

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

**Université 20 Août 1955-Skikda**  
**Faculté des Sciences – Département d'informatique**

**Master académique en informatique**  
**Option : Génie Logiciel Avancé et Applications**



**Thème**

**ANALYSE MULTICRITERES BASEE  
REGLES D'ASSOCIATION POUR LA PRISE DE DECISION  
MEDICALE**

**Réalisé par:**

BOUKEBIR Yasmine  
REBBOUH Sabrina

**Encadré par :**

A. MANSOUL

**Année Universitaire : 2023/2024**

# *Remerciements*

*Avant tout, nous remercions ALLAH qui nous a accordé cette faveur.*

*Toute notre gratitude vas à ceux dont les encouragements,  
les contributions et les idées nous furent si précieux durant  
l'élaboration de ce travail.*

*Parmi ceux qui nous ont patiemment aidé et soutenu dans cette  
épreuve nous citons notre encadrant Monsieur A.MANSOUL  
qui a supervisé la réalisation de ce travail dans les meilleures conditions.*

*Nous tenons aussi à remercier les membres du jury  
d'avoir accepté de juger notre travail et de nous honore de leur présence.*

*Nous voulons adresser nos vifs remerciements à  
tous les enseignants du département informatique  
de l'Université du 20 aout 1955 de Skikda nous remercions  
également tous ceux qui nous ont aidé de près ou de loin  
dans la réalisation de notre mémoire.*



## *Dédicaces*

*Je dedie ce modeste travail a ma mere qui m'a donne la vie  
et qui incarne la tendresse et le sacrifice pour mon  
bonheur et ma reussite.*

*A mon père, en tant qu'ecole de mon enfance,  
a ete mon ombre bienveillante tout au  
long de mes annees d'etudes,  
toujours la pour m'encourager, m'aider et me proteger.*

*A mes chers frères IMED et SALAH*

*Je tiens a exprimer ma gratitude envers mon coéquipier  
rebbouh Sabrina,  
avec qui j'ai partage des moments a la fois joyeux et stressants.*

*Je suis très reconnaissante pour votre soutien  
indéfectible et votre amour inconditionnel.*



*Yasmine boukebir*

## *Dédicaces*

*Je dédie cet humble travail à tous ceux qui m'ont  
encouragé à étudier quand j'avais cet âge, en particulier à ma chère mère, qui m'a  
prodigué chaque jour avec des conseils et des prières.*

*Je le dédie à toute ma famille, en particulier à mon mari, qui m'a encouragé et  
soutenu. moi et mes enfants : Alaa, Muhammad, Abdul-Mumen et Abdullah.*

*Et à mes frères, chacun par son nom,  
et à tous les amis, hommes et femmes, et à tous les honorables professeurs qui ne  
m'ont épargné aucune information.*

*Merci beaucoup*



**REBBOUH Sabrina**

# TABLE DES MATIERES

<b>Remerciements</b>	
<b>Dédicaces</b>	
<b>Liste des figures</b>	
<b>Résumé</b>	
<b>Introduction générale .....</b>	<b>3</b>

## **Chapitre 01 : L'analyse multicritères**

1.1 Introduction.....	5
1.2 Notions d'analyse multicritères.....	5
1.3 Les problématiques multicritères .....	7
1.4 Les méthodes multicritères .....	10
1.5 Conclusion .....	14

## **Chapitre 02 : Les règles d'association**

2.1 Introduction.....	16
2.2 La recherche d'associations .....	16
2.2.1 Définition et concepts de base.....	18
2.2.2 Algorithmes de recherche d'associations .....	18
2.2.3 Processus de recherche d'associations.....	19
2.3 État de l'art en analyse multicritères par les méthodes de fouille des données .....	20
2.4 Conclusion .....	22

## **Chapitre 03 : Conception : Approche de l'analyse multicritères basée règles d'association**

3.1 Introduction.....	24
3.2 Architecture générale du système .....	24
3.3 Le système d'analyse multicritères par les règles d'association .....	25
3.3.1 Le dataset .....	26
3.3.2 Le prétraitement des données.....	26
3.3.3 La phase d'apprentissage .....	27
3.4 Conclusion .....	31

