnels et répétitifs ; le marché des systèmes informatiques est dominé par quelques grands constructeurs ; les gros systèmes sont propriétaires et centralisent la puissance de traitement.

- les années 80 et les mini systèmes (informatique de département) : le marché de l'informatique s'ouvre aux PME ; l'ère des minis permet également à un grand nombre d'utilisateurs d'accéder aux données aux travers d'applications couvrant des besoins plus spécifiques ; l'utilisateur dispose d'une interface texte pour l'interaction avec le système (terminaux passifs).
- les années 90 et le client serveur (informatique individuelle): c'est le plein essor des ordinateurs personnels, de la bureautique, et l'explosion des réseaux et des télécommunications; le partage de l'information devient le défi majeur ; les applications bureautiques disponibles (traitement de texte et tableur) répondent à des besoins de traitement de l'information jusqu'à présent non pris en compte ; les terminaux passifs sont peu à peu remplacés par des micro-ordinateurs capables de les émuler; au-delà du réseau local, les applications client-serveur donnent l'occasion de répartir les traitements entre les machines les plus adaptées : les applications bureautiques sur les postes de travail et les applications critiques sur les serveurs ; la connexion entre le réseau local et le site central s'effectue par des liaisons téléinformatiques proposées par l'opérateur institutionnel ; le micro-ordinateur devient nomade grâce aux portables, capable ainsi de se connecter au réseau local ou à distance au système informatique de l'entreprise.
- les années 2000 et le boum de l'Internet : aujourd'hui le poste de travail, nomade ou fixe, est connecté au réseau local de l'organisation mais il est aussi ouvert sur l'extérieur grâce à l'Internet ; la communication entre les ordinateurs s'effectue grâce à un ensemble de protocoles normalisés (TCP/IP) ; les technologies de l'Internet deviennent des normes pour la mise en place tant des systèmes informatiques d'entreprise (intranet) que des systèmes informatiques interconnectés avec les partenaires (extranet) ; les applications métier obéissent au standard du web (HTTP, HTML) ; le poste de travail, équipé seulement d'un navigateur, peut accéder par le réseau à l'ensemble des applications « client »; le système d'information de l'entreprise est accessible depuis un poste de travail banalisé mais également depuis de nouveaux équipements comme le téléphone mobile.[5]

5. Système d'une entreprise :

La notion du système en générale représente un ensemble d'éléments en interaction dynamique permanente organisé en fonction d'un but à atteindre.

Un système est ensemble ordonné d'éléments liés les uns aux autres et qui interagissent entre eux. Le concept est employé aussi bien pour désigner tout un ensemble de concepts que des objets réels pourvus d'organisation [42].

Le périmètre du terme Système d'une entreprise peut être très différent d'une organisation à une autre et peut recouvrir selon les cas le tout ou une partie des éléments suivants :

5.1 Composants d'un système d'entreprise

Une organisation est composée d'un ensemble de systèmes. Il y a le système opérant, de pilotage et d'information, ce sont trois sous-systèmes qui interagissent entre eux :

- Le système opérant : c'est ce qui est à la base de toute organisation, c'est ce système qui permet la transformation de l'information dont l'objectif est de la restituer à la
- Bonne personne. Il correspond aux différents services d'une entreprise.
- Le système de pilotage : C'est ce qui va contrôler et piloter le système opérant. Il se situe donc à la tête du système d'information fixant les objectifs et prenant les décisions.
- Le système d'information : C'est ce qui intervient entre les deux autres systèmes. Ce système s'occupe de collecter, stocker, transformer et diffuser des données et informations dans le système opérant et de pilotage.

En résumé, un système d'information permet au système opérant de communiquer des informations qui ont été collectées et modifiées au système de pilotage qui est en charge de contrôler et prendre des décisions. [6]

5.2 Système d'information :

Le système d'information (SI) est un élément central d'une entreprise ou d'une organisation. Il permet aux différents acteurs de véhiculer des informations et de communiquer grâce à un ensemble de ressources matérielles, humaines et logicielles. Un SI permet de créer, collecter, stocker, traiter, modifier des informations sous divers formats.

L'objectif d'un SI est de restituer une information à la bonne personne et au bon moment sous le format approprié.

5.3 Systèmes experts :

Les systèmes experts (ou moteurs de règles) sont une technique de l'Intelligence Artificielle, qui permettent de reproduire les mécanismes cognitifs d'une personne dans un domaine bien particulier.

Ces logiciels permettent un gain de temps et/ou constituent une aide à la décision fondée sur un moteur d'inférence (i.e. partir de faits pour aller vers une conclusion) et sur des connaissances précises (celles que l'humain, expert, aura préalablement fait ingérer à la machine). Ils peuvent répondre à des questions en raisonnant à partir de faits et de règles connues. Aussi ont-ils souvent été employés dans l'industrie.

Un système expert se compose de 3 parties :

- Une base de faits.
- Une base de règles.
- Un moteur d'inférence.

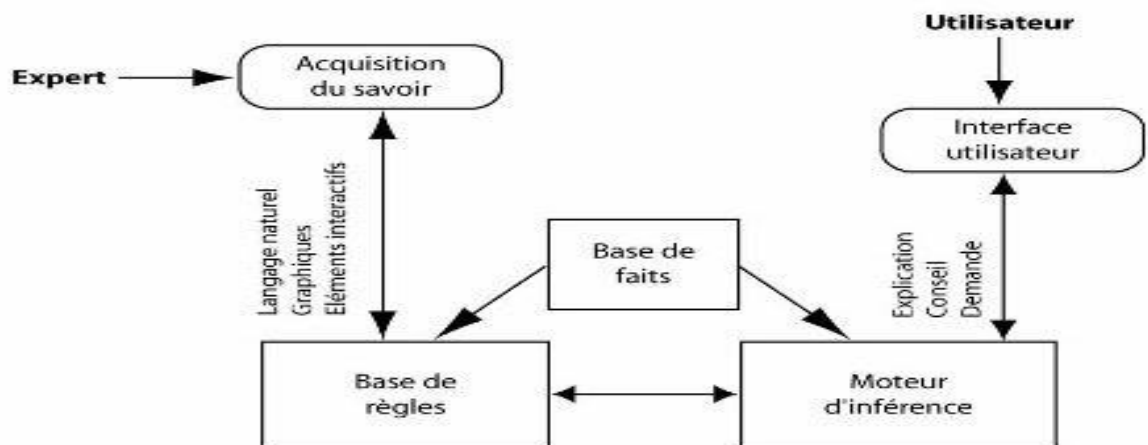


Figure 1 : Schéma de fonctionnement d'un système expert [7]

6. Solution logiciel pour la gestion de l'entreprise :

l'informatique avec l'avènement des technologies de l'information et de la communication a beaucoup aidé à faciliter les affaires des entreprises difficiles et a grandement contribué à la conduite de ses affaires, rendant les affaires difficiles et presque impossibles simples et quotidiennes pour ces entreprises, en plus d'augmenter les revenus et de gagner la confiance du client, parmi ces solutions, nous présentons ce qui suit :

6.1 LOGICIEL ERP :

6.1.1 définition :

Un ERP (Enterprise Resource Planning) ou encore parfois appelé PGI (Progiciel de Gestion Intégré) est un système d'information qui permet de gérer et suivre au quotidien, l'ensemble des informations et des services opérationnels d'une entreprise.

La vocation d'un logiciel ERP est d'homogénéiser le SI de l'entreprise avec un outil unique, capable de couvrir un large périmètre de gestion, c'est-à-dire :

- La gestion des achats
- Les ventes et la gestion commerciale
- La gestion comptable : comptabilité client, fournisseur, immobilisations, analytique,...
- Le contrôle de gestion : budgets,...
- La gestion des stocks : logistique, transport,...
- La gestion de production (planification, ...)
- La gestion du personnel et de la paie (GRH)

En général, un ERP est subdivisé en modules qui répondent chacun à un des domaines de gestion listés ci-dessus. On dit aussi que l'ERP est constitué de modules fonctionnels, chacun couvrant un périmètre de gestion de l'entreprise.

6.1.2 Caractéristiques de l'ERP

- Il est doté d'un éditeur unique
- En cas d'impact d'un module, l'information est mise à jour en temps réel dans l'ensemble des autres modules associés

- C'est un système qui garantit la piste d'audit : il est facile de retrouver et d'analyser l'origine de chaque information
- Il peut couvrir l'ensemble du Système d'Information (SI) de l'entreprise (sauf si l'entreprise ne choisit dans un premier temps d'implémenter que certains modules de l'ERP)
- Il garantit l'unicité des informations qu'il contient, puisqu'il n'a qu'une seule base de données au sens logique.

6.2 GRH :

On pourrait définir les ressources humaines comme l'ensemble de pratiques du management, ayant pour objectif de mobiliser et développer les ressources humaines, afin d'obtenir une plus grande productivité et une meilleure qualité de travail.

Elle vise principalement la valorisation des compétences, de la motivation, l'information et l'organisation.

Elle permet, également, d'aborder la relation de l'organisation avec son collaborateur depuis son recrutement jusqu'à son départ de l'entreprise (retraite, licenciement, démission ...).

6.3 SCM :

L'expression anglaise Supply Chain Management (SCM) peut se traduire par "gestion de la chaîne d'approvisionnement" (supplychain en anglais).

L'Object du SCM est la maîtrise de l'ensemble des flux logistiques de l'entreprise. Cette notion tend même vers la maîtrise des flux de l'ensemble des acteurs de la chaîne (flux des fournisseurs et des clients inclus).

Idéalement, le SCM souhaite coordonner la gestion des flux d'un réseau en termes de coûts, délais et qualité depuis l'approvisionnement des fournisseurs jusqu'à la distribution des clients des clients.

Le credo du SCM est que l'optimum global n'est pas obtenu en sommant des optimums locaux dont l'atteinte répond bien souvent à des intérêts contradictoires (en particulier en terme de coûts / recettes). (faq-logistique.com)[8]

6.3.1 Domaines d'intervention du SCM

- L'gestion de production et les approvisionnements

- La gestion de stock et des entrepôts
- Le transport / la distribution
- La traçabilité / la gestion des retours (reverse logistics)
- La gestion des évènements imprévus (Supply Chain Event Management)

6.3.2 Les logiciels de Supply Chain Planning (SCP) et Advanced Planning System (APS)

Les SCP et APS sont des outils de prévision qui permettent de mieux gérer vos stocks afin de faire face aux variations de la demande. La volatilité croissante des marchés rend ce type d'outil de plus en plus utile. En cas de hausse anticipée, il convient de relever votre niveau de stock optimal. Au contraire, si les prévisions sont faibles, réduire les quantités permet de diminuer les coûts. Les SCP et APS prennent en considération vos contraintes en matière de ressources, de capacités de production, de délais d'approvisionnement, de livraisons et de coûts.

6.4 CRM

6.4.1 Définition :

Le sigle CRM, Customer Relationship Management, désigne un ensemble de méthodes, de pratiques et d'outils technologiques qui judicieusement mis en œuvre et bien utilisés assurent une gestion plus efficace dans la durée des relations avec les clients et les prospects. En français, on préfère parler de Gestion de la Relation client (GRC) .[6]

6.4.2 Le principe

Dans les faits, en pratique, le Customer Relationship Management se résume bien trop souvent à la seule automatisation des forces de vente (SFA Sales Forces Automation). Le concept couvre bien d'autres fonctions décomposées en trois sous-ensembles.

- **Fonction Collaboratif**

Regroupe tous les canaux d'échanges avec le client et les partenaires. Ce sont tous les canaux d'échanges (e-mail, e-conférence...) avec le client ou entre tous les partenaires à propos du client.

Ces échanges sont essentiellement destinés à améliorer la rentabilité et la fidélité client.

- **Fonction Analytique**

Analyse des informations collectées au sein du data warehouse ou datamart, data mining et statistiques en sont les outils de prédiction.

- **Fonction Opérationnel**

Intégration et automatisation des processus horizontaux en liaison avec le client FrontOffice : ventes, marketing, services clients et back-office : ERP. On retrouve dans cette catégorie tous les outils de FrontOffice, de la force de vente et de gestion des campagnes marketing, sans omettre les centres d'appels.

6.4.3 Importance d'un CRM pour une entreprise :

L'automatisation croissante des processus marketing et e-commerce peut rapidement se faire au détriment des clients. Certains d'entre eux risquent de penser que leur prise en charge et leur suivi ne sont plus appropriés, en particulier lorsque la communication est uniformisée. Une certaine insatisfaction peut en résulter. Le CRM doit atténuer ce ressenti. Dans le cadre d'une approche orientée sur le client, il a pour objectif de personnaliser la communication avec les clients, actuels ou potentiels, le plus simplement et le plus efficacement possible.

6.4.4 Les avantages du CRM :

le CRM fournit les avantages suivants :

- Gestion des contacts améliorée.
- Collaboration entre équipes.
- Productivité accrue.
- Gestion des ventes dynamisée.
- Prévission des ventes précise.
- Reporting fiable.
- Indicateurs commerciaux optimisés.
- Meilleure satisfaction et fidélisation des clients.
- Retour sur investissement marketing plus rentable.
- Meilleurs produits et services.

- Base de données complète : grâce auxquels vous pouvez déterminer les tendances de consommation de vos clients et détecter de nouvelles opportunités commerciales.

6.4.5 La CRM pour mettre le client au premier

Orientée service, la CRM, appelés aussi GRC, consiste, à la Gestion de la Relation Client.

Cet outil d'aide à la vente rassemble puis traite les informations concernant les clients de la société. Directement liée à la mise en place d'actions de marketing relationnelle, la CRM poursuit l'objectif du développement commercial à travers :

- L'automatisation des forces de vente.
- La gestion des campagnes marketing.
- La gestion du support client.

Elle imprègne toutes les composantes de l'entreprise en vue de replacer le client au centre des préoccupations et accompagne les personnels dans leur démarche de qualité.

6.5 La gestion de CRM :

Le sigle CRM, Customer Relationship Management, désigne un ensemble de méthodes, de pratiques et d'outils technologiques qui sont judicieusement mis en œuvre et bien utilisés assurent une gestion plus efficace dans la durée des relations avec les clients et les prospects. En français, on préfère parler de Gestion de la Relation client (GRC).

6.5.1 La gestion de production

- La gestion de la production est l'ensemble des activités qui participent à la conception, la planification des ressources (matérielles, financières, ou les ressources humaines), leur ordonnancement, l'enregistrement et la traçabilité des activités de production, le contrôle des activités de production de l'entreprise.

La gestion de production manipule un nombre très important de données. Elle est donc par nature liée au système d'information de l'entreprise et à l'offre logicielle présente sur le marché qui a considérablement évoluée ces dernières années. En effet, alors que très longtemps elle a été concentrée autour de la GPAO (Gestion de production assistée par ordinateur) et des logiciels d'ordonnancement et de suivi de production, on a vu apparaître de nombreux sigles nouveaux (ERP, SCM, APS, MES...). Cette évolution correspond à la fois à une évolution des fonctions de bases intégrées dans les logiciels de GPAO, mais également à une intégration de fonctionnalités connexes qui a considérablement modifié la.[7]

6.5.2 La gestion commerciale, La logistique

L'ASLOG (Association française pour la logistique) définit la logistique comme Etant "l'art et la manière de mettre à disposition un produit donné au bon moment, au bon endroit, au moindre coût et avec a meilleure qualité".

La logistique regroupe l'ensemble des activités qui permettent de gérer les flux physiques et d'information dans le but d'en minimiser les coûts et ce de l'amont à l'aval de la "chaîne logistique" en respectant des conditions satisfaisantes en termes de délais et de qualité.[43]

6.5.3 La comptabilité

La comptabilité est la science des comptes, qui a pour but d'organiser les écritures par le choix judicieux des livres et des comptes nécessaires afin d'en suivre facilement les opérations et d'en présenter les résultats En tant que technique elle fournit un certain nombre de renseignements chiffrés d'ordre juridique et économique exprimé dans des états appelés comptes.

Conformément au code de commerce toute personne physique ou morale exerçant l'activité commerciale industriel etc...Doit tenir une comptabilité

La comptabilité est une technique quantitative de gestion normalisé qui consiste à saisir et mesurer des mouvements affectant l'exploitation ou la structure d'une entreprise et résultant soit de son activité interne, soit de ses relations avec l'extérieur .[8]

Dans la pratique le processus consiste à collecter, analyser évaluer enregistrer, calculer présenter et interpréter ces mouvements exprimés en termes monétaires.

Rôle de comptabilité : Le rôle de comptabilité se situe généralement a quatre niveaux :

6.5.3.1 Pour l'entreprise

- Elle est un outil de connaissance du résultat d'exploitation exprimé successivement en marge brute valeur ajoutée, résultat d'exploitation, résultat hors l'exploitation et résultat de l'exercice.
- Elle permet de connaître l'évolution de la situation patrimoniale.
- Elle fournit des informations à la comptabilité analytique.
- Elle est la base de l'analyse financière.

6.5.3.2 Pour les tiers

- Elle est un outil d'information des tiers avec lesquelles l'entreprise est en relation, clients, fournisseurs, banques, actionnaires, public....

6.5.3.3 Pour la nation

- Elle fournit à la comptabilité nationale les informations qui permettent l'évaluation du revenu national.

6.5.3.4 Pour l'administration fiscale :

- Elle est un outil de sémination du l'assiette et résultat fiscale imposable.

7. les paniers d'achat

7.1 Introduction

Servir le client Un enjeu que les entreprises mesurent, mais pas toujours avec une parfaite acuité. Ce qui paraît évident, c'est que plus on collecte de données, mieux on va ensuite pouvoir servir son client. Daniel Benoild, manager en consulting analytique chez PwC, relativise : « Ce n'est pas bêtement mathématique : 1% de données recueillies en plus, ça n'apporte pas forcément exactement 1% de service en plus. Tous ces éléments ne se valent

Pas : l'important, c'est donc de les collecter, certes, mais aussi et surtout d'avoir un modèle pertinent pour mieux les exploiter ». Concrètement, aucune donnée particulière n'a plus de valeur qu'une autre. Tout va en fait dépendre des problématiques que l'entreprise souhaite résoudre. En fonction d'elles, des consultants extérieurs pourront aider à la mise en place de modèles d'analyse adéquats, pour une bonne définition et une meilleure exploitation des différentes données collectées. Ces données permettent par exemple de mieux connaître le consommateur-type et son panier moyen ou d'identifier les attentes de ses cibles. [9]

7.2 Définition

Quand il y a une entreprise et des clients, c'est-à-dire qu'il y a des transactions entre eux, que ce soit à l'achat ou à la vente, et quel que soit le type de bénéfices, car ces transactions aboutissent à ce qu'on appelle un panier d'achats.