## **Chapitre 04 : implémentation et résultats**

4.1 Introduction.....	33
4.2 Environnement de développement.....	33

4.3 Interface principale du système.....	34
4.4 Le prétraitement des données .....	36
4.5 L'apprentissage.....	37
4.6 Les résultats et discussions .....	38
4.6.1 Les métriques de performance utilisées.....	38
4.6.2 Les résultats et discussion.....	38
4.7 Conclusion .....	40
<b>Conclusion générale.....</b>	<b>41</b>
<b>Bibliographie</b>	

## LISTE DES FIGURES

<b>Figure 1</b> : Classement des méthodes selon le type de problématique multicritère de décision .....	8
<b>Figure 2</b> : Problématiques de choix $P.\alpha$ .....	8
<b>Figure 3</b> : Problématiques de tri $P.\beta$ .....	9
<b>Figure 4</b> : Problématiques de rangement $P.\gamma$ .....	9
<b>Figure 5</b> : Graphe des Solution .....	13
<b>Figure 6</b> : Le processus de décision multicritères .....	14
<b>Figure 7</b> : Architecture générale du système.....	24
<b>Figure 8</b> : Processus du système .....	25
<b>Figure 9</b> : Dataset.....	26
<b>Figure 10</b> : Prétraitement.....	27
<b>Figure 11</b> : Analyse Multicritère .....	30
<b>Figure 12</b> : Ajout de règles d'association dans le tableau des performances.....	31
<b>Figure 13</b> : Code python interface.....	35
<b>Figure 14</b> : Interface du système .....	36
<b>Figure 15</b> : Le prétraitement .....	36
<b>Figure 16</b> . Code python pour l'extraction des règles d'associations .....	37
<b>Figure 17</b> : Résultats des règles d'assotions .....	38
<b>Figure 18</b> : Le décision final .....	40

## LISTE DES Tableaux

<b>Tableau 1</b> : Tableau de performance .....	12
<b>Tableau 2</b> : Matrice de concordance .....	12
<b>Tableau 3</b> : Matrice de discordance .....	12
<b>Tableau 4</b> : sur-classement .....	13
<b>Tableau 5</b> : Tableau des données .....	28
<b>Tableau 6</b> : Tableau de performance.....	29

## **RESUME**

L'aide à la décision, l'analyse multicritère et les règles d'association sont interconnectées dans le contexte d'une prise de décision médicale basée sur des données complexes. L'aide à la décision utilise des outils et des techniques analytiques pour faciliter la sélection de la solution optimale parmi plusieurs alternatives. L'analyse multicritère permet de poser un problème et de définir une solution à ce problème à travers un tableau de performances dans lequel est inscrit le problème posé. Dans ce contexte, nous travaillerons à définir une solution d'aide à la décision médicale. s'exprime, dans un premier temps à l'aide d'une règle d'association dérivée d'un ensemble de données médicales. Nous avons développé le système qui nous permet de travailler sur la base d'un ensemble de règles médicales, qui alimentent un processus d'analyse multicritère dans lequel la meilleure solution est calculée (règle) et cela nous donnera la meilleure règle à appliquer pour la meilleure aide à la décision.

Le système que nous avons proposé décide quel médicament prescrire pour le malade avec des caractéristiques bien déterminées

Ce système nous a donné un résultat satisfaisant

**Mots Clés : aide à la décision, analyse multicritères, règles d'association, apprentissage, aide à la décision médicale, apriori**

## **ABSTRACT**

Decision support, multi-criteria analysis and association rules are interconnected in the context of medical decision-making based on complex data. Decision support uses analytical tools and techniques to facilitate the choice of an optimal solution among several alternatives. Multi-criteria analysis allows us to pose a problem and identify a solution to this problem through a performance table which writes the problem posed, it is in this context, which we will work to identify a solution to aid decision-making. medical, which is expressed, initially using an association rule drawn from a medical data set. From this party we have developed a system which allows us to work on the basis of a set of medical rules. association, which feeds a multi-criteria analysis process which will be a calculation of the best solution (rule) and this will give us, the best rule applied to have the best decision aid The system that we have proposed decides which medicine prescribed for the patient with well-defined characteristics

This system gave us a satisfactory result

**Keywords: decision support, multi-criteria analysis, association rules, learning, medical decision support, apriori**

## المخلص

إن دعم القرار والتحليل متعدد المعايير وقواعد الارتباط مترابطة في سياق اتخاذ القرارات الطبية بناءً على البيانات المعقدة. يستخدم دعم القرار الأدوات والتقنيات التحليلية لتسهيل اختيار الحل الأمثل من بين عدة بدائل. يتيح لنا التحليل متعدد المعايير طرح المشكلة وتحديد حل لهذه المشكلة من خلال جدول الأداء الذي يكتب فيه المشكلة المطروحة، وفي هذا السياق سنعمل على تحديد حل يساعد على اتخاذ القرار الطبي يتم التعبير عنها، في البداية باستخدام قاعدة الارتباط المستمدة من مجموعة البيانات الطبية، قمنا بتطوير نظام يسمح لنا بالعمل على أساس مجموعة من القواعد الطبية، والتي تغذي عملية تحليل متعددة المعايير يتم حساب أفضل حل (قاعدة) وهذا سيعطينا أفضل قاعدة مطبقة للحصول على أفضل مساعدة في اتخاذ القرار. النظام الذي اقترحناه يقرر الدواء الموصوف للمريض بخصائص محددة جيدًا لقد أعطانا هذا النظام نتيجة مرضي!

الكلمات المفتاحية: دعم القرار، التحليل متعدد المعايير، قواعد الارتباط، التعلم، دعم القرار الطبي، البديهية

# INTRODUCTION GÉNÉRALE

### Problématique

Dans un monde de plus en plus complexe et interconnecté, les décideurs sont confrontés à des choix multiples et souvent conflictuels. L'aide à la décision devient ainsi un enjeu crucial, où l'analyse multicritères (AMC) joue un rôle fondamental en permettant de prendre en compte plusieurs critères simultanément pour évaluer les options possibles.

Cependant, l'AMC peut devenir rapidement complexe à gérer, surtout lorsqu'on traite de grandes quantités de données. C'est ici qu'interviennent les règles d'association, une technique de data mining qui permet de découvrir des relations cachées entre les variables dans de vastes ensembles de données. La problématique centrale de cette recherche est donc de comprendre comment combiner efficacement l'analyse multicritères et les règles d'association pour améliorer les processus de prise de décision.

Comment ces deux approches peuvent-elles être intégrées pour offrir une solution robuste et flexible aux problèmes décisionnels complexes ?

### Objectifs de recherche.

L'objectif principal de cette recherche est de développer une approche innovante qui intègre les règles d'association dans l'analyse multicritères pour améliorer la qualité et l'efficacité de l'aide à la décision. Plus précisément, nous visons à :

1. Développer une méthode systématique pour combiner les critères d'analyse multicritères avec les règles d'association (malade  $\Rightarrow$  médicament).
2. Concevoir et implémenter un système capable de réaliser cette intégration.
3. pour faciliter le choix d'une solution optimale parmi plusieurs alternatives
4. Évaluer l'efficacité du système à travers des études de cas et des simulations pour la décision médicale.
5. Proposer des améliorations et des perspectives futures pour l'application médicale.

### L'organisation de mémoire.

Ce document est divisé en quatre chapitres organisés comme suit :

- **Introduction générale**

La section d'introduction générale présente le contexte et la problématique de la prise de décision multicritères et l'utilisation des règles d'association. Elle expose également les objectifs de recherche et donne un aperçu de la structure du mémoire.

- **Chapitre 1 : L'analyse multicritères**

Dans ce chapitre, nous introduisons les notions fondamentales de l'analyse multicritères. Nous définirons les actions, les critères et les poids utilisés dans cette méthode. Nous discuterons de la problématique sous-jacente et présenterons les principales méthodes multicritères existantes avant de conclure ce chapitre.