Le panier peut être une véritable affaire entre le vendeur et l'acheteur ou un espace virtuel sur des sites de e-commerce qui recueille tous les achats que le client a choisi et qui ont suscité son intérêt

7.3 Comment l'entreprise comprend les besoins du client via le panier d'achat

Ce qui compte pour chaque entreprise est d'augmenter ses profits et ses revenus et son émergence sur le marché. Pour que cet aspect soit atteint, l'entreprise doit comprendre les désirs et les actions de ses clients ce qu'ils aiment et quels sont les biens les plus demandés par eux en plus de distinguer les moments de leurs besoins pour des produits spécifiques et les moments où ils n'en ont pas besoin.

Par exemple, l'entreprise de télécommunications a pu connaître le comportement de ses clients, par exemple s'ils veulent renoncer à leurs services, en étudiant les changements survenus aux clients qui les ont précédés, qui ont effectivement renoncé à leurs services, et donc accorder plus d'attention à cette catégorie et trouver des moyens de regagner leur confiance, ainsi qu'étudier dans les magasins les habitudes et les comportements de leurs clients pour les comprendre. Plus et ainsi les gagner.

Nous ne pouvons pas comprendre le client et prédire les comportements à venir, sauf en triant ses paniers d'achat, c'est-à-dire que plus il est analysé, plus il y a de paniers, plus les clients comprennent leur impact et donc plus de profit

8. Conclusion

Dans ce chapitre, nous avons parlé du côté économique des entreprises, de la façon dont elles ont démarré et des problèmes les plus importants auxquels elles étaient confrontées, et nous avons mis en lumière le taux d'assistance et de solutions que l'informatique a fourni à ces entreprises pour aider à réduire considérablement leurs problèmes et contribuer à gagner la confiance des clients, dont le plus important est le data mining , qui a beaucoup contribué. En augmentant les connaissances et en facilitant les affaires. Alors, qu'est-ce que le datamining et quelles sont ses taches les plus importantes.

Tout cela va le connaître dans le prochain chapitre.

Chapitre 2 :

Data mining

1.Introduction

Comme nous l'avons appris auparavant, l'objectif principal de chaque entreprise est de dégager des bénéfices, et pour y parvenir, elle doit obtenir la satisfaction de ses clients, et pour cela, elle collecte le plus grand nombre possible des données, mais il n'est pas nécessaire que toutes les données soient utiles et puissent aider et, par conséquent, les entreprises ici recourent à des spécialistes dans ce domaine pour effectuer le processus de filtrage.

Les données et la sortie sont plus avantageuses pour y appliquer des études pour pouvoir analyser le comportement du client, l'un des moyens les plus utilisés pour rechercher dans ces données et prospecter en elle est ce qu'on appelle l'ECD rencontre lors de l'extraction des données. Enfin nous terminons par la conclusion.

2. Processus d'extraction de connaissances(ECD) :

En 1996, dans les actes de la première conférence internationale sur le KDD a donné l'une des définitions les plus connues de KDD : «Le non trivial processus d'identification de modèles valides, nouveaux, potentiellement utiles et finalement compréhensibles dans les données. »[15]

Dans cette partie, nous allons donner un aperçu général sur le processus ECD (définition, étapes. . .) notamment sur l'étape fouille de données (data mining).., les techniques utilisées (motif, règle d'association, classification . . .) et différents types de base de données qu'on Ce processus est cyclique et permanent.

”L'Extraction de Connaissances à partir des Données (ECD) est un processus itératif et interactif d'analyse d'un grand ensemble de données brutes afin d'en extraire des connaissances exploitables par un utilisateur analyste qui y joue un rôle central”[16]

Il s'agit de proposer des connaissances nouvelles qui enrichissent les interprétations du champ d'application, tout en fournissant des méthodes automatiques qui exploitent cette information.

2.1 Le principe

Le data mining recouvre des techniques d'analyse pour trouver des tendances ou des corrélations cachées parmi des masses importantes de données ou encore pour détecter des

informations stratégiques ou de nouvelles connaissances, en s'appuyant sur des méthodes de traitement statistique avec un zeste d'intelligence artificielle.

Le data mining peut, selon les besoins, intervenir sur votre connaissance client (valeur et potentiel) et/ou sur l'optimisation de vos produits et actions commerciales. [44]

2.2 Les étapes principales d'ECD

- **Sélection** : La sélection des données est le processus où les données pertinentes pour la tâche d'analyse sont récupérées de la base de données.

- **Prétraitement** : en raison de la taille énorme, des multiples sources hétérogènes et de diverses formes, les données sont très sensibles au bruit, manquant et incohérent. Des données de faible qualité entraîneront des résultats d'extraction de faible qualité. Dans cette étape, nous essayons de prétraiter les données pour améliorer leur efficacité et faciliter le futur processus d'exploration. Il existe plusieurs techniques de prétraitement :

1. **Nettoyage des données** : pour supprimer le bruit et corriger les incohérences
2. **Intégration de données** : pour incorporer plusieurs formulaires et sources de données dans un magasin de données cohérent tel qu'un entrepôt de données.
3. **Réduction des données** : pour réduire la taille des données.

- **Transformation** : Dans cette étape, les données seront transformées ou consolidées sous des formes appropriées pour l'exploration, en effectuant des opérations de résumé ou d'agrégation.

- **Fouille de données** : est la partie principale du processus KDD, ici nous utilisons des techniques et des outils efficaces pour exploiter des connaissances efficaces, ces techniques et outils et d'autres aspects de la fouille de données sont discutés à l'étape suivante.

- **Interprétation** : à cette étape certains modèles sont obtenues, tous les modèles ou résultats ne sont pas tolérables, nous devons donc évaluer l'utilité des résultats. Sinon, les modèles sont présentés dans le format, qui est compris par l'utilisateur, et de précieuses connaissances peuvent être découvertes.

Il y a un point fondamental dans l'ECD, qui est le point le plus important, et qui nous aide dans les manières que nous aborderons plus tard dans l'apprentissage, la lecture du client, l'étude de son comportement et la prévision de ce qu'il fera plus tard, ce qui est la fouille de données.

La figure 2 présente ces étapes ainsi que les enchaînements possibles.

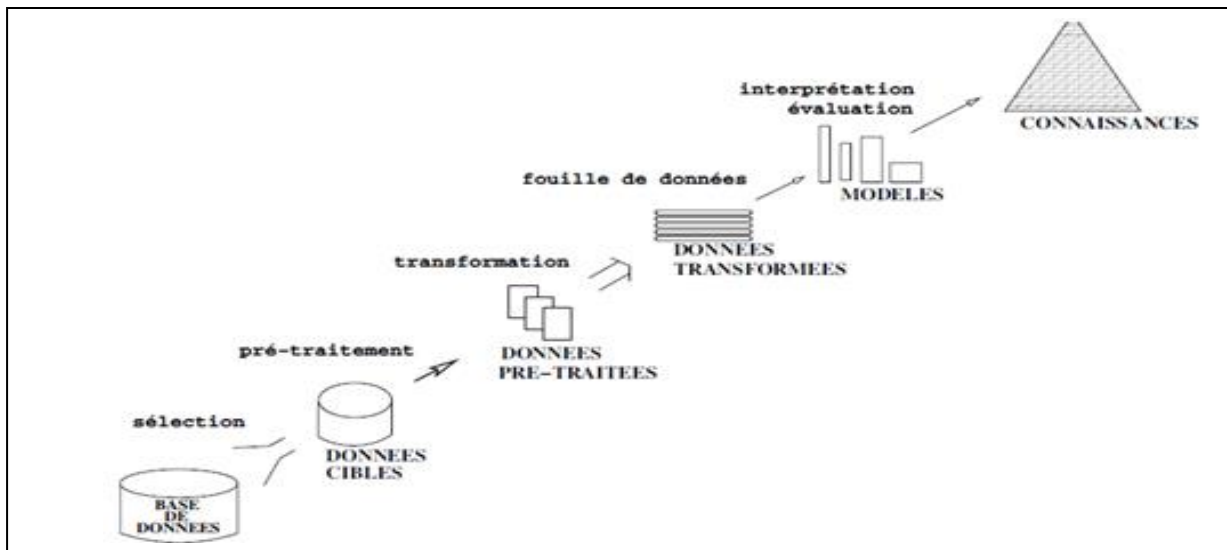


Figure 2 : processus de l'extraction de connaissance [17]

3.Fouille de données (Data mining)

3.1. Introduction :

Historiquement, le Data Mining est très jeune. Le concept apparaît en 1989 sous un premier nom de KDD (Knowledge Discovery in Data bases, en français ECD pour Extraction de Connaissances à partir des Données), avant qu'en 1991 apparaisse pour la première fois le terme de Data Mining Comme l'expliquent fort bien Michael Berry et Gordon Linoff, ce concept tel qu'on l'entend aujourd'hui, et surtout tel qu'on l'applique dans les services marketing.

Il est étroitement lié au concept du « one-to-one relationship ». C'est à dire la personnalisation des rapports entre l'entreprise et sa clientèle.

La Fouille de Données, ou de manière générique l'Extraction de Connaissance à partir de Données (en anglais Data Mining and Knowledge Discovery in Data bases), est un domaine de recherche qui a véritablement pris son essor au milieu des années 90. S'il est toujours possible de discuter quant à sa véritable originalité par rapport au traitement statistique des données qui existe déjà depuis longtemps [18], il est indéniable en revanche que son avènement s'est accompagné d'une forte accélération de la diffusion de logiciels spécialisés

Est pampilles Data Mining. Les raisons sont multiples ; on pourra citer entre autres : la rencontre entre des Communautés différentes (apprentissage automatique, bases de Données, analyse de données, statistiques) ; le développement d'Internet qui a permis la diffusion à peu de frais des logiciels, avec pour certains des codes sources ; l'élargissement du champ d'application du traitement des données telles que la catégorisation de textes et d'images, la bio-informatique, etc. [19]

3.2. Définition :

Le Data mining, est le cœur de processus ECD, et c'est un domaine qui est apparu avec l'explosion des quantités de données stockées, Il consiste à l'analyse de grands ensembles de données observationnelles pour découvrir des modèles et des relations intéressantes et utiles. Le domaine combine des outils de statistiques et d'intelligence artificielle (tels que les réseaux de neurones, les règles d'associations, les arbres de décisions,..) avec la gestion de bases de données pour analyser de grandes collections numériques, appelées ensembles de données. L'exploration de données (fouille de données) est largement utilisée dans les domaines des affaires (assurances, banque, commerce de détail), de la recherche scientifique (astronomie, médecine) et de la sécurité publique (détection des criminels et des terroristes). [20]

3.3 Le Data Mining un avantage compétitif pour toutes les entreprises

A l'ère du Big Data où l'espace de stockage des données n'est plus vraiment un problème, toutes les entreprises veulent désormais tirer parti de leurs grands volumes de données. Ces données peuvent les aider à comprendre leur environnement interne (RH, organisation, process...), leur environnement externe (type de clients, parcours clients, image de l'entreprise...) et à anticiper les phénomènes qui s'y rattachent. Elles deviennent par conséquent une grande richesse si elles sont bien exploitées. C'est justement l'objectif du Data Mining.

Les études de Data Mining menées dans les entreprises ont des objectifs très variés.

Elles peuvent par exemple aboutir :à une stratégie marketing différenciée par types de clients grâce à l'élaboration d'une segmentation comportementale, à l'optimisation de l'efficacité des actions marketing et commerciales grâce à une segmentation stratégique, à une

efficacité accrue des campagnes marketing : e-mailing, SMS réseaux sociaux... grâce au ciblage des clients à fort potentiel, à un investissement commercial adapté et optimisé grâce à une prédiction du potentiel de vente par zone géographique, à un diagnostic de la relation clients grâce à une analyse TextMining de postes sur les réseaux sociaux [21].

3.4 Domain d'application

La fouille de données a aujourd'hui une grande importance économique du fait qu'elle permet d'optimiser la gestion des ressources humaines et matérielles. Elle est utilisée par exemple dans différents domaines on cite :

3.4.1 La banque

En ce qui concerne la banque, il existe plusieurs utilisations pour l'extraction de données, notamment :

- Identification des clients les plus actifs
- Identification des clients stagnants qui ne remboursent pas les prêts
- Tentative de connaître la réaction des clients face à la variation du taux d'intérêt Détection des fraudes en modifiant les relations du client avec les cartes de crédit
- Essayer de savoir les clients dont la fréquence peut changer d'affiliation dans la période à venir [45].

3.4.2 L'assurance

Dans le domaine de l'assurance, nous pouvons utiliser les éléments suivants :

- Des indicateurs pour détecter les risques tels que les produits les plus à risque en plus de la fraude
- Déterminer les produits obligatoires tels que les voitures et non obligatoires tels que le logement
- Extraire les propriétés de ces publicités (type d'accident, blessures, etc.)[46].

3.4.3 .Le commerce

En ce qui concerne le commerce, il peut être utilisé pour en apprendre davantage sur les risques découlant de chaque transaction, ses avantages et ses inconvénients, connaître les clients les plus fidèles et leurs crédits en ce qui concerne les opérations de crédit, évaluer les concurrents et essayer de les identifier [47].

3.4.4 .La détection de fraude

- Détection des fraudes par carte de crédit
- Fausse demande de compensation médicale
- La détection des fraudes dans les demandes de passeport par la conception d'un système de diagnostic par apprentissage en ligne [46].

3.4.5 La médecine et la pharmacie

Dans le domaine médical, le data mining est utilisée pour :

- Déterminer les caractéristiques de chaque maladie et les indicateurs pour chacune d'elles
- Distinguer les symptômes rares qui peuvent distinguer les symptômes graves et ainsi déterminer le traitement le plus approprié
- Prédire les symptômes et les effets des nouveaux médicaments
- La possibilité de découvrir de nouvelles maladies en observant les symptômes similaires aux cas [48].

3.5. Processus de fouille de données

Il passe par plusieurs étapes, qui sont :

- 1-comprendre le problème.
- 2-collecte des données.
- 3-prétraitement.
- 4-estimer le modèle.
- 5-interpréter le modèle et établir les conclusions.

3.6. Principales taches de fouille de données

Nous avons des données structurées. Les objets sont représentés par des enregistrements (ou descriptions) qui consistent en un ensemble de champs (attributs) prenant des valeurs dans un champ. Plusieurs tâches peuvent être liées au "data mining", on peut par exemple indiquer :

- 1-La classification.
- 2-L'estimation.
- 3- La prédiction.

4- Les règles d'associations.

5- La segmentation.

La figure suivante présente les taches de la fouille de donnés

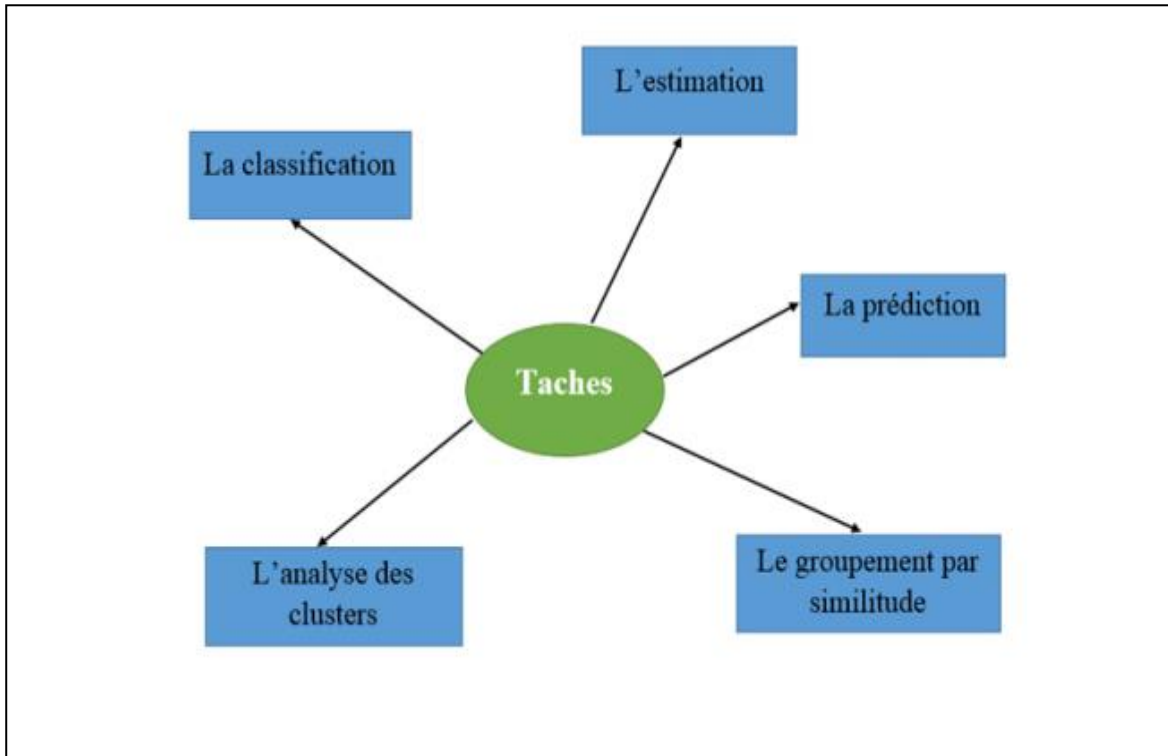


Figure 3 : Les taches de la fouille de données [22].

La fouille de données ou bien le data mining définit plusieurs techniques et ses différentes méthodes et cela est illustré dans La figure suivante La figure suivante présente :

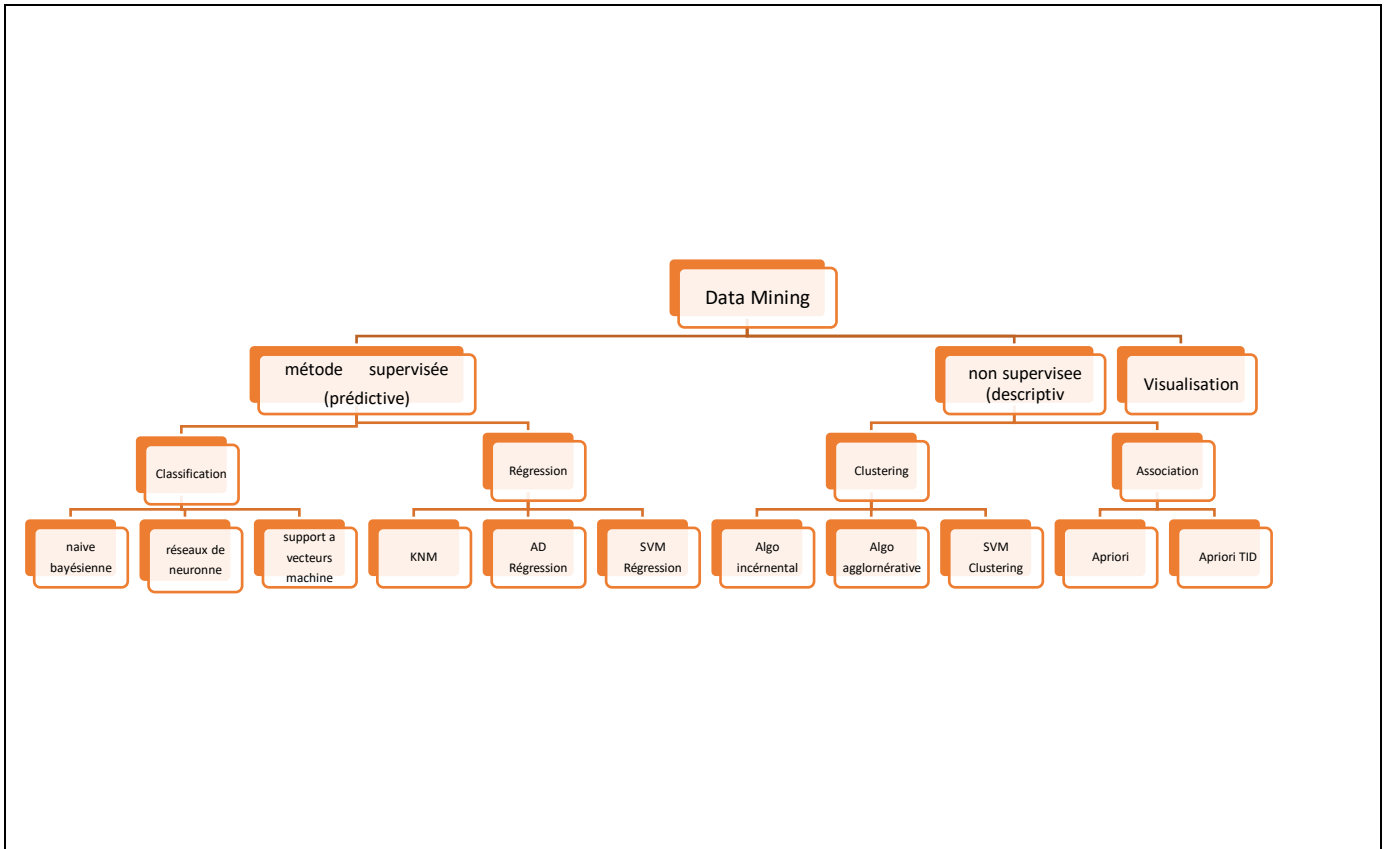


Figure 4 : Les Technique [23].

3.6.1 .Classification

La classification est la tâche la plus commune de la fouille de données. Elle consiste à examiner les caractéristiques d’un objet et lui attribuer une classe, la classe est un champ particulier à valeurs discrètes. Des exemples de tâche de classification sont :

- 1– attribué ou non un prêt à un client
- 2– établir un diagnostic
- 3– accepter ou refuser un retrait dans un distributeur

Il existe deux types :

3.6.1.1 Classification supervisée (apprentissage)

Elle sert à obtenir un critère de séparation destiné à prédire l'appartenance à une classe tel que le nombre de classe est défini a priori.

On présente quelque méthode de classification supervisée :

- 1- Analyse discriminante.
- 2- k-Plus proches voisins.
- 3- Arbre de décision.
- 4- Réseau de neurones.

3.6.1.2 Classification non supervisée

Elle sert à établir des représentations des données dans des espaces à faible dimensions pour y lire des typologies d'individus tels que le nombre de classes n'est pas connu.

On présente quelque méthode de classification non supervisée :

- 1-Analyse en composante principale (ACP).
- 2-Analyse des associations.
- 3-Analyse en clusters[24].

3.6.2. L'estimation

Elle consiste à estimer la valeur d'un champ à partir des caractéristiques d'un objet. Le champ à estimer est un champ à valeurs continues. L'estimation peut être utilisée dans un but de classification. Il suffit d'attribuer une classe particulière pour un intervalle de valeurs du champ estimé. Des exemples de tâche d'estimation sont :

- 1- Estimer les revenus d'un client.
- 2- Estimer les risques[25].

3.6.3. La prédiction

Cela consiste à estimer une valeur future. En général, les valeurs connues sont historiques. On cherche à prédire la valeur future d'un champ. Cette tâche est proche des précédentes. Les méthodes de classification et d'estimation peuvent être utilisées en prédiction. Des exemples de tâches de prédiction sont :

- 1 - Prédire les valeurs futures d'actions.
- 2 - Prédire, au vu de leurs actions passées, les départs de clients. [2]

3.6.4. Les règles d'associations

Cette tâche, plus utilisée dans l'analyse du panier de la ménagère, consiste à déterminer les variables qui sont associées. Elle consiste à rechercher les relations ou les dépendances existantes entre plusieurs caractéristiques d'un individu. L'exemple type est la détermination des articles qui se retrouvent ensemble sur un même ticket de supermarché. Cette tâche peut être effectuée pour identifier des opportunités de vente croisée et concevoir des groupements attractifs de produit [26].

3.6.5. La segmentation

Consiste à former des groupes (clusters) homogènes à l'intérieur d'une population. Pour cette tâche, il n'y a pas de classe à expliquer ou de valeur à prédire définie a priori, il s'agit de créer des groupes homogènes dans la population définie par un ensemble des enregistrements. Il appartient ensuite à un expert du domaine de déterminer l'intérêt et la signification des groupes ainsi constitués. Cette tâche est souvent effectuée avant les précédentes pour construire des groupes sur lesquels on applique des tâches de classification ou d'estimation.

3.7 Apprentissage supervisé Vs Apprentissage non supervisé

3.7.1. Les techniques supervisées :

Appartiennent à la modélisation prédictive. Dans l'apprentissage supervisé (classification), les données d'apprentissage obtenues par des observations, ou des mesures sont accompagnées d'étiquettes indiquant la classe de l'observation. Les nouvelles données sont classées en fonction de l'ensemble de formation.

3.7.2. Les techniques non supervisées :

Appartiennent à la modélisation descriptive. Dans l'apprentissage non supervisé (clustering), la classe de données d'apprentissage est inconnue. Nous cherchons à trouver des classes ou des groupes d'observations ou de mesures.

3.8. Les méthodes de Data Mining

Dans le domaine de datamining il existe des méthodes spécifiques pour un problème particulier. On peut les classer en fonction de :

- 1-La tâche à réaliser.
- 2-Le type et la nature des données.
- 3-L'ensemble des connaissances et des compétences disponibles
- 4-La finalité du modèle construit

5-L'environnement social, technique, philosophique de l'entreprise

On distingue deux grandes catégories de méthodes d'analyse consacrées à la fouille de données [27] :

3.8.1. Les méthodes classiques :

On y retrouve des outils généralistes de l'informatique ou des mathématiques :

- Les requêtes dans les bases de données, simples ou multicritères, dont la représentation est une vue,
- Les requêtes d'analyse croisée, représentées par des tableaux croisés.
- Les différents graphes, graphiques et représentations.
- Les statistiques descriptives.
- L'analyse de données : analyse en composantes principales(ACP).

3.8.2. Les méthodes sophistiquées :

Elles ont été élaborées pour résoudre des tâches bien définies, nous citons :

- Les algorithmes de segmentation.
- les règles d'associations.
- Les algorithmes de recherche du plus proche voisin.
- Les arbres de décision.
- Les réseaux de neurones.

- Les algorithmes génétiques.

Nous présentons par la suite les méthodes de fouille de données sophistiquées les plus utilisées actuellement dans l'analyse de données surtout pour ce qui concerne la relation client.

3.8.2.1. Arbre de décision

Développées autour de 1960 (AID de Morgan et Sonquist) et très utilisées en marketing, ces méthodes délaissées par les statisticiens ont connu un regain d'intérêt avec les travaux de Breiman et al. (1984) qui en ont renouvelé la problématique : elles sont devenues un des outils les plus populaires du data mining en raison de la lisibilité des résultats [28].

Fournit des règles explicites de classement et supporte bien les données hétérogènes, manquantes et les effets non linéaires, actuellement la seule grande concurrente de l'arbre de

décision est la régression logistique, cette méthode étant préférée dans la prédiction du risque en raison de sa plus grande robustesse. Remarquons que les arbres de décision sont à la frontière entre les méthodes prédictives et descriptives, puisque leur classement s'opère en segmentant la population à laquelle ils s'appliquent : ils ressortissent donc à la catégorie des classifications hiérarchiques descendantes supervisées.

3.8.2.2. Les réseaux de neurones

Les réseaux de neurones sont apparus dans les années cinquante avec les premiers perceptrons, et sont utilisés industriellement depuis les années quatre-vingt. Un réseau de neurone a une architecture calquée sur celle du cerveau, organisée en neurones et synapses, et se présente comme un ensemble de nœuds ou neurones formels, ou unités connectés entre eux, chaque variable prédictive continue correspondant à un nœud d'un premier niveau, appelé couche d'entrée, et chaque variable prédictive catégorique correspondant également à un nœud de la couche d'entrée.

Le cas échéant, lorsque le réseau est utilisé dans une technique prédictive, il y a une ou plusieurs variables à expliquer, elle correspondant alors chacune à un nœud (ou plusieurs dans le cas des variables catégorielles) d'un dernier niveau : la couche sortie. Les réseaux prédictifs sont dits à apprentissage supervisé et les réseaux descriptifs sont dits à apprentissage non supervisé. Entre la couche d'entrée et la couche de sortie sont parfois connectés à des nœuds appartenant à un niveau intermédiaire : la couche cachée. Il peut [29].

Un nœud de perceptron se présente donc comme suit :

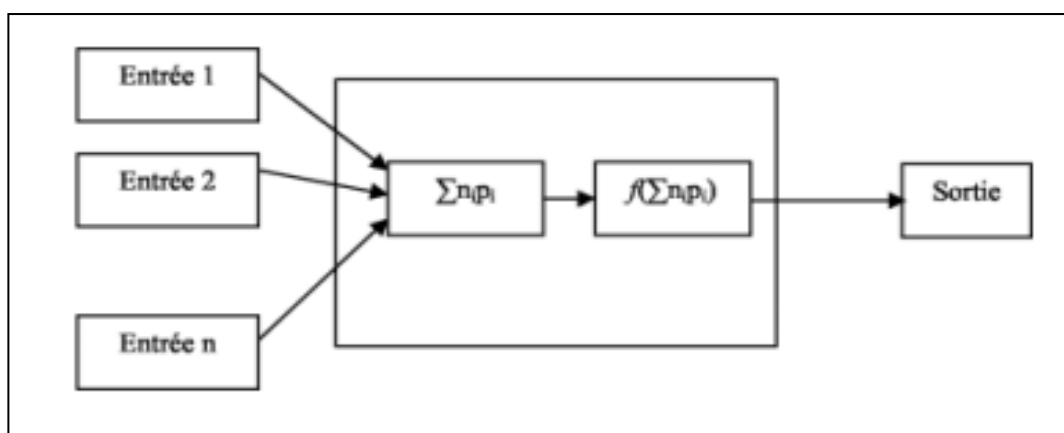


Figure 5 : Nœud [32].

Dans cette figure, on utilise les notions suivantes :

- n_i est la valeur du nœud i du niveau précédent (la sommation sur i correspond à l'ensemble des nœuds du niveau précédent connectés au nœud observé).
- p_i est le poids associé à la connexion entre i et le nœud observé.
- f est la fonction de transfert associée au nœud observé.

Parmi les applications des réseaux de neurones on cite :

- Statistiques : analyse de données, prévision, classification
- Robotique : contrôle et guidage de robots ou de véhicules autonomes.
- Imagerie, reconnaissance de formes.
- Traitement du signal.
- Simulation de l'apprentissage [31].

Les réseaux de neurones sont des outils très utilisés dans les méthodes de data mining puisque qu'ils sont un outil d'apprentissage par excellence.

3.8.4. Règles d'associations

Les règles d'associations présentent une méthode d'apprentissage automatique basée sur des règles permettant de découvrir des relations intéressantes entre des variables dans de grandes bases de données. Il est destiné à identifier les règles fortes découvertes dans les bases de données en utilisant certaines mesures d'intérêt tel que la confiance, ou la conviction qui permet de présenter la coexistence des deux parties (gauche et droite) de la règle et d'après cette mesure on obtient un ensemble de règles solides qui dépassent un minimum de support et un minimum de qualité. Sur la base du concept de règles strictes, l'apprentissage par l'association a introduit des règles d'association permettant de détecter les régularités entre les observations dans des données de rangées à grande échelle.

La recherche des règles d'associations parmi les méthodes non supervisées car on dispose en entrée que de la description des achats [33].

Les règles d'associations sont principalement faciles à l'utilisation et définissent comment les produits ou des services se situent les uns par rapport aux autres.

3.9 Etapes de création d'un projet data mining

- Choix du sujet

- Inventaire des données existantes
- Collectes, nettoyage et mise en forme des données
- Etudes statistique de base d'analyse
- Mise en œuvre des algorithmes
- Elaboration des modèles
- Validation et choix d'un modèle
- Déploiement du modèle
- Formation des utilisateurs
- Suivi des modèles

4. Conclusion

L'extraction de données est un vaste domaine qui permet d'analyser de grandes et énormes quantités de données, car elles sont utilisées dans divers domaines de l'informatique et de l'intelligence artificielle pour l'extraction des connaissances et la prise de décision, que nous borderons plus en détail dans le prochain chapitre.

Chapitre 3:

**Les règles d'associations
pour l'analyse de panier
d'achat**

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

1. Introduction

Aujourd'hui, l'extraction de règles d'association est l'une des tâches les plus répandues dans l'extraction de données en raison de son rôle important et de son importance.

L'une des applications pour extraire ces règles est l'analyse de paniers d'achats et son objectif est de mettre en évidence les relations entre les fonctionnalités de la base de données pour comprendre les habitudes de consommation des clients en fonction de leurs besoins, en mettant en avant les offres promotionnelles afin d'améliorer les intrants.

Dans la base de données, la transaction est un groupe de produits que le client a acquis, appelés attributs, et donc la règle de corrélation est un lien entre plusieurs attributs avec l'existence de la condition. En d'autres termes, le but de l'extraction de règles est de découvrir comment un ensemble de fonctionnalités est lié à l'autre.

Il existe plusieurs techniques et algorithmes utilisés pour l'extraction de règles d'associations, dans ce qui suit nous allons présenter les plus connus.

2. Les règles d'associations

Une règle d'association est une relation d'occurrence qui lie deux ou plusieurs items particuliers. Ceci s'illustre plus facilement à l'aide d'un exemple. Si le contenu des paniers d'achats révèle qu'un grand nombre de clients qui achètent du pain, achètent également du beurre, cela signifie qu'il existe une relation d'occurrence simultanée forte entre ces deux items. Cette relation est appelée association et de celle-ci s'induit la règle d'association suivante : pain => beurre.

Cette règle d'association est dite classique ou traditionnelle, puisque les deux événements, l'achat du pain et l'achat du beurre, se déroulent simultanément lors de la même transaction.

Lorsque l'occurrence des items est étalée selon un ordre spécifique, par exemple les clients qui achètent du pain, achètent également du beurre dans les jours qui suivent, les règles d'associations entre ces items ordonnés sont qualifiées d'associations séquentielles.

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

L'objectif des règles d'associations, traditionnelles ou séquentielles, consiste à découvrir les associations fréquentes dans les données étudiées. [34]

Les règles d'associations sont utilisées principalement dans l'analyse du panier d'achat

3. Concepts de support et de confiance

Les règles d'association sont des une techniques qui tentent de révéler et de trouver les relations importantes et distinctes entre différents ensembles de données, ou ce que l'on appelle des grandes bases de données [35].

Les règles d'association sont définies comme des techniques qui tentent de trouver des relations importantes et distinctes entre un groupe dans un ensemble de données [35].

Ces règles sont devenues la façon dont nous connaissons actuellement l'analyse des données par deux articles publiés par Agrawal et al. En 1993 et 1994. Au début, dans le premier article, ils ont discuté des concepts et des problèmes entourant les règles d'association, tandis que dans le second, ils ont parlé de l'algorithme apriori TID et à partir de ce moment, cet algorithme joue un grand rôle dans les règles d'association afin que nous le trouvions dans les cas les plus compliqués

Une règle d'association s'énonce comme suit :

Pain \Rightarrow beurre [support = 2%, confiance = 60 %].

La règle d'association précédente se lit de la manière suivante 60% des clients qui ont acheté le pain ont acheté du beurre avec lui, et 2% des transactions de la base de données correspondent à ce qui a été mentionné précédemment.

La confiance se définit comme la probabilité d'obtenir dans une transaction l'item résultant de la règle, dans ce cas-ci du beurre, sachant qu'un ou que d'autres items, dans ce cas-cile pain, s'y retrouvent également.

Le support, quant à lui, est donné par la proportion des transactions de la base de données transactionnelles qui contiennent tous les items présents dans la règle.

Comment le calculer le support et la confiance :

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

Soit A un motif ou l'itemset, et nous avons une base de données de 20 transactions

Et il a répété l'élément A dans 7 transactions dans cette base, puis

$$\text{Support (A)} = \frac{\text{Le nombre de transactions qu'il contient A}}{\text{Le nombre de transactions dans la base de données}} = \frac{7}{20}$$

Soit la règle $A \rightarrow B$, à partir d'une base de données de 20 transactions, nous avons 5 transactions dans lesquelles l'article A été mentionné

Et 3 transactions dans lesquelles A est mentionné avec B. donc

$$\text{Confiance (A} \rightarrow \text{B)} =$$

$$\frac{\text{Le nombre de transactions mentionnées par les deux parties de la règle (AB)}}{\text{Le nombre de transactions dans lesquelles le côté gauche de la règle est mentionné (A)}} = \frac{3}{5}$$

Ce petit exemple démontre l'avantage principal des règles d'associations : leur simplicité. Contrairement à d'autres méthodes d'analyse de données, les règles d'associations fournissent des réponses simples et facilement interprétables quel que soit le degré de complexité de la règle [36].

4. Algorithmes :

L'extraction de règles d'association est l'un des principaux problèmes du KDD. Ce problème fut développé pour l'analyse de bases de données de transactions de ventes, chacune constituée d'une liste d'articles achetés, afin d'identifier les groupes d'articles achetés le plus fréquemment ensemble. Une règle d'association permettra de définir une stratégie commerciale dans le but d'augmenter les ventes, en plaçant les articles qui ont tendance à être acheté ensemble dans le même rayonnage du magasin. Afin de ne générer que les relations significatives entre les ensembles d'articles, des mesures d'utilité (le support) et de précision (la confiance), empruntées aux statistiques, sont associées à chaque règle d'association. Les règles générées sont celles dont le support et la confiance sont supérieurs ou égaux à des seuils minimaux définis par l'utilisateur en fonction de ses objectifs et du type de données traitées [49].

La technique d'extraction de règles d'association introduite par Agrawal est considérée comme l'une des techniques les plus importantes dans la découverte de connaissances dans les bases de données. Deux étapes principales composent ces algorithmes :

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

1. La recherche des itemsets fréquents ;
2. La génération des règles d'association.

Différentes études soulignent que ce découpage correspond à la manière optimale face à l'explosion combinatoire de l'algorithme. Ainsi, la recherche des itemsets fréquents sur un ensemble m d'attributs binaires est un espace de dimension 2^m . L'étape de génération des règles à partir d'un itemset fréquent de largeur k peut engendrer la génération de $2^k - 2$ règles [37].