- **Chapitre 2 : Les règles d'association**

Ce chapitre débutera par une introduction à la recherche d'associations, incluant les définitions et concepts de base. Nous décrirons ensuite les algorithmes de recherche d'associations, les processus impliqués, et les différentes applications des règles d'association. Enfin, nous fournirons un état de l'art en combinant l'analyse multicritères avec les règles d'association avant de conclure.

- **Chapitre 3 : Conception : Approche de l'analyse multicritères par les règles d'association**

Dans ce chapitre, nous présenterons l'architecture générale du système développé. Nous décrirons le système d'analyse multicritères basé sur les règles d'association, en détaillant le dataset utilisé, le prétraitement des données et la phase d'apprentissage. Ce chapitre se terminera par une conclusion.

- **Chapitre 4 : Implémentation et Résultats**

Le dernier chapitre décrit l'environnement de développement et l'interface du système implémenté. Il présente en détail le prétraitement des données, les étapes d'apprentissage et discute des résultats obtenus. Les métriques de performance utilisées sont expliquées, suivies d'une discussion sur les résultats et leurs implications. Enfin, une conclusion synthétise les principaux apports de cette recherche.

- **Conclusion générale**

Enfin, ce chapitre résume les principales découvertes de l'étude, discute des limitations et des implications pratiques de l'intégration de l'analyse multicritères et des règles d'association, et propose des pistes pour des recherches futures

# **CHAPITRE 1 : ANALYSE MULTICRITERES**

## 1.1 Introduction

L'analyse multicritère AMC se présente comme une alternative aux méthodes classiques basées sur la définition d'une fonction unique, souvent exprimée en terme économique et qui reflète la prise en compte de plusieurs critères. L'intérêt des méthodes multicritères est de considérer un ensemble de critères de différente nature (exprimés en unités différentes), sans nécessairement les transformer en critères économiques ni en une fonction unique. Il ne s'agit pas de rechercher un optimum, mais une solution compromise qui peut prendre diverses formes : choix, affectation ou classement.

Plusieurs méthodes existent dans la littérature, ces méthodes multicritères connaissent de multiples applications dans différents domaines tels que le transport, agriculture, le management de territoire, la politique et la conception de réseaux de télécommunication. Dans ce chapitre nous allons définir le cadre théorique et les aspects méthodologiques des méthodes multicritères, ensuite nous allons illustrer leurs approches en étudiant un types de méthode multicritère : ELECTRE TRI.

Selon Roy [1] : « L'aide à la décision est l'activité de celui qui prenant appui sur des modèles clairement explicités, mais non nécessairement complètement formalisés, aide à obtenir des éléments de réponses aux questions que se pose un intervenant dans le processus de décision, éléments concourant à éclairer la décision et normalement à prescrire, ou simplement à favoriser un comportement de nature à accroître la cohérence entre l'évolution du processus d'une part, les objectifs et le système de valeurs au service desquels cet intervenant se trouve placé d'autre part »

## 1.2 Notions d'analyse multicritères

### • Actions, critères et poids

Nous avons vu que l'action représentait l'objet de la décision, mais nous n'avons pas évoqué sa réalisabilité. Afin de différencier les actions réalisables de celles qui ne le sont pas, [3] nomme action potentielle ou alternative une action-réalisable, c.-à-d. une action dont la mise en œuvre en pratique est envisageable. En aide multicritère à la décision l'ensemble des alternatives A est généralement construit sous forme d'une liste :  $A = \{a_1, a_2, \dots\}$ .

**Action** : Une action peut être [1] :

- **Réelle** : Elle correspond à un projet complètement élaboré susceptible d'être mis à exécution. Mais nous serons souvent amenés à considérer, dans l'aide à la décision, les actions fictives.

- **Fictive / Imaginaire** : est considérée soit comme une action idéalisée pour connaître la réaction du décideur, soit comme objet de référence (ou prototype) dans les problèmes de classification multicritère, les actions fictives peuvent être réalistes ou non réalistes.
- **Réaliste** : correspond à un projet dont la mise à exécution peut être raisonnablement envisagée.
- **Potentielle** : est une action réelle ou fictive provisoirement jugée réaliste par un acteur au moins ou présumée telle par l'homme d'étude en vue de l'aide à la décision.