À partir de l'algorithme de référence Agrawal. Différents algorithmes ont également été développés pour réduire le nombre d'éléments en créant des éléments fermés, maximaux ou optimaux. D'autres algorithmes ont été développés pour réduire un certain nombre de règles. Enfin, diverses méthodes complémentaires ont été suggérées après le traitement, telles que l'élagage et les résumés ou les combinaisons [38].

4.1. Les itemsets

Les algorithmes d'extraction considèrent un ensemble d'item sets d'une taille donnée lors de chaque itération. Ces algorithmes limitent le nombre d'item sets candidats considérés, en les générant à partir des items sets fréquents de l'itération précédente : « tous les sur-ensembles d'un itemset infrequenté sont infrequentés et tous les sous-ensembles d'un itemset fréquent sont fréquents ». [40]

La recherche des itemsets fréquents est un problème non trivial car le nombre d'itemsets potentiellement fréquents est exponentiel par rapport au nombre d'items considérés dans la base de données. On appelle itemset fréquent un itemset dont le support en d'autres termes, le nombre de lignes dupliquées pour cet article ou itemset est supérieur à un seuil minsup fixé. Si le support n'est pas atteint, on dit que l'itemset est infrequenté.

Soit I , l'ensemble des items, et minsup un seuil minimal de support=2. L'ensemble F des itemsets fréquents. [50]

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

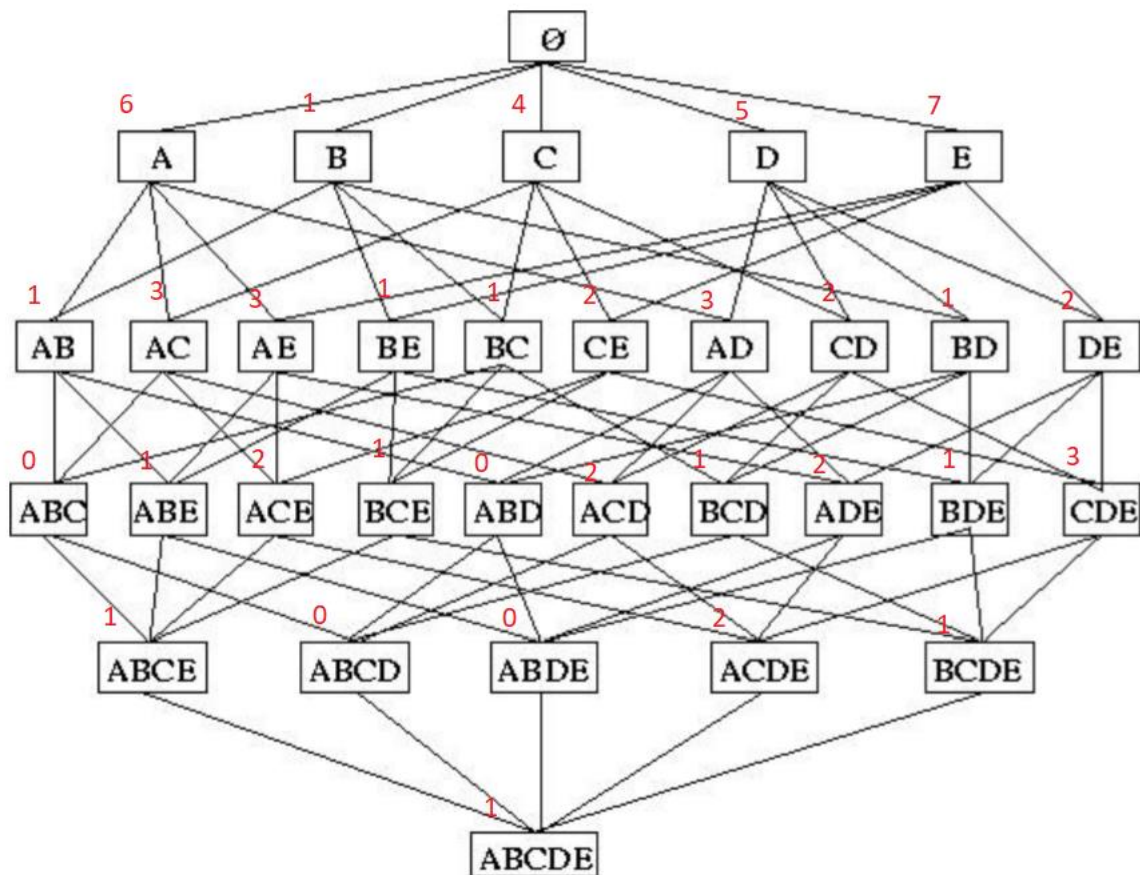


Figure 6 : Treillis d'itemsets

La figure 1 présente le treillis des itemsets pour l'ensemble $I = \{A, B, C, D, E\}$.

Après avoir formé un Treillis d'itemsets de sorte que chaque ligne du Treillis représente une génération d'itemsets et le support pour chaque itemset est calculé. On supprime les itemset dont le support est inférieur au minsup, ne laissant que les itemsets fréquents. Comme le montre la figure ci-dessous :

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

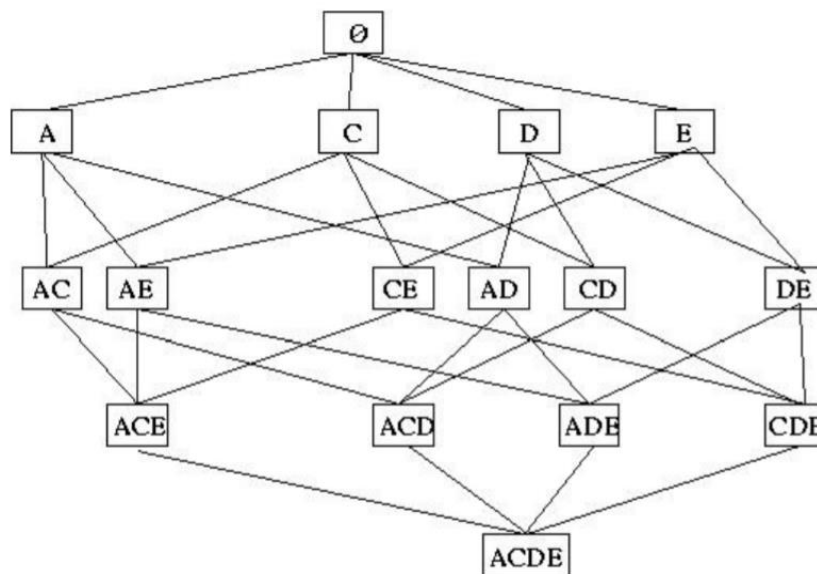


Figure 7 : Treillis composé des itemsets fréquents.

Des itemsets fréquents maximaux

Un motif fréquent est dit Maximal si aucun de ses sur-motifs immédiats n'est fréquent

Des itemsets fréquents fermés

Un motif fréquent est dit fermé s'il ne possède aucun sur-motif qui a le même support.

Paramètre commun

Le paramètre commun de ces trois concepts est la détermination du seuil de support minimal assurant la génération des itemsets fréquents. Cette approche présente tout de même un inconvénient : un seuil de support trop faible peut conduire à la génération d'un nombre trop importants d'itemsets tandis qu'un seuil trop important peut minimiser le nombre de règles. Par conséquent, la recherche des top-k itemsets fréquents fermés a été proposée, de longueur minimale min_l , où k est un nombre déterminé par l'expert permettant de fouiller k itemsets fréquents fermés. Top-k fait référence aux k motifs fermés les plus fréquents et min_l la longueur minimale des motifs fermés. Pour répondre à cette approche, plusieurs algorithmes ont été proposés, dont TFP (Top-k FrequentClosedItemsets) dans lequel toutes les transactions plus courtes que min_l sont exclues de la recherche et le support minimal est déterminé dynamiquement lors de la construction du FP-tree, permettant ainsi d'élaguer le treillis avant l'extraction [42].

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

L'exemple suivant illustre la relation entre les trois itemsets :

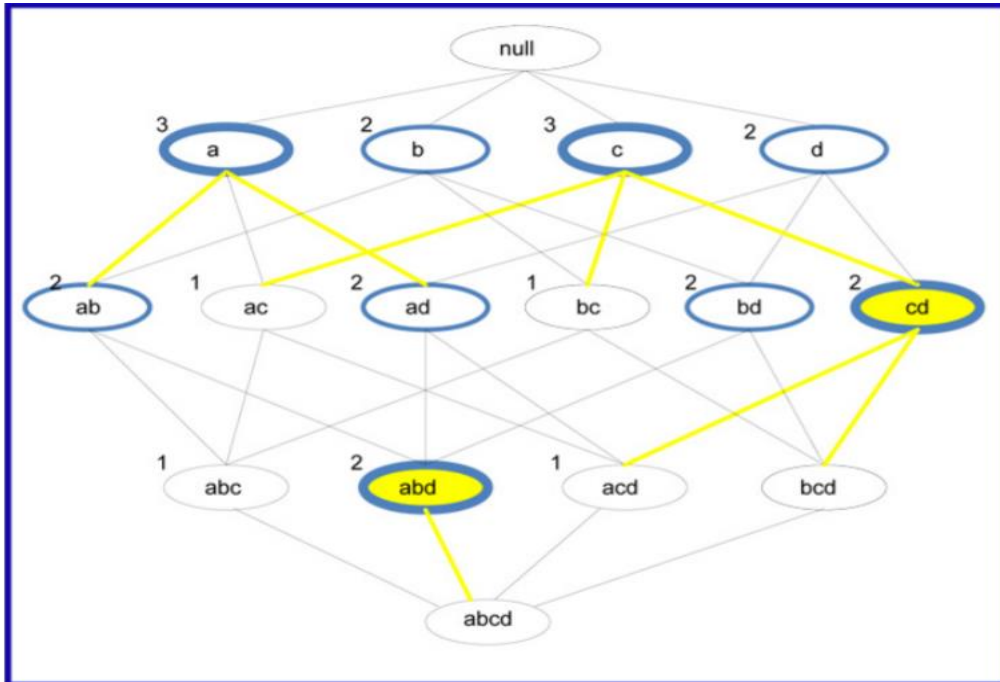
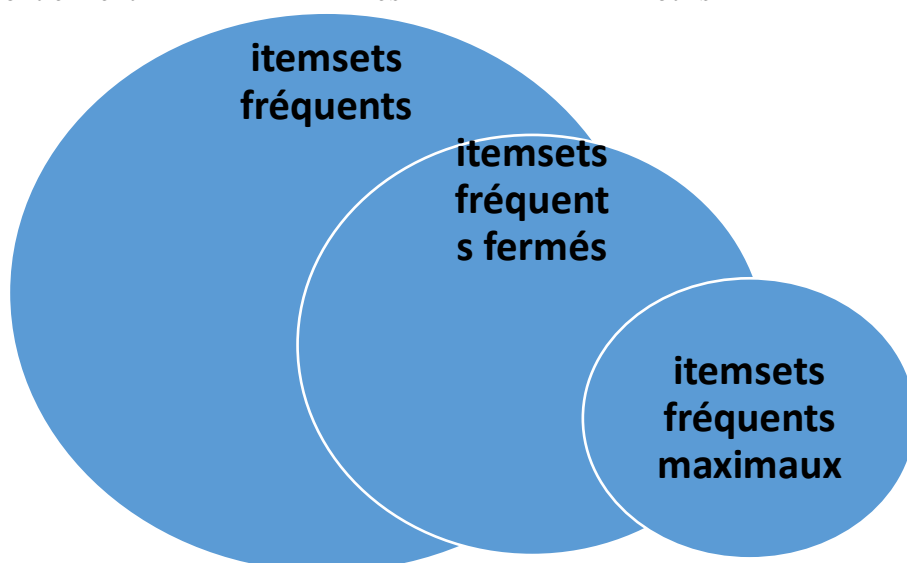


Figure 8 : Treillis des itemsets [40]

- Les motifs encerclés par les lignes minces ne sont pas fréquents, les autres le sont.
- Les motifs encerclés par des lignes plus épais sont fermés.
- Les motifs colorés sont maximaux.

Il est clair que les motifs fréquents contiennent les motifs fermés et les motifs fermés contiennent les motifs maximaux.



Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

4.2 L'algorithme :

L'extraction des itemsets fréquents est la première étape de base pour former les règles d'associations, et l'algorithme Apriori est considéré comme le premier algorithme d'extraction des itemsets fréquents pour former des règles d'associations.

4.2.1 Apriori

L'algorithme Apriori développé par [Agrawal1994] est une grande réussite dans l'histoire des règles d'associations [Cheung1996c]. C'est l'algorithme de règle d'association le plus connu. [33]

L'algorithme Apriori, est un algorithme clé pour les règles d'associations car il est à la base de la majorité des algorithmes servant à découvrir des règles d'associations plus complexes telles que les associations séquentielles et multidimensionnelles. Il tire son nom de son heuristique qui utilise l'information connue a priori sur la fréquence des items. Cette heuristique stipule que si A , un sous-ensemble d'items de l'ensemble I , ne possède pas le support minimal, il ne peut être engagé dans une règle d'association avec tout autre item ij de l'ensemble I , $ij \notin A$. Ainsi, si A est peu fréquent la règle $A \Rightarrow ij$ l'est également et il est donc inutile d'examiner toute règle d'association où A est impliqué.

4.2.1.1 Les propriétés de l'algorithme Apriori

L'Apriori génère les itemsets candidats en joignant Les itemsets fréquents de la passe précédente et supprimer les sous-ensembles qui ne sont pas fréquents lors de la passe précédente sans tenir compte des transactions dans la base de données. En ne considérant que les itemsets fréquents de la passe précédente, le nombre d'itemsets fréquents candidats est considérablement réduit. Lors du premier passage, les itemsets avec un seul item sont comptés. Les itemsets fréquents découverts du premier passage sont utilisés pour générer les ensembles candidats de la seconde passe en utilisant la fonction apriori gen (). Une fois que les itemsets candidats sont trouvés, leur les supports sont comptés pour découvrir les itemsets fréquents de taille deux en analysant la base de données. Au troisième passage, les itemsets fréquents de la deuxième passe sont considérés comme les ensembles candidats pour découvrir des itemsets fréquents de cette passe. Ce processus itératif se termine lorsqu'aucun

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

nouvel itemsets fréquents n'est trouvé. Chaque passage i de l'algorithme scanne une fois la base de données et détermine ensembles itemsets de taille i . L_i désigne itemsets fréquents de taille i , tandis que C_i est des candidats de taille i . [52]

Propriété 1 : Tous les sous-ensembles d'un ensemble fréquent sont fréquents. Cette propriété permet à l'algorithme Apriori de limiter la construction des ensembles "candidats", c'est à dire susceptibles d'être fréquents, de taille k à partir des ensembles fréquents de taille $k-1$.

Propriété 2 : Tous les sur-ensembles d'un ensemble non fréquent sont non fréquents. Cette propriété permet à l'algorithme Apriori d'éliminer un ensemble candidat si un seul de ses sous-ensembles est non fréquent. [51]

4.2.1.2 Principe d'algorithme apriori :

Les algorithmes d'extraction d'éléments répétitifs sont répétitifs, ce qui signifie que ces éléments sont examinés en fonction du niveau dans lequel ils se trouvent :

C'est-à-dire que dans l'itération M , nous recherchons des itemsets fréquents de longueur M .

En $M + 1$, nous recherchons des itemsets fréquents de longueur $M + 1$.

Il existe deux algorithmes consécutifs et complémentaires [43] :

1- Algorithm Apriori - Generation

Apriori-gen (F).

Entrée : F : ensembles de mots fréquents de cardinal k

⇒ **Début**

- 1- $C \leftarrow \{c = f_1 \cup f_2 \text{ telsque } (f_1, f_2) \in F \times F, \text{card}(c) = k + 1\}$
- 2- pour chaque $c \in C$ faire
- 3- pour chaque $s \subset c, \text{card}(s) = k$ faire
- 4- si $s \notin F$ alors
- 5- $C \leftarrow C \setminus \{c\}$ 7 Retourner C

⇒ **Fin**

2- Algorithm Apriori

Apriori (B, minsup).

Entrée : B : Base de données de transaction D , seuil de support minimum minsup : entier.

Sortie : Ensemble des itemsets fréquents

⇒ **Début**

- 1- $C_1 \leftarrow \{\text{ensemble des motifs de taille } 1\}$
- 2- $k \leftarrow 1$

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

- 3- tantque $C_k = \emptyset$ faire
 - 4- pour chaque $c \in C_k$ faire
 - 5- pour chaque $t \in B$ faire
 - 6- si $c \subset t$ alors
 - 7- $\text{support}(c) \leftarrow \text{support}(c) + 1$
 - 8- $F_k \leftarrow \{c \in C_k \mid \text{support}(c) > \text{minsup}\}$
 - 9- $k \leftarrow k + 1$
 - 10- $C_k \leftarrow \text{Apriori-gen}(F_{k-1})$
 - 11- Retourner Ensemble des itemsets fréquents
- ⇒ Fin

4.2.1.3 Inconvénient de l'algorithme Apriori :

Il y a un inconvénient majeur à l'algorithme Apriori, qui est le point le plus important pour toutes les entreprises et institutions du plus petit au plus grand et c'est le temps. Il faut un grand nombre d'opérations pour accéder à la base de données, et la base formelle contient beaucoup de zéros et donc à chaque fois on les parcourt, d'où une perte de temps.

De plus, pour révéler le motif récurrent en taille 100 (par exemple) $v_1, v_2 \dots v_{100}$, 2100 ensembles d'éléments filtrants doivent être générés, ce qui entraîne le coût et la perte de temps de la génération candidate. Par conséquent, il vérifiera plusieurs groupes de groupes candidats et analysera la base de données plusieurs fois à plusieurs reprises pour trouver des groupes candidats. Apriori sera très faible et inefficace lorsque la capacité de la mémoire est limitée avec un grand nombre de coefficients.

4.2.2 Améliorations proposées à Apriori :

- **DHP (Direct Hashing and Pruning)** proposé par Park et al. DHP est une légère modification d'Apriori qui utilise des tables de hachage pour diminuer le nombre d'itemsets candidats générés. Cet algorithme propose également une diminution de la taille de la base de données, au fur et à mesure des itérations. Lors des deux premières itérations (gestion des 1-itemsets et des 2-itemsets), on utilise une table de hachage sur les numéros des itemsets. A chacune des cases de la table de hachage, on associe un compteur. A chaque fois qu'une instance contient l'un des itemsets associés (par la fonction de hachage) à une case de la table de hachage, on incrémente le compteur de cette case. On garde les itemsets associés aux cases dont le compteur est suffisant (pour une case associée à 3 itemsets, on garde les itemsets si le compteur associé à la

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

case possède une valeur supérieure à trois fois le seuil de support fixé). A partir de la deuxième itération, on effectue un pruning des instances de la base. Pour chaque instance, si le nombre de bits de cette instance est supérieur à k , l'instance peut être utile à la k ème itération, sinon, on la supprime [53].