L'ensemble des actions noté  $A$  est l'ensemble des objets, décisions, candidats, que l'on va explorer dans le processus de décision.

**Critère** : Un critère  $g$  est une fonction à valeurs réelles définie sur l'ensemble  $A$  des actions potentielles de telle sorte qu'il soit possible de raisonner ou de décrire le résultat de la comparaison de deux actions  $a$  et  $b$  à partir des deux nombres  $g(a)$  et  $g(b)$ .

On distingue plusieurs types de critères [2] :

- **Vrai-critère** : la structure de préférence sous-jacente est une structure de Pré-ordre total ("modèle traditionnel").
- **Quasi-critère** : la structure de préférence sous-jacente est une structure de quasi-ordre ("modèle à seuil").
- **Critère d'intervalle** : la structure de préférence sous-jacente est une structure d'ordre d'intervalle ("modèle à seuil variable").
- **Pseudo-critère** : la structure de préférence sous-jacente est une structure de pseudo-ordre.

**Poids** : Les poids sont des coefficients attribués à chaque critère pour refléter leur importance relative dans la décision. Par exemple, si la durabilité est considérée comme un critère très important lors du choix d'un fournisseur, on peut lui attribuer un poids plus élevé par rapport à d'autres critères comme le coût.

### 1.3 La problématique

Dans un processus d'aide à la décision, lors de la construction du modèle d'évaluation, il est rare d'aboutir uniquement à un seul critère correspondant à un point de vue unique sur lequel le décideur exprimera ses préférences. Il est donc nécessaire de considérer plusieurs points de vue (coûts, ressources humaines, sécurité, environnement, etc.) dans la suite de la construction du modèle d'évaluation. La décision en présence de critères multiples est difficile car les critères sont souvent conflictuels.

Pour cela, plusieurs méthodologies d'aide multicritère à la décision ont été développées [3, 4]. Dans ces méthodologies, les préférences du décideur sont d'abord élicitées sur chaque point de vue correspondant à un critère exprimé sur une échelle cardinale ou bien ordinale. Ces critères sont, ensuite, agrégés au moyen de méthodes d'agrégation multicritères. Pour cela, ces méthodologies offrent également plusieurs outils afin d'éliciter les préférences globales des décideurs pour paramétrer ces méthodes.

Nous avons vu à la section précédente qu'il y avait quatre étapes principales pour représenter de manière formelle un processus d'aide à la décision. Nous allons nous concentrer, dans cette section, sur les produits de ce processus dans le cas de l'aide multicritère à la décision : l'ensemble des actions  $A$ , la problématique, l'ensemble des dimensions (noté  $D$ ), l'ensemble des critères (noté  $F$ ) et l'ensemble des méthodes d'évaluation.

L'analyse multicritère permet de prendre en compte un nombre important d'avis et de motivations. Parallèlement, l'utilisation d'une méthode d'analyse multicritère confère de l'autorité aux décideurs et constitue un bon soutien de démonstration en direction des tiers.

La problématique en AMCD est la manière dont l'aide à la décision doit être envisagée [4]. Nous avons vu à la figure 1 qu'elle pouvait consister à partitionner l'ensemble des alternatives  $A$ . Elle peut également consister à construire l'ensemble  $A$  ou la famille de critères  $F$ . Nous pouvons donc distinguer quatre problématiques en aide multicritère à la décision [4] : les problématiques de choix, les problématiques de tri, les problématiques de rangement et les problématiques de description.