- **Apriori TID :**

L'amélioration qu'apporte cet algorithme par rapport au précédent est le fait de stocker à chaque itération les identificateurs des transactions contenant les sous-ensembles fréquents dans l'ensemble C_k . La propriété des sous-ensembles fréquents nous permet de voir clairement que les transactions sollicitées à l'itération $k + 1$ seront parmi celles identifiées à l'itération k . Par conséquent, l'accès à D est effectué seulement pour l'itération 1 [44].

- **AprioriHybrid** proposé par Agrawal et Srikant.

Il constitue un mélange d'Apriori et AprioriTID en fonction de l'itération à laquelle on se trouve. [51]

Cet algorithme est basé sur l'idée qu'il n'est pas nécessaire d'utiliser le même algorithme dans tous les passages de données. Comme mentionné dans [Agrawal1994], Apriori a de meilleures performances dans les passages précédents, et Apriori TID surclasse Apriori plus tard passe. Sur la base des observations expérimentales, la technique Apriori-Hybrid a été développée utilise Apriori dans le passe initial et bascule sur Apriori-TID lorsqu'il s'attend à ce que l'ensemble C_k à la fin de la passe tienne en mémoire. Par conséquent, une estimation de C_k à la fin de chaque passe est nécessaire. En outre, le passage d'Apriori à Apriori-TID. La performance de cette technique a également été évaluée en menant des expériences pour de grands ensembles de données. C'était observé qu'Apriori-Hybrid fonctionne mieux qu'Apriori sauf dans le cas où le changement se produit à la toute fin de les passes. Apriori-Hybrid est utilisé pour la classification du cancer. [52]

- **MaxEclat / MaxClique** proposé par Zaki et al. A chaque itemset on associe une liste Des identifiants des instances qui contiennent cet itemset. Ces algorithmes effectuent ensuite des jointures sur les tidlistes des instances associés à chaque itemset. Le nombre d'instances contenus dans une tid-liste est le support de l'itemset associé à cette tid-liste. [54].

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

Et nous résumerons la différence entre max Clique et max Eclat Dans le tableau suivant [66] :

MaxEclat	MaxClique
Eclat utilise des classes basées sur des préfixes et des recherches ascendantes. MaxEclat utilise classes basées sur des préfixes et une recherche hybride. Dans une classe basé préfixe on traite d'abord la racine puis on traite, de manière récursive et dans l'ordre, les sous-arbres pointés de racine.	Clique utilise des classes basées sur la clique et une recherche ascendante. MaxClique utilise des classes basées sur la clique et recherche hybride. Une clique, C , dans un graphe non orienté $G = (V, E)$ est un sous-ensemble des sommets, $C \subseteq V$, tel que tous les deux sommets distincts sont adjacents.

- **Algorithme de comptage dynamique** 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 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 [55].

- **Apriori Partition** proposé par Savasere et al. Cet algorithme ne nécessite que deux Balayages du jeu de données. Durant le premier balayage, la base est divisée en n partitions disjointes qui seront considérées une à une successivement. Pour chaque partition, l'ensemble des itemsets fréquents dans la partition (les itemsets fréquents locaux) sont extraits. Les ensembles d'itemsets fréquents de chaque partition sont ensuite fusionnés pour obtenir un

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

ensemble d'itemsets candidats à être fréquents sur la base entière. On a ainsi un surensemble de l'ensemble des itemsets fréquents. Durant le second balayage les supports pour ces candidats sont calculés sur toute la base, les itemsets fréquents sont identifiés. [56].

Dans notre travail, nous avons utilisé l'algorithme Apriori TID.

4.2.3 Principe de l'Algorithme Apriori TID

Il utilise une fonction apriori-gen pour déterminer les itemsets candidats avant le début de la passe. La caractéristique intéressante est que la base de données D n'est pas utilisée pour compter le support après le premier passage. Au contraire, l'ensemble $|C_k|$ est utilisé à cette fin. Chaque set $|C_k|$ est de la forme $\langle TID; \{X_k\} \rangle$, où chaque X_k est un k -itemset potentiellement fréquent présent dans la transaction avec identifiant TID. Pour $k = 1$, C_1 correspond à la base de données D , bien que conceptuellement chaque item i soit remplacé par itemsets $\{i\}$. Pour $k > 1$, $|C_k|$ est généré par l'algorithme. Le membre de C_k correspondant à la transaction t est $\langle t.TID, \{c \hat{C}_k \mid c \text{ contenu dans } t\} \rangle$.

Si une transaction ne contient aucun k -itemset candidat, alors $|C_k|$ n'aura pas d'entrée pour cette transaction. Donc, le nombre d'entrées dans $|C_k|$ peut être inférieur au nombre de transactions dans la base de données, en particulier pour les grandes valeurs de k . De plus, pour de grandes valeurs de k , chaque entrée peut être plus petite que la transaction correspondante peu de candidats peuvent être contenus dans la transaction. Cependant, pour les petites valeurs de k , chaque entrée peut être plus grande que la valeur correspondante transaction car une entrée dans C_k inclut tous les k -itemsets candidats contenus dans la transaction [52].

Algorithm Apriori TID:

→ Début

- 1) $L_1 = \{1\text{-itemsets frequent}\}$;
- 2) $C_1 = \text{base de données } D$;
- 3) Pour ($k = 2$; $L_{k-1} \neq \emptyset$; $k++$) fair
- 4) $C_k = \text{apriori-gen}(L_{k-1})$;
- 5) $|C_k| = \emptyset$;
- 6) pour entries $t \in |C_{k-1}|$ fair

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

7) // déterminer les itemsets candidats dans Ck Contenu dans la transaction avec identifiant t.TID

$C_t = \{c \in C_k \mid (c - c[k]) \in \text{set-of-itemsets} \wedge (c - c[k-1]) \in \text{set-de-itemsets}\}$;

8) pour candidates $c \in C_t$ faire

9) $c.\text{count}++$;

10) si $(C_t \neq \emptyset)$ then $|C_k| += \langle t.\text{TID}, C_t \rangle$;

11) Fin

12) $L_k = \{c \in C_k \mid c.\text{count} \geq \text{minsup}\}$

13) Fin

14) réponse = $\cup_k L_k$;

→ Fin

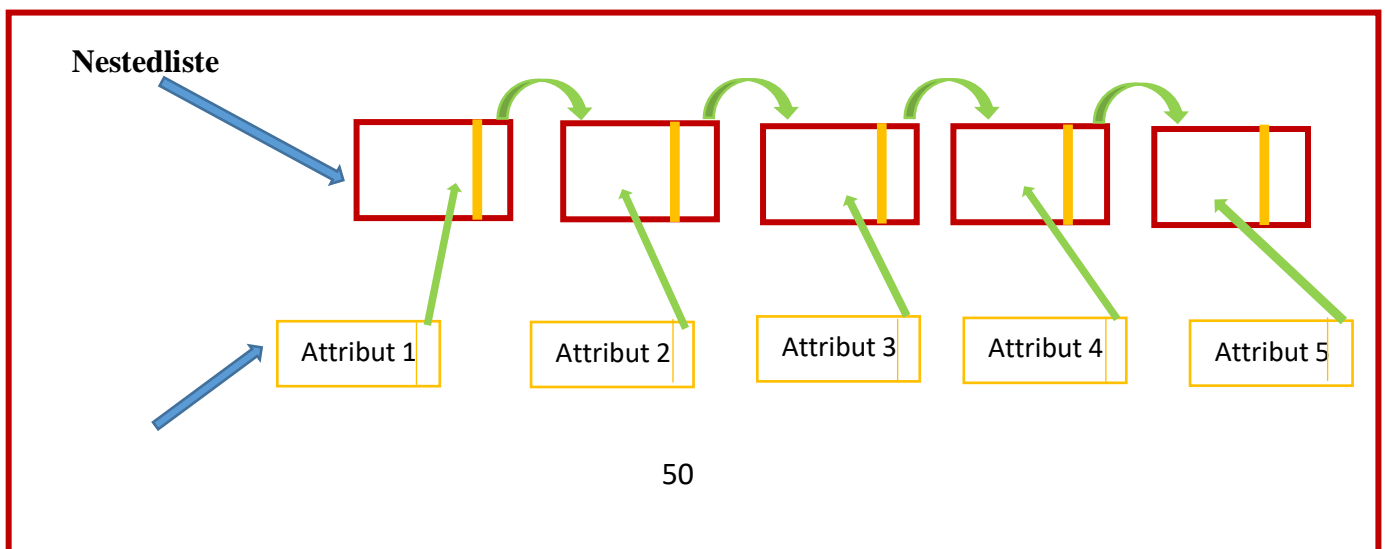
Cet exemple expliquera le travail d'Apriori TID simplement :

	Attribut 1	Attribut2	Attribut3	Attribut4	Attribut5
	Farine	Lait	Chocolat	Huile	Sucre
Transaction1	1	0	1	1	0
Transaction2	0	1	1	0	1
Transaction3	1	1	1	0	1
Transaction4	0	1	0	0	1
Transaction5	1	1	1	0	1
Transaction6	0	1	1	0	1

- ÉTAPE 1 :

- Créer Nested liste contient le premier ensemble des items

Nested liste : Une liste que chaque case contient l'ID de la liste d'attribut de la base de données utilisée



Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

Liste d'attribut 1

(TID de premier attribut)

- Créer la première liste de TID

Itemsets C(1) + liste de TID :

Farine	[1, 3, 5]	
Lait	[2, 3, 4, 5, 6]	
Chocolat	[1, 2, 3, 5, 6]	L(1)= 4
Huile	[1]	
Sucre	[2, 3, 4, 5, 6]	

- ÉTAPE 2 :

- Obtenez les articles fréquents L (i) de C (i):

Ensembles itemsets de 1_ itemsets :

Farine	3
Lait	5
Chocolat	5
Sucre	5

- ÉTAPE 3:

- Faire toutes les combinaisons possibles à partir de L (i-1)

- Faire des intersections entre les listes de TID

Itemsets de C(2) + liste de TID

Farine Lait	[3, 5]	
Farine Chocolat	[1, 3, 5]	
Farine Sucre	[3, 5]	L (2) = 6
Lait Chocolat	[2, 3, 5, 6]	
Lait Sucre	[2, 3, 4, 5, 6]	
Chocolat Sucre	[2, 3, 5, 6]	

- ÉTAPE 4:

- Répéter ÉTAPE 2

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

- Répéter ÉTAPE 03
- Arrêter lorsque $(L) > 0$

Ensembles itemsets de 2_ itemsets :

FarineLait	2
FarineChocolat	3
Farine Sucre	2
LaitChocolat	4
Lait Sucre	5
Chocolat Sucre	4

Itemsets de C(3) + liste de TID

Farine Lait Chocolat	[3, 5]	
Farine Lait Sucre	[3, 5]	L(3)= 4
Farine Chocolat Sucre	[3, 5]	
Lait Chocolat Sucre	[2, 3, 5, 6]	

Ensembles itemsets de 3_ itemsets :

Farine Lait Chocolat	2
Farine Lait Sucre	2
Farine Chocolat Sucre	2
Lait Chocolat Sucre	4

Itemsets de C(4) + liste de TID

Farine Lait Chocolat Sucre	[3, 5]	L(4)= 1
----------------------------	--------	---------

Ensembles itemsets de 4_ itemsets :

Farine Lait Chocolat Sucre	2
----------------------------	---

Nous avons donc tous les itemsets fréquents et nous pouvons extraire toutes les règles D'association solide.

8. Conclusion

Les règles d'association depuis leur création sont considérées comme l'un des outils les plus efficaces et les plus puissants pour identifier et distinguer les relations entre les différentes données, indépendamment de la facilité et de la simplicité de leur lecture et de leur interprétation afin qu'elles puissent aider les analystes à découvrir des relations et des connexions.

Chapitre 3 : Les règles d'associations pour l'analyse de panier d'achat

Elles permettent de bénéficier d'une grande quantité de connaissances cachées Ce qui n'est pas facile de découvrir d'une autre manière qui peut vraiment être utilisée pour améliorer l'efficacité de l'entreprise et ainsi augmenter ses revenus sans toucher à certains problèmes liés à la taille des méga données.

L'algorithme Apriori TID est utilisé en particulier sur le problème de l'analyse du panier d'achats. Dans ce cas, chaque personne est considérée comme un panier et chaque article est un produit, l'entreprise est toujours intéressé à connaître les produits qui sont souvent achetés ensemble et par conséquent les préférences des clients.

Dans ce travail, nous avons choisi l'algorithme Apriori TID pour extraire les règles d'associations sur une base de données qui regroupe des paniers d'achats afin d'extraire des connaissances qui contribuent grandement à améliorer ses opérations en quelque sorte il s'agit d'une prédiction du comportement du client.



Chapitre 4:

Conception

Chapitre 4Conception:

1. Introduction :

Dans ce chapitre, nous présentons la vision conceptuelle de notre système qui tente de trouver des voies et des connexions entre les produits dans le but de les utiliser de manière optimale pour comprendre le comportement des consommateurs.

Où nous expliquerons les étapes et les unités qui composent notre système. Et soulignons les principales étapes en expliquant les différents éléments du système.

L'objectif de notre travail est de concevoir un système qui soit capable d'accomplir plusieurs objectifs et de les considérer, tel que :

- la possibilité de créer une nouvelle base de données de produits ou de commodités
- dans laquelle l'utilisateur est capable de traiter une base de données pré-préparée en organisant son étude et en apprenant dessus
- enfin en extrayant ses propres règles et donc la capacité de prédire et d'essayer de connaître le comportement des consommateurs.

Ce chapitre est consacré pour clarifier notre ordre d'extraction de règles d'association et pour mieux comprendre la méthodologie du système et son fonctionnement.

2. Description de notre travaille

Notre système est un système qui s'inscrit dans le cadre des CRM, il se base sur l'analyse des achats trouvés dans le panier de différents clients, donc il dépend des données enregistrées des biens achetés pendant une certaine période de temps, dans le but de trouver des liens et des relations entre eux pour aider à comprendre le mode de consommation ou trouver un schéma qui montre les biens liés afin que le vendeur puisse plus tard utiliser ces résultats Pour prendre une décision, qu'il s'agisse de réorganiser ses étagères, de proposer différentes offres promotionnelles, ou d'augmenter un bien ou au contraire le diminuer.

Dans ce but nous avons choisi la technique d'extraction de règle d'associations qui se matérialise par l'algorithme d'optimisation avancé Apriori TID, il s'agit de rechercher et d'extraire toutes les règles d'associations et de déterminer celles qui sont fortes pour les données entrées qui sont regroupées dans une base de données transactionnelle sous forme d'informations d'achat pour les clients d'un supermarché.

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

3. Conception globale du système

Notre système est un système d'analyse de paniers d'achats qui va nous aider à mieux répondre aux besoins de la clientèle il s'agit d'extraire les règles de liens d'une énorme base de données en découvrant les motifs fréquents à l'aide de l'algorithme Apriori TID qui vont être interpréter par des règles de liens.

En général, notre application comprend les principales étapes du processus de prédiction, représentées par trois étapes de base pour le fonctionnement du système, et chaque étape contient un ensemble des modules.

- La première étape est prétraitement chargée de convertir la base de données transactionnelle en base formel.

Les entrées de cette étape sont des données transactionnelles, tandis que les sorties sont des données structurées

- La deuxième étape est responsable de l'extraction des itemsets Ensuite, les itemsets fréquents leurs prises en charge et stockage de chaque génération.
- Quant à la troisième étape extraction des règles, elle est chargée de transformer les itemsets fréquentes extraites de la deuxième étape en un ensemble de règles d'associations.

Nous pouvons décrire la méthodologie d'extraction de règles d'association ainsi que son utilisation dans ce qui suit.

La structure de notre système est représentée dans le schéma suivant :

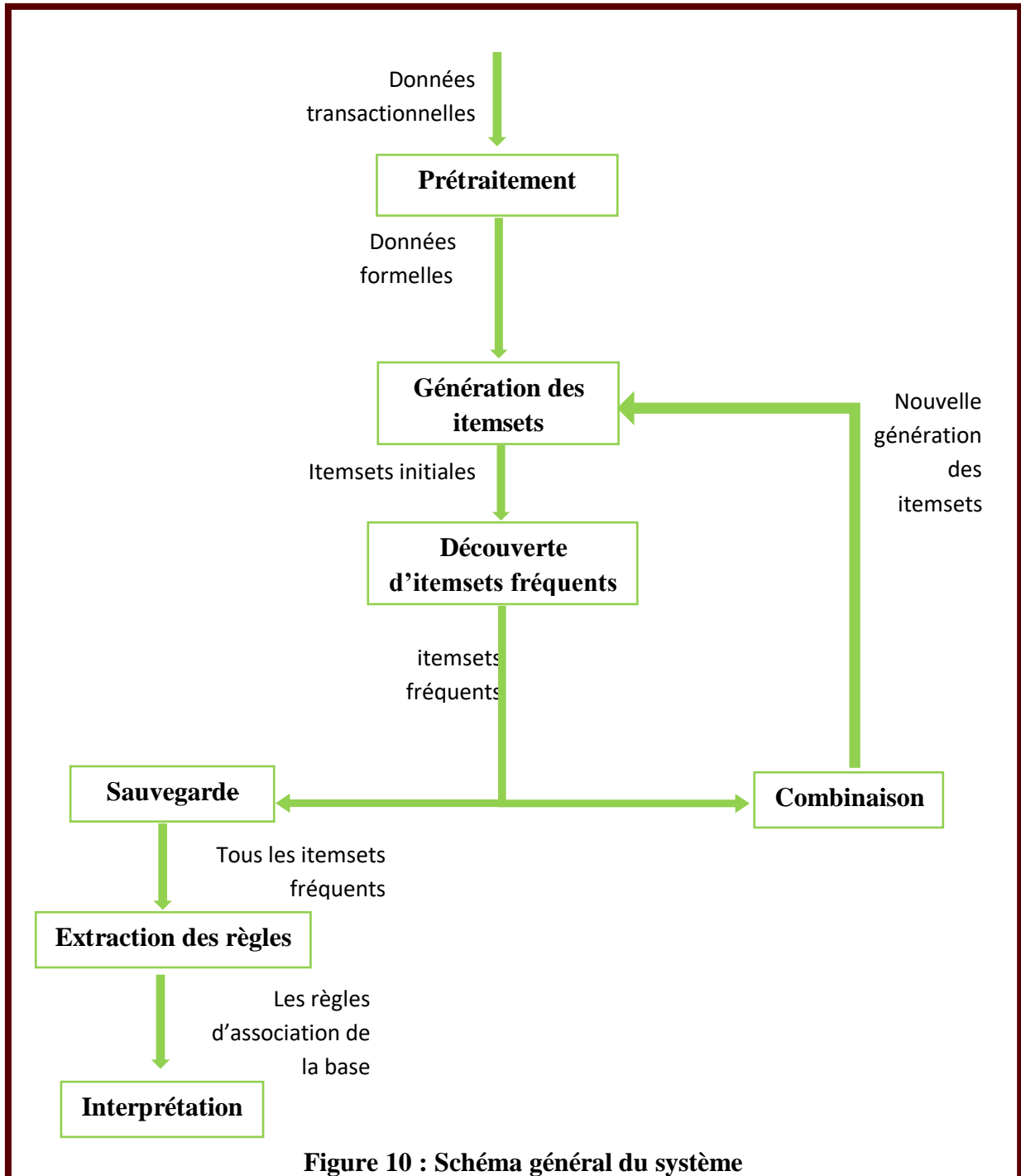


Figure 10 : Schéma général du système

Dans la conception détaillée, nous expliquerons chaque point mentionné ci-dessus dans le.

4. Conception détaillée :

Comme nous l'avons expliqué dans le schéma précédent, nous pouvons diviser le système en 3 modules :

4.1 L'étape 1

C'est l'étape responsable du prétraitement elle contient un seul module qui a pour rôle la préparation des données afin qu'elles soient exploitables par l'algorithme a priori TID

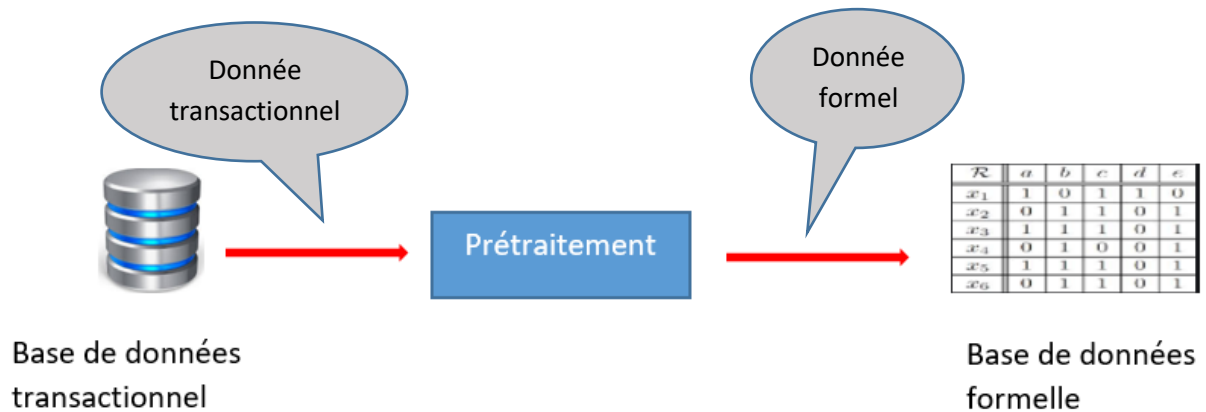


Figure 11 : le prétraitement

4.1.1 Prétraitement :

L'étape prétraitement consiste à changer l'état des données saisies d'un état à un autre, afin que l'application puisse les lire et en extraire ce qui est nécessaire.

La base de données dans le premier cas est une base de données transactionnelle, de sorte que ses données sont des lignes, et chaque ligne contient un ensemble d'éléments séparant chaque élément et un élément par une virgule, et après traitement, elle devient une base de données organisée sous la forme d'une grande table qui contient un zéro ou un selon la présence ou l'absence de l'élément dans chaque ligne de la base initiale.

4.1.2 Les données transactionnelles :

Dans le contexte de la gestion de données, les données transactionnelles sont les informations enregistrées à partir des transactions.

Chapitre 4 Conception:

Une transaction, dans ce cas, est une séquence d'échange d'informations et de travail connexe qui est traitée comme une seule unité dans le but de répondre à une demande. Les données transactionnelles peuvent être financières, logistiques ou professionnelles et tout englober, du bon de commande aux coûts et indemnités d'assurance en passant par le statut d'une livraison et les heures de travail d'un employé.

En tant que composants d'enregistrements transactionnels, les données transactionnelles sont groupées avec les données de référence qui y sont associées. Les données transactionnelles enregistrent une heure et des données de référence utiles et nécessaires pour un enregistrement transactionnel donné [65].

Comme nous l'avons dit précédemment, notre base de données utilisée regroupe les données qui concernent les achats effectués par les clients, car elle contient un grand ensemble de lignes et chaque ligne représente une commande pour un client ou un panier. Chaque panier contient un groupe de marchandises. Lorsque le panier se termine, la ligne se termine.

Panier1 :Prod1, Prod4
Panier2 :Prod1, Prod2, Prod3, Prod4
Panier3 :Prod1, Prod2, Prod4
Panier 4: Prod3,Prod4

4.1.3 Les données formelles :

La version de base de l'extraction de motifs fréquents permet de faire la fouille dans une table d'une base de données relationnelle dont les valeurs sont des booléens indiquant la présence ou l'absence d'une propriété.

Une telle base est appelée base de données formelle. Une base de données formelle est définie par un triplé (O, P, R) où :

- O est un ensemble fini d'objets.
- P est un ensemble fini de propriétés.

Chapitre 4 Conception:

– R est une relation sur $O \rightarrow P$ qui permet d'indiquer si un objet x a une propriété p (noté xRp) ou non. [40]

Par exemple dans le cas d'analyse du panier dans un supermarché, O est l'ensemble des transactions d'achat, P est l'ensemble d'articles et R est la relation indiquant si un article Prod est acheté Il marque dans la case de transaction 1 et en son absence, il marque 0.

R	Prod1	Prod2	Prod3	Prod4
Panier1	1	0	0	1
Panier2	1	1	1	1
Panier 3	1	0	1	0
Panier4	0	0	1	1

4.1.4 Fonctionnement :

La base de données transactionnelle est parcourue une fois afin de collecter tous les produits existants dans tous les paniers d'achat de cette base et placés en première ligne dans la nouvelle base de données, Le reste des lignes restent les mêmes que le nombre des paniers existants.

Et enfin les cases sont remplies en fonction de la présence ou de l'absence du produit, c'est-à-dire que la case de l'intersection du panier avec le produit détermine la présence ou l'absence de ce produit dans ce panier.

4.2 L'étape 2

C'est l'étape responsable de l'extraction des motifs fréquents, les modules qui assurent le fonctionnement de cette étape sont :

4.2.1 Génération des itemsets :

Dans ce module, tous les itemsets présents 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

Chapitre 4 Conception:

Contiennent les numéros ou les ID des transactions qui contiennent ceitemset comme expliqué dans l'exemple suivant :

L'exemple ci-dessous montre un mécanisme pour passer de La base formelle aux motifs et comment les extraire :

I. On a la base formelle

R	Prod1	Prod2	Prod3	Prod4
Panier1	1	0	0	1
Panier2	1	1	1	1
Panier 3	1	0	1	0
Panier4	0	0	1	1

Création des listes par colonne (chaque produit)

Items de longueur 1 et liste TID :

Prod1 = [1, 2, 3]

Prod2 = [2]

Prod3 = [1, 3, 4]

Prod4 = [1, 2, 4]

4.2.2 Découverte d'itemsets fréquents:

Après avoir extrait les itemsets, nous nous débarrassons de tous les itemsets non fréquents, à partir de support donné au début, ne laissant que les itemsets fréquents pour cette génération

Dans cet exemple, nous effaçons des listes inférieure ou égale à 1 comme dans l'exemple le cas du prod2

4.2.3 Sauvegarde:

Nous conservons tous les itemsets fréquents pour chaque génération, donc dans cet exemple, nous sauvegardons Prod1, Prod3, Prod4.

4.2.4 Combinaison :

Après avoir supprimé les listes dont la fréquence est inférieure ou égale à un comme Prod2, l'intersection entre les listes qui ont les mêmes numéros de client c'est-à-dire le ID d'un client commun Sur cette base, une intersection se produit entre la liste courante et le reste des listes de la même génération dans le cas d'un ou plusieurs clients communs.

Chapitre 4 Conception:

Le nom de la nouvelle listes (nouvel itemsets) résultant de l'intersection de deux séries devient la fusion des noms des deux listes précédentes par exemple Prod1 et Prod4 Deviens en Prod1Prod4.

On a itemsets de longueur 2 et liste TID :

Prod1Prod3= [2, 3]

Prod1Prod4= [2]

Prod3Prod4 = [2, 4]

Après avoir obtenu les itemsets de la deuxième génération, nous répétons les mêmes étapes précédentes, c'est-à-dire que nous nous débarrassons des itemsets non fréquents et ne conservons que les itemsets fréquents.

Nous sauvegardons Prod1Prod3, Prod3Prod4, puis formons les itemsets de la troisième génération à travers la combinaison.

itemsets de longueur 3 et liste TID :

Prod1Prod3Prod4= [2]

Étant donné que l'itemset de cette génération n'est pas fréquente, nous nous arrêtons ici et considérons cela comme un point d'arrêt.

Enfin, nous sortons l'étape avec les itemsets fréquents suivantes Prod1, Prod3, Prod4, Prod1Prod3, Prod3Prod4, dont nous extrayons les règles d'association.

4.3 L'étape 3

Les modules qui assurent le fonctionnement de cette étape sont :

4.3.1 Extraction des règles :

Consiste à générer Les règles d'association à partir des itemsets fréquents :

Une fois l'algorithme Apriori TID terminée, nous avons la base de données des itemsets fréquents, et à partir d'eux, nous pouvons extraire les règles d'association comme indiqué ci-dessous :

Après l'obtention des itemsets fréquents, le processus d'extraction des règles d'association est simple et dépend de l'extraction de toutes les règles entre les itemsets fréquents.

Chapitre 4 Conception:

C'est ce que nous montrerons dans la figure suivant :

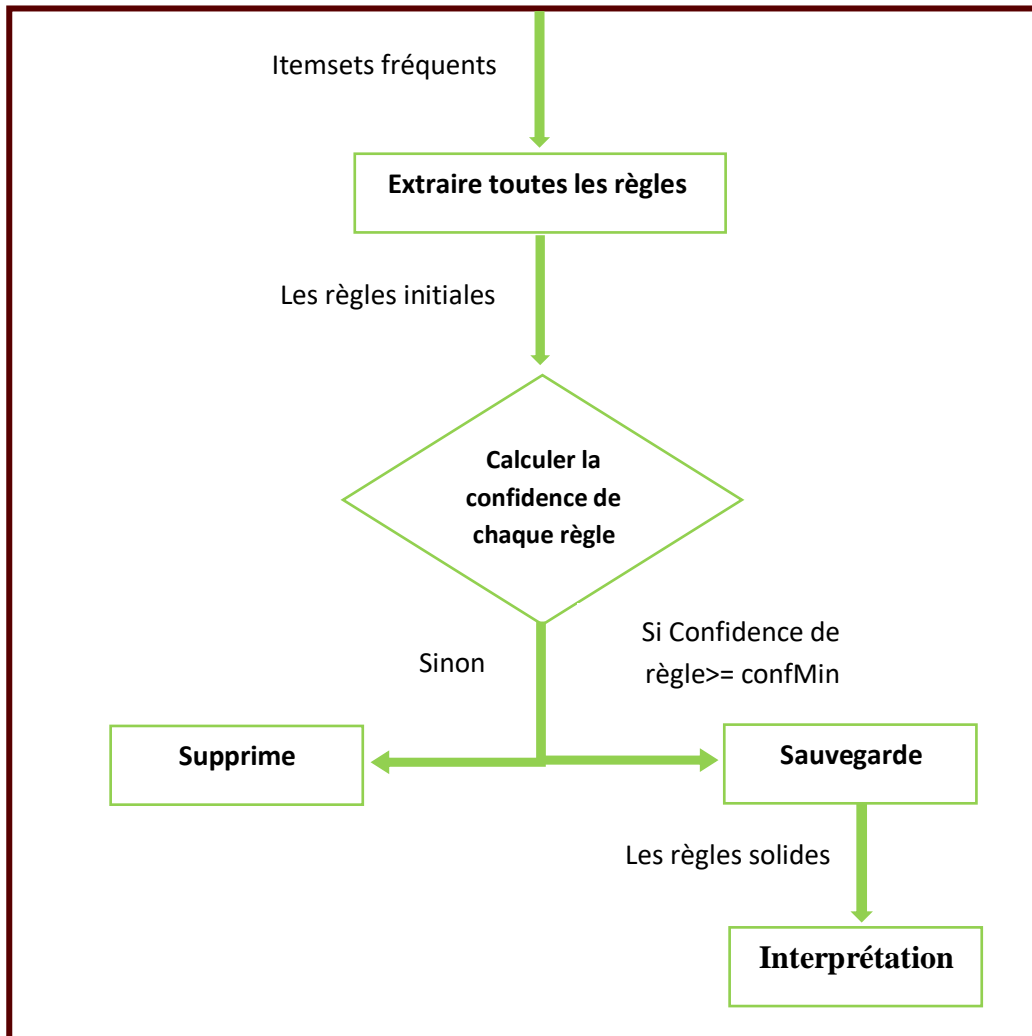


Figure 12 : processus d'Extraction des règles

En appliquant à l'exemple précédent, nous extrayons les règles d'association des itemsets fréquents comme suit :

Le confMin= 50%

L'itemset fréquent Prod1Prod3 :

$$\text{Prod1} \rightarrow \text{Prod3} \quad \text{confiance} = \frac{\text{support de Prod1Prod3}}{\text{support de Prod1}} = \frac{2}{3} = 66\%$$

$$\text{Prod3} \rightarrow \text{Prod1} \quad \text{confiance} = \frac{\text{support de Prod1Prod3}}{\text{support de Prod3}} = \frac{2}{3} = 66\%$$

L'itemset fréquent Prod1Prod3:

Chapitre 4 Conception:

$$\text{Prod3} \rightarrow \text{Prod4} \quad \text{confidence} = \frac{\text{support de Prod3Prod4}}{\text{support de Prod3}} = \frac{2}{3} = 66\%$$

$$\text{Prod3} \rightarrow \text{Prod4} \quad \text{confidence} = \frac{\text{support de Prod3Prod4}}{\text{support de Prod4}} = \frac{2}{3} = 66\%$$

Puisque toutes les règles extraites ont une plus grande confiance que la confMin, toutes les règles extraites sont acceptables.

4.3.2 Interprétation :

Nous pouvons utiliser notre modèle de prise de décision pour découvrir les liens cachés entre les produits et savoir comment se déroulent les achats et nous aider à comprendre le comportement des consommateurs.

5. Conclusion

Dans ce chapitre, nous avons présenté la conception de notre système, où nous avons détaillé les différentes étapes de cette conception, préparer les données et les étudier et extraire les parties prises en charge, en extraire les règles d'associations et comment les appliquer réellement au profit du vendeur pour augmenter les taux de profit.

Dans le chapitre suivant, nous décrirons l'implémentation de notre système et les résultats obtenus et leurs discussions.

Chapitre 5:

Implémentation

1. Introduction

Dans le chapitre précédent, nous avons présenté notre méthode proposée Pour trier les grandes bases de données pour trouver des associations entre les données et Les routes qui nous permettent de tirer profit de cette base, ces routes sont appelées les règles d'association.

Nous avons utilisé l'algorithme apriori TID pour extraire les motifs fréquents suivi de la déduction des règles d'association. Nous avons expliqué comment notre application peut extraire ces règles du début à la fin pour nous aider à prendre une décision.

Dans ce chapitre, nous allons présenter l'environnement de travail, le langage de programmation et les outils que nous avons utilisés pour construire ce système. Par la suite, nous allons expliquer son mécanisme d'action et les résultats obtenus.

2. Outils et langages de développement

Lors de la phase de mise en œuvre, nous avons eu l'occasion d'apprendre davantage sur différents outils et techniques de développement logiciel qui sont décrits ci-dessous :

2.1 Environnement et outils de programmation

Le choix du bon environnement de programmation est très important pour le développement des projets. Afin de réaliser notre système et son interface, nous avons eu recours aux outils suivants :

2.1.1 Netbeans :



Figure14. Interface Netbeans[57]

NetBeans est un environnement de développement intégré (IDE) open source pour le développement avec Java, PHP, C ++ et d'autres langages de programmation. NetBeans est également appelé plate-forme de composants modulaires utilisés pour développer des

applications de bureau Java[57], L'EDI est conçu pour limiter les erreurs de codage et faciliter la correction des erreurs avec des outils tels que NetBeansFindBugs pour localiser et résoudre les problèmes de codage Java courants et Debugger pour gérer le code complexe avec des surveillances de terrain, des points d'arrêt et la surveillance de l'exécution ; il prend également en charge C / C ++, PHP , Groovy et HTML5 en plus de Java , JavaScript et JavaFX[58].

2.1.2 Scence builder



Figure 15 : interface de SceneBuilder[59]

SceneBuilder est un outil de mise en page visuelle qui permet aux utilisateurs de concevoir rapidement des interfaces utilisateur d'application JavaFX, sans codage.

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 avec un projet Java en liant l'interface utilisateur à la logique de l'application [59].

2.2 Langages de programmation

2.2.1 Java



Figure 16 : interface de Java[60]

C'est un langage de programmation orienté objet, développé par Sun Microsystems. Il permet de créer des logiciels compatibles avec de nombreux systèmes d'exploitation (Windows, Linux, Macintosh, Solaris). Java donne aussi la possibilité de développer des programmes pour téléphones portables et assistant personnels [60].

Java est un langage de programmation inspiré du langage C++, avec un modèle de programmation orienté objet. Il permet de créer des applications complètes. Il peut également servir à créer un petit module d'application, dit applet, à intégrer dans une page Web [61].

2.2.2 Javafx



Figure 17 : interface de JavaFx[62]

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 [62].

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.

2.2.3 Différents entre Java swing et Java fx et pour quoi le Java fx

	Java Swing	Java FX
Composants	Swing a un certain nombre de composants	Moins de composants par rapport aux API Swing héritées
Interface utilisateur	Moins de composants par rapport aux API Swing héritées	Des composants GUI riches peuvent être créés avec une apparence et une sensation avancées
Développement	Les API Swing sont utilisées pour écrire des composants d'interface utilisateur	Scripts JavaFX et développement rapide de l'interface utilisateur avec le générateur d'écran
Fonctionnalité	Aucune nouvelle présentation de fonctionnalités pour l'avenir	JavaFX dispose d'une nouvelle boîte à outils riche, qui devrait se développer à l'avenir
Catégorie	Bibliothèque d'interface utilisateur héritée complète	À venir pour proposer des composants d'interface utilisateur riches
Prise en charge MVC	La prise en charge MVC entre les composants manque de cohérence	Convivial avec le modèle MVC

3. Description des différentes classes

L'ensemble des classes sont présentées dans le tableau suivant :

Classe	Signification
Application.fxml	code XML contenant l'interface de contrôle et les formalités de saisie des informations (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)
ApplicationController.java	<p>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 fréquents et enfin de l'extraction des règles d'association.</p> <p>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.</p> <p>La deuxième classe est appelée règle, 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.</p> <p>Ce code est principalement basé sur une liste imbriquée (nestedliste), c'est-à-dire une liste qui contient un groupe de listes.</p>
ApprioriTidApplication.java	Main de la largeur de l'interface d'entrée les informations initial.
FXMLDocument.fxml	code XML qui contient les formalités de la première interface.
FXMLDocumentController.java	Le code permettant de relier l'interface de saisie au code principal (Controller), afin de respecter les données saisies au début.