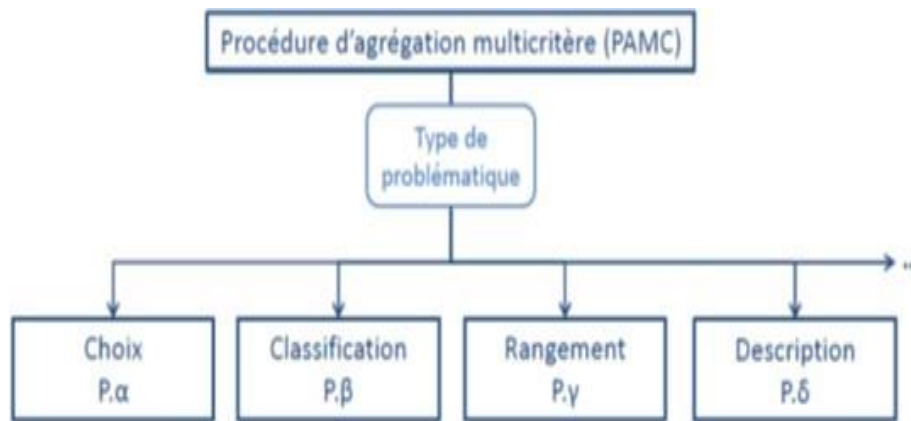


Figure 1 – Classement des méthodes selon le type de problématique multicritère de décision [1]

❖ **Problématiques de choix P.α**

Dans ce type de problématiques, l'aide à la décision est orientée de telle sorte à ce que le résultat soit une sélection d'un ensemble de « bonnes » alternatives qui soit de plus petite cardinalité possible. Bien que l'alternative qui sera recommandée soit dans cet ensemble, cela ne veut pas dire que la procédure de sélection utilisée est orientée vers cette solution. Elle consiste plutôt à éliminer les mauvaises alternatives en les comparants. L'ensemble sélectionné contient des alternatives qui sont équivalentes ou incomparables entre elles, mais ne contient pas nécessairement que de bonnes alternatives.

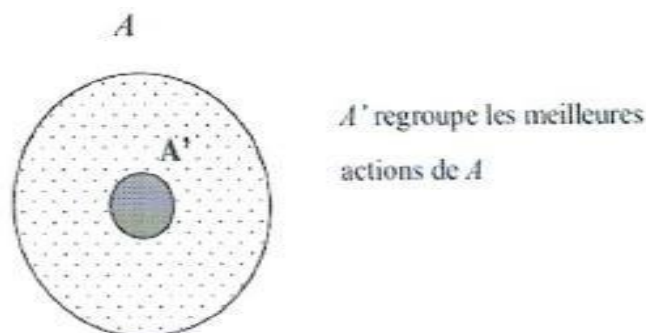


Figure 2 : Problématiques de choix P.α. [1]

❖ **Problématiques de tri P.β**

Dans ce type de problématique, l'aide à la décision est orientée de telle sorte que chaque alternative soit placée dans une catégorie jugée « appropriée » parmi des catégories prédéfinies généralement ordonnées. Le choix de cette catégorie est justifié par le type de jugement que l'on voudrait porter sur les actions et par les traitements que l'on souhaiterait faire. Par exemple, définir des catégories de confort pour les offres (très inconfortable, inconfortable, confortable, très confortable) et préciser les recommandations nécessaires afin d'encourager les fournisseurs à passer à la catégorie supérieure.

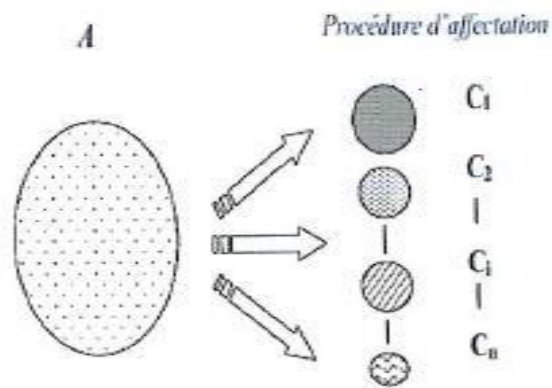


Figure 3 : Problématiques de tri P.β. [1]

❖ **Problématiques de rangement P.γ**

Dans ce type de problématiques, l'aide à la décision est orientée de telle sorte à ce que l'on ait un préordre partiel ou total sur l'ensemble A (voir l'annexe de ce chapitre pour la définition de ces notions). Il est possible de considérer les classes d'équivalence induites par ce préordre comme des catégories ordonnées.