Main.css	Spécial pour l'esthétique de nos interfaces d'application.
Transf.java	Le code responsable de la conversion de la base de données d'une base de données transactionnelle en une base de données formel.

4. Base de données utilisé

Nous avons utilisé dans ce mémoire la base grocerie [64] sous forme d'un fichier (groceries.csv) qui contient un nombre énorme de transactions (9835 transactions) afin que chaque transaction soit un panier d'achat contenant un groupe de marchandises et des produits (Provisions alimentaires).

La figure suivant énonce une partie de cette base de données :

	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,Jong life bakery product											
6	whole milk,butter,yogurt,rice,abrasive cleaner											
7	rollsbuns											
8	other vegetables,UHT-milk,rollsbuns,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,rollsbuns,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,rollsbuns,flour,sweet spreads,salty snack,waffles,candy,bathroom cleaner											
26	bottled water,canned beer											
27	yogurt											
28	sausage,rollsbuns,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,rollsbuns,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 18 : partie de base de données groceries

5. Présentation de l’application

On présente dans cette section les différentes interfaces de l’application.

5.1. Première interface :

L’interface principale

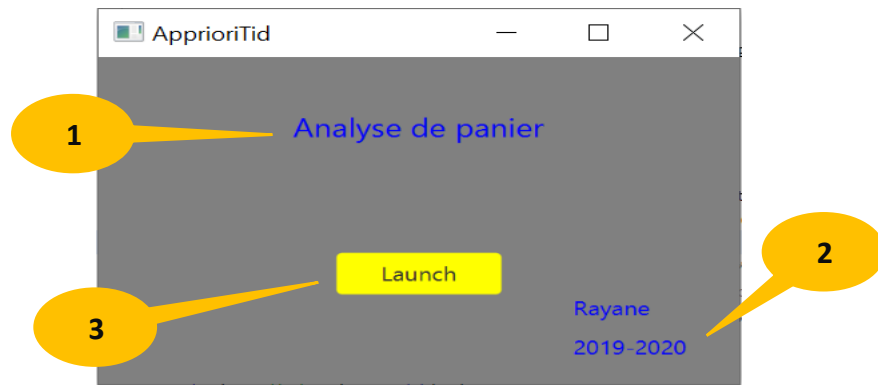


Figure19 : La forme principale

1. La première interface est une interface pratique qui contient un nom qui exprime le travail effectué par cette application (Analyse de panier), qui trie et analyse les paniers d'achat des clients.
2. L'année au cours de laquelle cette application a été complétée.
3. Un bouton pour démarrer cette application.

5.2 Deuxième interface :

Interface de traitement

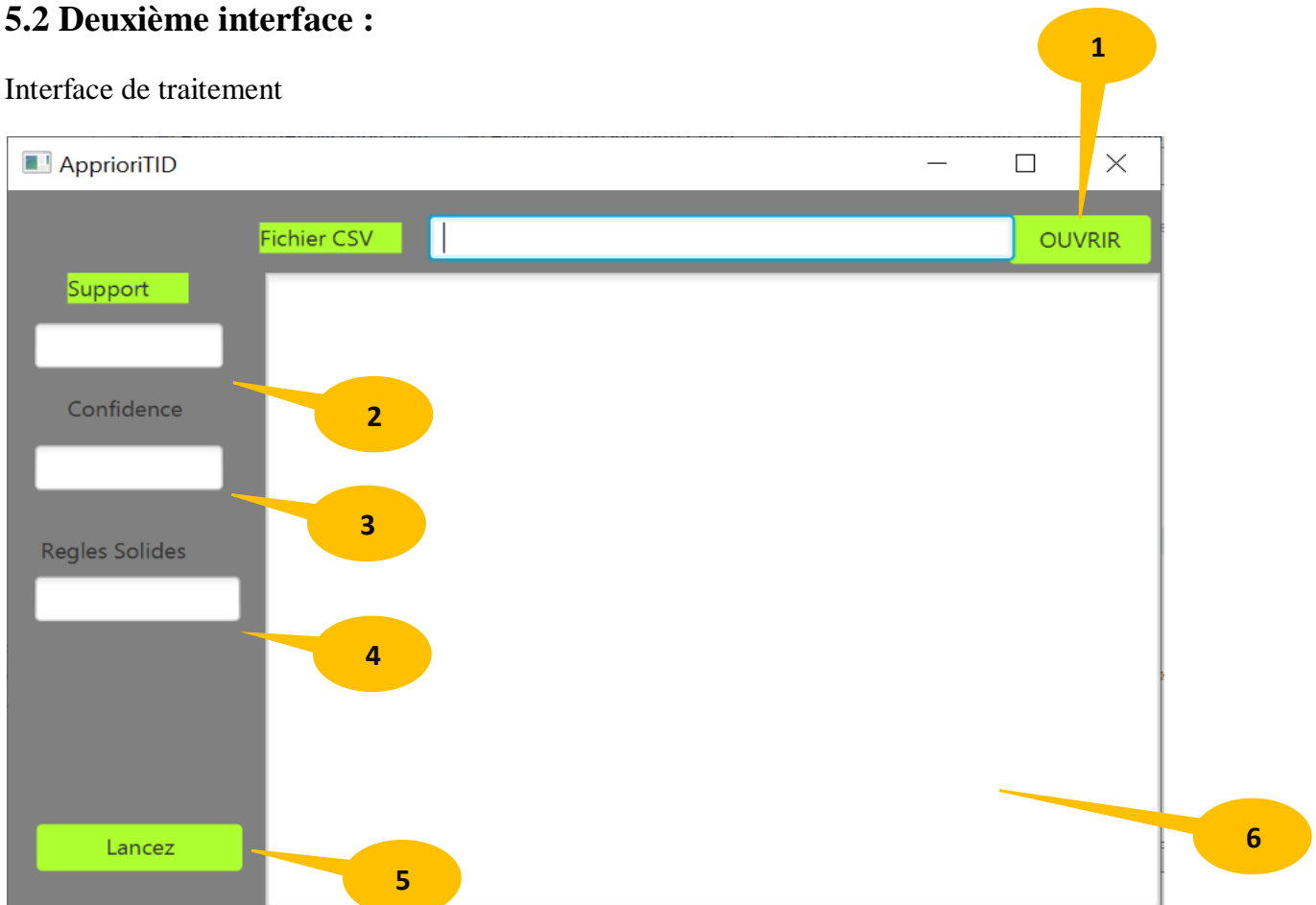


Figure 20 : interface de traitement

1. Le bouton qui nous permet de définir et de choisir la base de données à partir de laquelle nous allons extraire les règles d'association, en téléchargeant son lien.
2. L'espace alloué pour écrire le support minimum, pour déterminer les itemsets fréquents dans la base de données.
3. L'espace désigné pour rédiger la confiance minimale des règles d'association.
4. Déterminer le nombre des règles d'association à apparaître dans l'interface, qui apparaissent classées en fonction de Leur dureté et de leur rigidité.
5. Bouton pour Exécuter l'application.
6. La zone dans laquelle les résultats sont présentés après le traitement analyse de la base de données.

Nous avons mené une expérience sur une mini base de données de forme CSV extraite de la grande base de données groceries pour montrer des résultats simples et faciles à expliquer. La base que nous avons utilisée est la suivante :

	A	B	C
1	milk,chocolate,diaper		
2	milk,bread,butter,diaper		
3	milk,bread,butter,chocolate,diaper		
4	bread,butter,chocolate,diaper		
5	bread,butter,chocolate		

Figure21 : base de données transactionnel

Après avoir sélectionné la base de données et commencé le travail, le premier processus consiste à convertir la base de données de transactionnel en formel et le résultat est affiché dans le console d'application Netbeans, comme indiqué ci-dessous :

```

bread butter milk diaper chocolate
0 0 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0
1 1 0 1 1 0 0 0 0 0 0 0 0 0 0
1 1 0 0 1 0 0 0 0 0 0 0 0 0 0
    
```

Figure 22 : base de données formel

Après avoir obtenu la base de données formel, le travail de l'algorithme Apriori TID commence maintenant, c'est-à-dire extraire les itemsets fréquents de la base de données et les

afficher à l'interface dans l'ordre de la première génération à la dernière génération, comme indiqué ci-dessous.

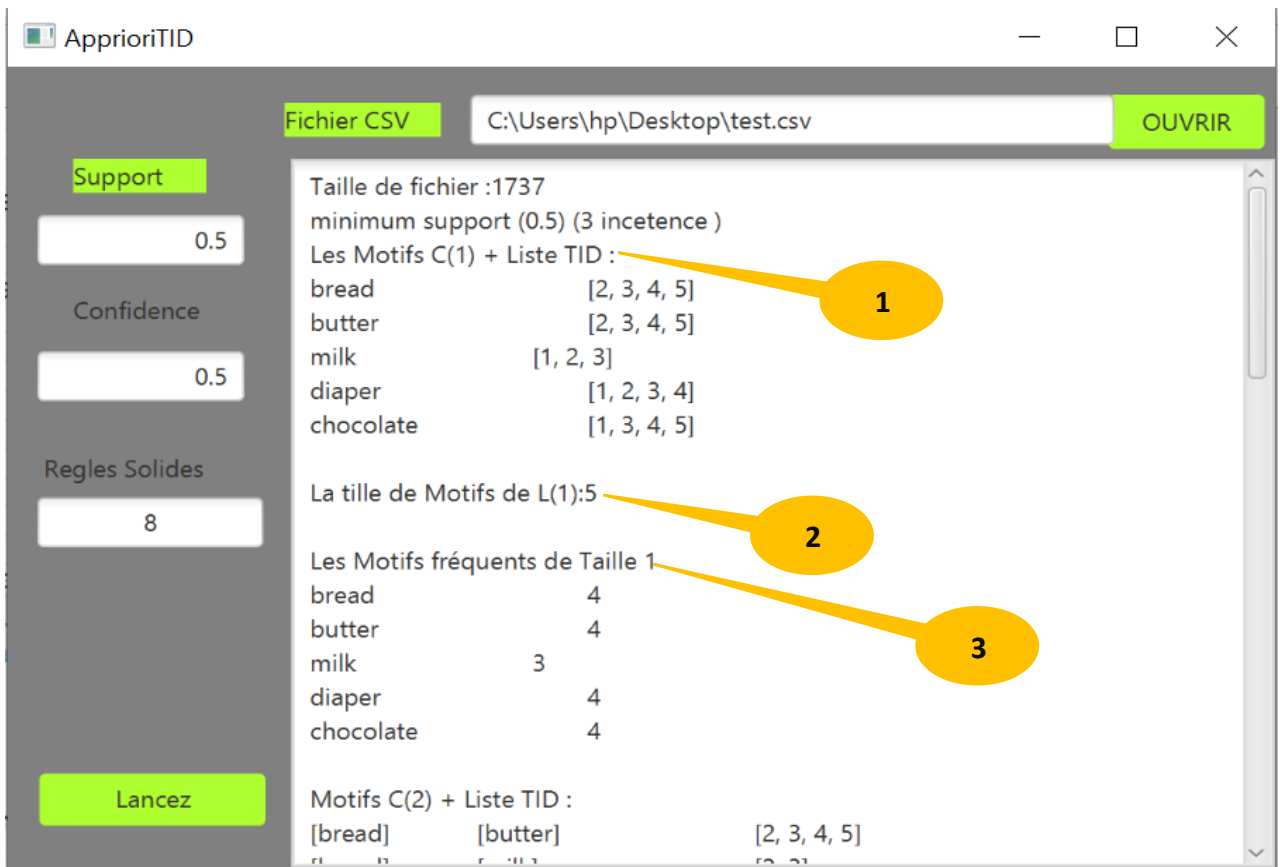


Figure 23 : génération 1 et leur itemsets fréquents

1. Génération de taille1.
2. 5 itemsetsfréquents dans cette génération.
3. Les itemsets fréquents de cette génération.

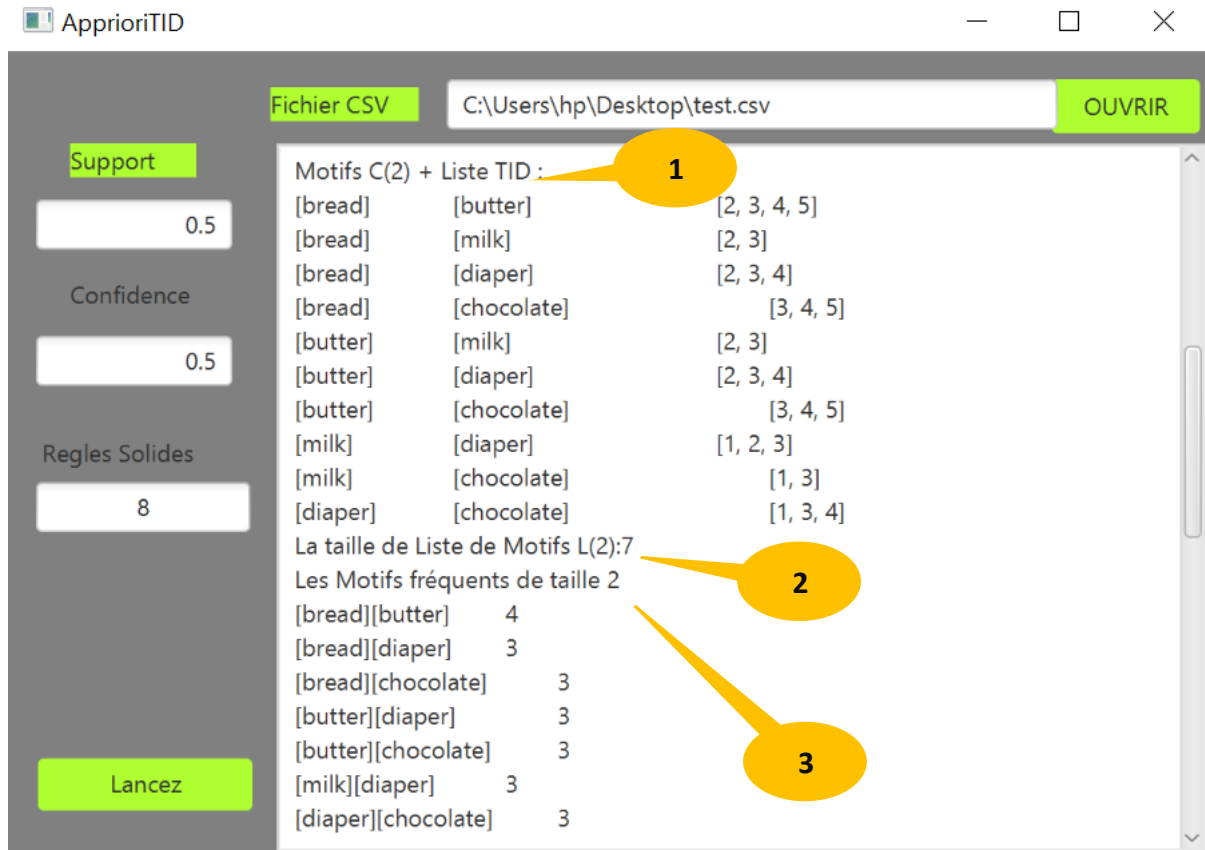


Figure 24 : génération 2 et les itemsets fréquents

1. Génération de taille2.
2. 7 itemsets fréquents dans cette génération.
3. Les itemsets fréquents de cette génération.

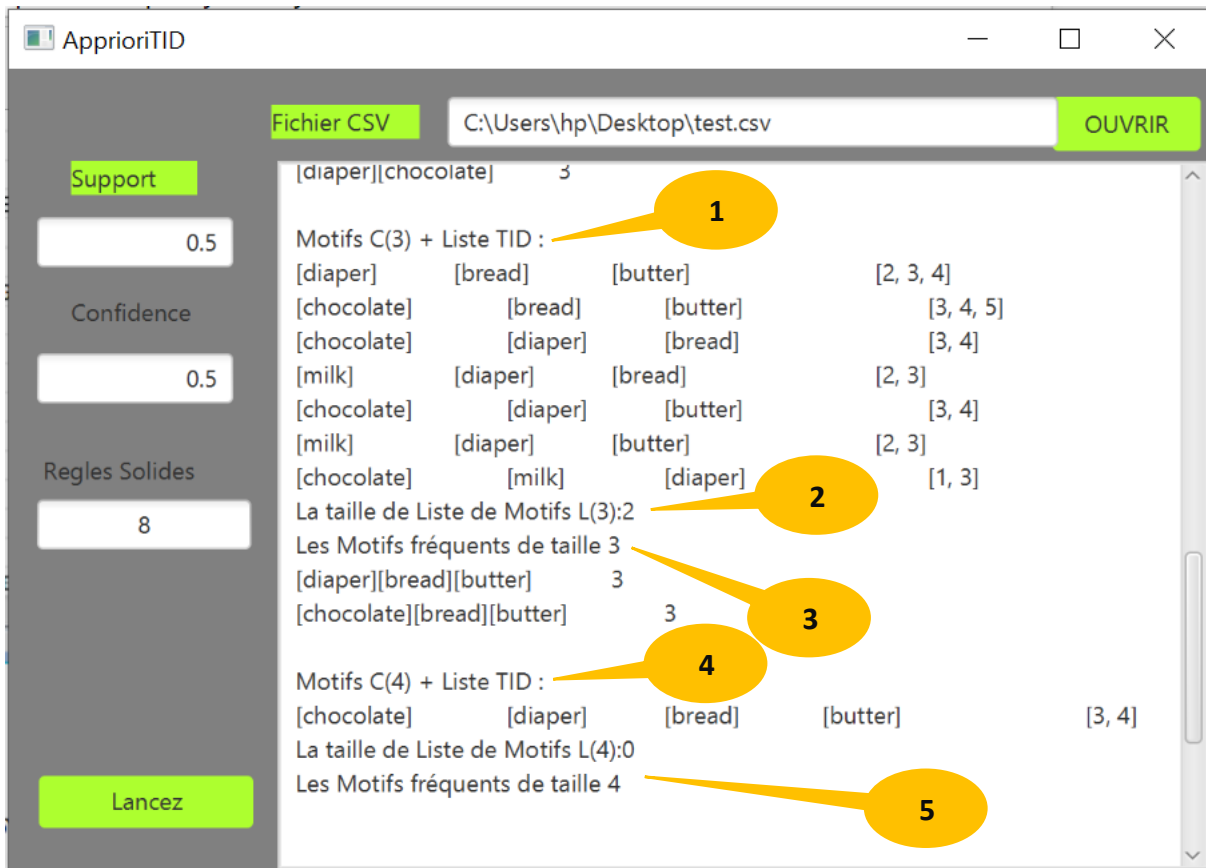


Figure 25 : génération 3 et les itemsets

1. Génération de taille3.
2. 2 itemsets fréquents dans cette génération.
3. Les itemsets fréquents de cette génération.
4. Génération de taille4.
5. Puisque la quatrième génération contient un itemset et que cette itemset n'est pas fréquent, la quatrième génération ne compte pas.