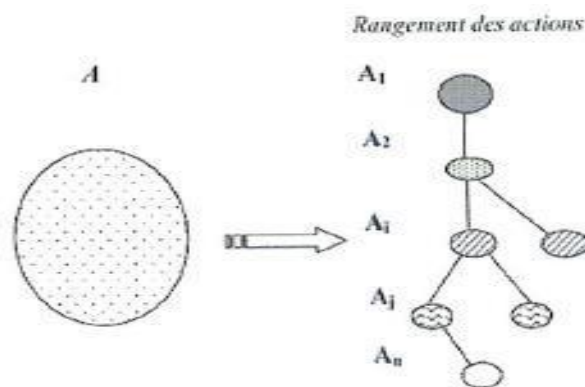


Figure 4 : Problématiques de rangement P.γ. [1]

### ❖ **Problématiques de description P.5**

Il n'est pas nécessaire que le résultat d'une aide à la décision soit une prescription (choix d'une ou d'un ensemble d'actions comme recommandation au décideur) mais il peut s'agir aussi de déterminer l'ensemble des actions potentielles  $A$ , une famille de critères  $F$  ou des valeurs possibles que peuvent prendre certains paramètres (seuil d'indifférence et de préférence, poids, niveau d'aspiration). Ce type de problématique est approprié lorsque le décideur rencontre des difficultés à définir le problème, à exprimer ses points de vue ou le type de résultat qu'il souhaiterait obtenir.

## **1.4 Les méthodes multicritères**

Les problèmes de décision multicritères opèrent, habituellement, en 4 étapes selon MAYSTRE et al. [5]. Les deux premières étapes sont communes pour toutes les méthodes multicritères à l'inverse des deux dernières qui dépendent de la méthode choisie :

### **1. Création d'une liste des actions potentielles.**

Au cours de cette étape, on établit une liste des actions potentielles qui vont rentrer en concurrence. Cette liste n'est pas exhaustive et définitive. Elle peut évoluer tout au long de l'étude (suppression ou ajout d'actions).

### **2. Création d'une liste des critères et pondération**

Il s'agit d'élaborer la liste des critères à prendre en considération. Un critère peut être plus important qu'un autre. Cette importance relative est exprimée par un nombre appelé poids.

### **3. Définition de la matrice de performances**

N'importe quelle méthode multicritère agit sur la matrice des performances. Il s'agit de juger chaque action par rapport à chacun des critères.

### **4. Agrégation des performances**

Il s'agit d'établir un modèle des préférences globales, c'est-à-dire une représentation qui formalise les préférences relativement à un ensemble  $A$  d'actions potentielles, que l'homme d'étude juge approprié au problème d'aide à la décision. Cette opération passe par la comparaison des actions entre elles.

Cela se traduit par l'utilisation de relations de comparaison. On trouve alors les notions suivantes :

(a) **Le Sur-classement.** Une action  $a_1$  surclasse une action  $a_2$ , noté «  $a_1 S a_2$  », si elle est au moins aussi bonne qu' $a_1$  relativement à une majorité de critères, sans être trop nettement plus mauvaise que  $a_2$  relativement aux autres critères.

(b) **Préférence :** cette relation permet de traduire une situation dans laquelle il existe des raisons claires et suffisantes pour mettre en évidence une préférence entre deux actions  $a$  et  $b$ .

(c) **Indifférence :** cette relation traduit une situation dans laquelle il n'existe pas de raisons suffisamment fortes pour confirmer une préférence dans un sens ou dans l'autre.

(d) **Incomparabilité :** entre les deux actions notée  $aRb$  : signifie que les deux actions sont tellement différentes l'une de l'autre qu'il est difficile de pouvoir les comparer.

\_ Les problèmes l'aide multicritère à la décision passe par la phase d'agrégation qu'est une opération multi forme (plusieurs méthode), ou l'utilisateur aura, à réduire les espaces des solutions, pour cela est vas utiliser des **seuils** (de valeurs) qui vents illimité les possibilités des combinaisons des solutions

\_ Il utilisera alors, la concordance et discordance

#### • Concordance :

La concordance entre deux critères ou alternatives se produit lorsque ces critères ou alternatives sont cohérents ou en accord les uns avec les autres. En d'autres termes, une concordance signifie que les critères ou alternatives se soutiennent mutuellement ou qu'ils sont similaires dans leur évaluation.