Après avoir atteint toutes les itemsets fréquents de la base de données, l'application extrait les règles d'associations associées, comme indiqué ci-dessous :

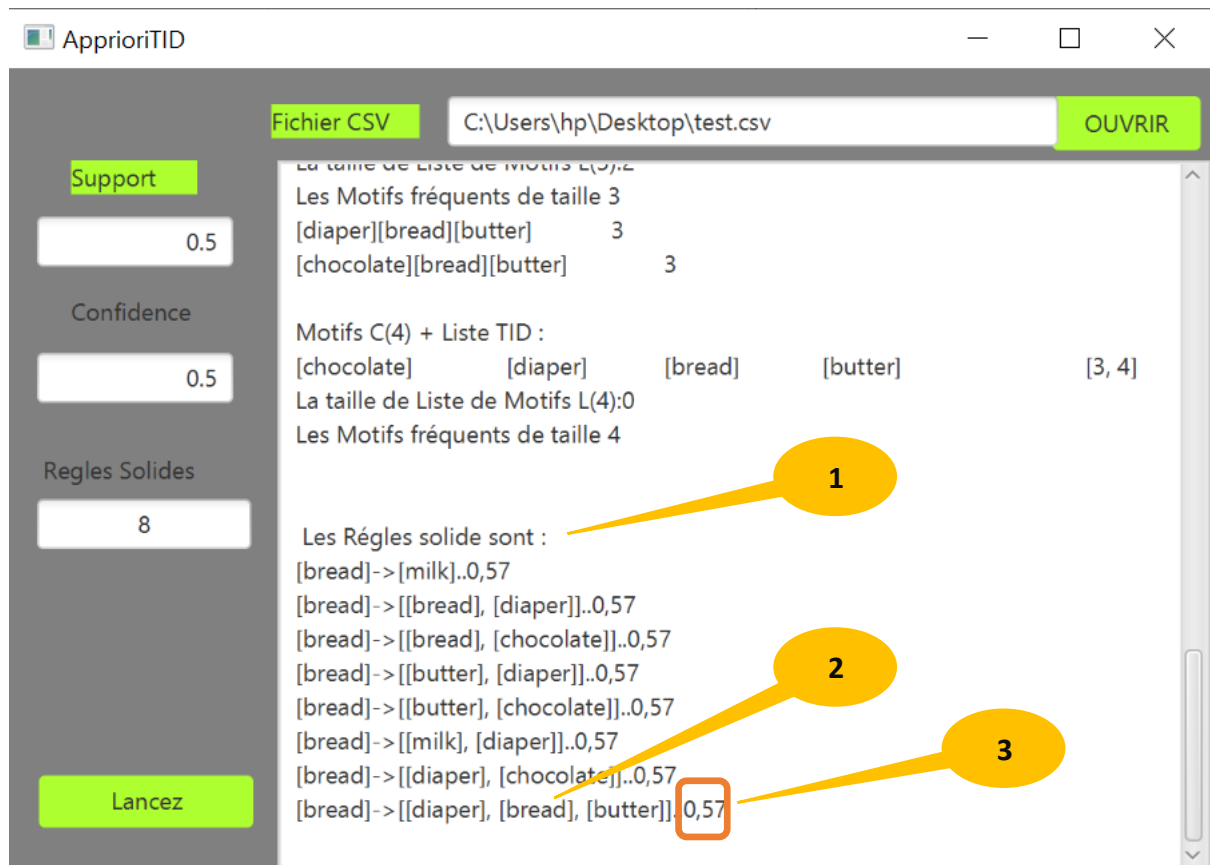
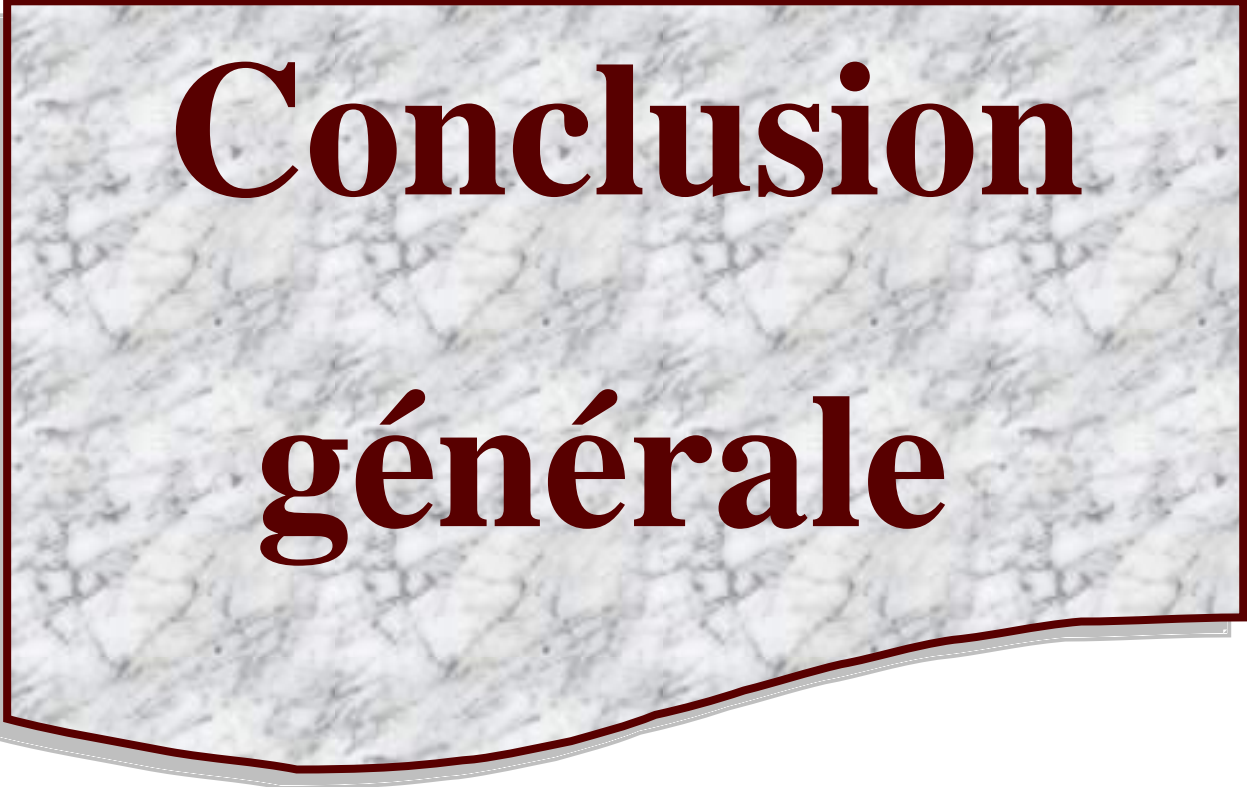


Figure 26 : les règles d’associations extraites

1. Les règles d’association.
2. La règle numéro 8.
3. La confiance de la règle numéro 8.

6. Conclusion :

Nous avons présenté dans ce chapitre l’application logicielle de notre système en respectant la modélisation qui a été établie précédemment et ce, pour répondre, aux objectifs déclarés, et Nous avons clarifié le mécanisme de notre flux de travail.



Conclusion générale

Conclusion générale

Conclusion générale

Dans ce travail, nous avons introduit une approche assez récente de fouille de données. Cela dépend de la découverte et de l'extraction des liens cachés et significatifs, appelés les règles d'association, de la grande base de données, que nous appelons les bases de données de transaction ou transactionnel.

Cette technique est considérée comme un coin de l'approche d'apprentissage symbolique, et elle est utilisée dans le domaine de l'exploration de données ou le data mining en général, elle possède un grand rôle en aidant à extraire les connaissances à travers la base utilisée.

Dans les chapitres précédents nous avons appris le data mining, nous nous sommes spécialisés dans la fouille de données de ses propriétés et méthodes, puis nous avons mis en lumière les règles d'association et les mécanismes pour les extraire.

Les règles d'association, en tant que méthode d'apprentissage non supervisé, permettent de découvrir un ensemble de règles basées sur un ensemble de transactions. Ces règles expliquent les relations et les interactions qui existent entre les différents composants de ces transactions.

La transaction se compose d'un groupe d'éléments ou bien des items, et l'un diffère de l'autre par le nombre d'éléments qu'il contient. La base des transactions que nous avons utilisée appartient au supermarché s'appelle grosseries de sorte que ses transactions contiennent les besoins d'achat des clients et chaque transaction représente le panier d'un client.

Ici, les règles d'association aident à connaître les produits vendus ensemble au fil du temps, ce qui permet au propriétaire du supermarché de connaître les intentions des clients et de comprendre leurs besoins et ainsi mieux en répondre pour l'objectif est d'augmenter les profits.

Les avantages que l'on peut tirer de l'utilisation de cet algorithme :

- Assurer la disposition et la collecte optimales des produits en fonction de l'achat fréquent de ceux-ci ensemble afin de réduire la difficulté et le temps de recherche.
- Assurer une meilleure gestion des commandes, où cette application nous permet de connaître les meilleures ventes et les plus utilisées sur le marché.

Conclusion générale.

- Connaître les articles les moins vendus.
- Mettez en place des publicités sur les produits les moins vendus et essayez de les promouvoir et d'offrir des rabais sur eux.
- Donner une vue et une vision des différents produits et de leur relation avec d'autres facteurs, tels que les produits pour les fêtes et événements religieux. A l'approche de ces fêtes, les produits qui y sont liés sont placés dans un lieu spécifique afin que des offres et des remises puissent être mises en place pendant ces périodes.

D'autre part, notre système permet de faciliter la gestion au niveau du supermarché et de comprendre comment les choses se passent en peu de temps.

Par manque de temps nous n'avons pas pu accomplir certains points que nous comptons programmer comme perspectives ou suite de ce travail :

- Etudier le changement des paramètres de l'algorithme et ainsi pouvoir les optimiser
- Distribuer la solution pour optimiser le facteur de temps
- Comblent les manques de cet algorithme par une combinaison avec un autre qui les prend en charge.
- Mettre au point un système de prise de décision intelligent qui utilise les règles d'association résultante pour prédire le comportement du client dont la satisfaction est le but de ce domaine de recherche.



BIBLIOGRAPHIE

1. G.BRESSY et C.KONKUYT ,2000
2. Prof. Rachid ZAMMAR Université Mohammed V-Agdal Faculté des Sciences Rabat
Année universitaire 2011-2012
3. Paul Guiton-Berion-05.03.2018
4. Tchahou, Herbert Nkwimi, Claude Girard, and Martin Hamel
5. Publié le 14 janvier 2012 par pgi-erp
6. <http://www.zeblogsante.com/systemes-experts-les-outils-de-la-medecine->
7. <https://www.piloter.org/techno/CRM/CRM.htm>
8. Jean--Luc Koehl – Professeur de chaire supérieure – Lycée RenéCassin – Strasbourg
– Juillet
9. Courtois, Alain, et al. Gestion de production. Les Ed. d'Organisation
10. par assamlassouad ordre des expert comptable algérien - expert comptable stagiaire
2009
11. Martin de Kerimel <https://www.cademploi.fr>
12. Y. Kodratoff. “techniques et outils de l’extraction de connaissances à partir des
données”. In : 92 (Mars 1998), 38–43
13. UM. Fayyad. Data mining and knowledge discovery: Making sense out of data. IEEE
Expert: Intelligent Systems and Their Applications, 1996 - dl.acm.org
14. DA Zighed, Y Kodratoff et Napoli. “Extraction de connaissance à partir d’une base
de données”. In : Bulletin LFIA 1 (2001)
15. Usama Fayyad, Gregory Piatetsky-Shapiro et Padhraic Smyth. “From data mining to
knowledge discovery in databases”. In: AI magazine 17.3 (1996)
16. Rakotomalala, Ricco. "TANAGRA: une plate-forme d’expérimentation pour la
fouille de données." Revue MODULAD 32 (2005): 70-85
17. Christopher Clifton. Encyclopædia Britannica : Definition of Data Mining. 2010
18. <https://www.coheris.com/logiciel-data-mining/quest-ce-que-le-data-mining>
19. Y. Kodratoff. “techniques et outils de l’extraction de connaissances à partir des
données”. In : 92 (Mars 1998), 38–43
20. DRID ABOU BAKR SEDDIK. “PATTERN RECOGNITION FOR HEALTHCARE
ANALYTICS”. Thèse de doct. University Mohammedkhider –BISKRA,
2017.HEALTHCARE ANALYTICS”. Thèse de doct. University Mohammedkhider
–BISKRA, 2017
21. Gordon S Linoff et Michael JA Berry. Data mining techniques: for marketing, sales,
and customer relationship management. John Wiley, 2011

22. AbdelHamid Djeflal. Fouille de données Avancé. cours 2018/2019
23. Michael Berry et Gordon Linoff. Mastering data mining : The art and science of customer relationship management. John Wiley, 1999
24. Gilbert Saporta. Introduction au Data Mining et à l'apprentissage statistique. Paris
25. StephaneTuffery. Data mining et statistique décisionnelle : l'intelligence dans les bases de données. Editions Technip, 2005
26. StephaneTuffery. Data mining et statistique décisionnelle : l'intelligence dans les bases de données. Editions Technip, 2005
27. Y. KodratoffCornuéjols L. Miclet. Apprentissage artificiel. Eyrolles
28. StephaneTuffery. Data mining et statistique décisionnelle : l'intelligence dans les bases de données. Editions Technip, 2005
29. LAMICHE Chaabane. "Fusion et fouille de données par les connaissances :Applicationa l'analyse d'image". Thèse de doct. Universite Mohamed Khider, 2013
30. Blanchard. Un Système de Visualisation Pour l'Extraction, l'Évaluation, et l'Exploration interactives des Règles d'Association. Thèse, Sciences et Technologies de l'Information et des Matériaux, Nantes, November 2005. Henri Briand (Dir.)]
31. Han ,Kamber,(2006) .Data Mining : Concepts and Techniques, 2nd ed. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor
32. Han ,Kamber,(2006) .Data Mining : Concepts and Techniques, 2nd ed. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor
33. R. Agrawal and T. Imielinski and A. N. Swami (1993). Mining Association Rules between Sets of Items in Large Databases .Dans Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pages 207_216
34. R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases. In Proceedings of the 20th International Conference on VLDB, pages 487–499, Santiago, Chile, September 1994
35. J. Blanchard. Un Système de Visualisation Pour l'Extraction, l'Évaluation, et l'Exploration interactives des Règles d'Association. Thèse, Sciences et Technologies de l'Information et des Matériaux, Nantes, November 2005. Henri Briand (Dir.)
36. R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases. In Proceedings of the 20th International Conference on VLDB, pages 487–499, Santiago, Chile, September 1994
37. N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Efficient Mining of Association Rules Using Closed Itemset Lattices. Inf. Syst., 24(1) :25–46, 1999

38. J. Wang, J. Han, Y. Lu, and P. Tzvetkov. TFP : An Efficient Algorithm for Mining Top-K Frequent Closed Itemsets. IEEE Transactions on Knowledge and Data Engineering, 17(5) :652–664, 2005
39. blog.khaledtannir.net
40. AbdElHamid,DjeflalCh2-Recherche-des-modèles-fréquents-corrélations-et-associations.pdf Cours2018/2019
41. <https://www.maxicours.com/se/cours/les-partenaires-de-l-entreprise/>
42. Berrah, Lamia. La quantification de la performance dans les entreprises manufacturières: de la déclaration des objectifs à la définition des systèmes d'indicateurs. Diss. 2013.
43. Berrah, Lamia. La quantification de la performance dans les entreprises manufacturières: de la déclaration des objectifs à la définition des systèmes d'indicateurs. Diss. 2013.
44. Stéphane, Tufféry. Data mining et statistique décisionnelle: l'intelligence des données. Editions Technip, 2012.
45. all& BELVIDI Lind gifiane Université d'Oran Faculté des Sciences Economiques, de Science de Gestion et des Science Commerciales
<https://www.memoireonline.com/04/12/5648/Traitement-d-un-dossier-de-credit-d-exploitation-Cas-Banque-exterieure-d-Algerie.html>
46. ftp://public.dhe.ibm.com/software/data/sw-library/cognos/fr/pdfs/brochures/br_detection_fraude_assurance.pdf
47. Badot, Olivier, and Jean-François Lemoine. "Les stratégies d'innovation dans le commerce indépendant de proximité." Décisions Marketing (2010) : 63-66.
48. Sandra Bringay. Fouille de données de santé. Informatique [cs]. Université de Montpellier, 2015.
49. Pasquier, Nicolas. Data mining: algorithmes d'extraction et de réduction des règles d'association dans les bases de données. Diss. 2000.
50. BEN NEDJEMA BELKACEM et BENLAITER AMEURE UNIVERSITE DE M'SILA FACULTE DES MATHEMATIQUES ET DE L'INFORMATIQUE Département d'Informatique 2011 /2012
51. R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases. In Proceedings of the 20th International Conference on VLDB, pages 487–499, Santiago, Chile, September 1994.
52. International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 4, Issue 7 (November 2012), PP. 53-57
53. BEN NEDJEMA BELKACEM et BENLAITER AMEURE,UNIVERSITE DE M'SILA FACULTE DES MATHEMATIQUES ET DE

54. M. J. Zaki and C.-J. Hsiao. CHARM: An Efficient Algorithm for Closed Itemset Mining. In Proceedings of the Second SIAM DM, Arlington, VA, 2002. SIAM
55. S. Brin, R. Motwani, J. D. Ullman, and S. Tsur. Dynamic Itemset Counting
56. A. Savasere, E. Omiecinski, and S. Navathe. An Efficient Algorithm for Mining Association Rules in Large Databases. In Proceedings of the 21st VLDB
57. <https://www.techopedia.com/definition/24735/netbeans>
58. <https://www.theserverside.com/definition/NetBeans>
59. <https://github.com/gluonhq/scenebuilder/wiki>
60. <https://www.futurasciences.com/tech/definitions/internet-java-485/>
61. <https://www.lemagit.fr/definition/Java>
62. <https://docs.oracle.com/javafx/2/overview/jfxpub-overview.htm>
63. M. KHODJA 2 ème année LMD Gestion des entreprises <https://www.ummtto.dz/wp-content/uploads/2018/02/Gestion-d-entreprise-Chapitre-1-2-ann%c3%a9e-SFC-Section-B.pdf>
64. <https://www.kaggle.com/irfanasrullah/groceries?fbclid=IwAR2AQohbegl8wmc19lpqH3GTyn2VMEt7IDQbDOFC-q4NpzZfrlxZcNE27c8>
65. whatis.techtarget.com/fr/d%C3%A9finition/donn%C3%A9es-transactionnelles
66. Mohammed J. Zaki Rensselaer Polytechnic Institute October–December 1999 Parallel and Distributed Association Mining.