- **Méthodes de concordance :** On utilise souvent des méthodes numériques pour mesurer la concordance, telles que les indices de similarité ou de corrélation entre les critères ou alternatives.

#### • Discordance :

La discordance, en revanche, se produit lorsque deux critères ou alternatives sont en désaccord ou divergent l'un par rapport à l'autre. Cela indique qu'il y a une différence significative dans la manière dont ils sont évalués ou notés par rapport aux autres critères ou alternatives.

- **Méthodes de discordance :** Pour mesurer la discordance, on utilise souvent des méthodes basées sur des seuils ou des indices de dis similarité qui quantifient l'écart ou l'écart entre les critères ou alternatives.

**Exemple :**

On veut choisir un fournisseur pour une machine industrielle. Les 4 fournisseurs désignés par la direction de l'entreprise ont été retenus sur la base de 6 critères et ont reçu les notes suivantes.

**1-Le tableau de performance**

Critères	Délais livraison	Fiabilité machine	Disponibilité pièces rechange	Proximité pays	Coûts livraison	Coûts transport
Fournisseur A	9	9	5	8	6	7
Fournisseur B	9	8	5	7	6	5
Fournisseur C	7	9	7	7	6	7
Fournisseur D	10	9	7	8	8	3
<b>Poids</b>	Important 3	Important 3	Faible 1	Moyen 2	Moyen 2	Faible 1

**Tableau 1 :** tableau de performance

**2. Matrice de concordance**

	A	B	C	D
A		1,00	0,92	0,50
B	0,50		0,58	0,08
C	0,58	0,75		0,42
D	0,92	0,92	0,92	

**Tableau 2 :** Matrice de concordance

« a surclasse b »,

$$C(a, b) = \sum_{\substack{K \\ \forall j : g_j(a) \geq g_j(b)}} k_j \quad \text{avec } K = \sum_{j=1}^n k_j$$

**3. Matrice de discordance**

	A	B	C	D
A		0	0,2	0,2
B	0,2		0,2	0,2
C	0,2	0,2		0,3
D	0,4	0,2	0,4	

**Tableau 3 :** Matrice de discordance

$$D(a,b) = \begin{cases} D(a,b) = 0 & \text{si } \forall j, g_j(a) \geq g_j(b) \\ D(a,b) = \frac{1}{\delta} \max [g(b) - g(a)] & \end{cases}$$

avec  $\delta$  est la différence maximale entre le même critère pour 2 actions données.

**4. sur-classement et Graphe des Solution**

Le filtrage correspondant, avec 0.9 et 0.2, donne la matrice de sur classement

	A	B	C	D
A		x	x	
B				
C				
D		x		

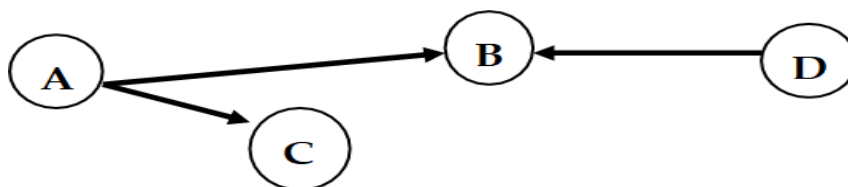
**Tableau 4 :** sur-classement

La relation de sur-classement est construite par la comparaison des indices de concordance et de discordance à des seuils limites de concordance  $c^{\wedge}$  et de discordance  $d^{\wedge}$ .

En considérant :  $c^{\wedge} = 0.9$  et  $d^{\wedge} = 0.2$

$$A \mathbf{S} b \Leftrightarrow C(a,b) \geq 0.9 \quad \text{et} \quad D(a,b) \leq 0.2$$

**A S B      A S C      D S B**



**Figure 5 :** Graphe des Solution

L'application que nous nous proposons de développer peut se voir comme étant l'interconnexion de plusieurs modules complémentaires dépendants l'un de l'autre. La figure 5 illustre les différents modules qui représentent des échanges de l'information.

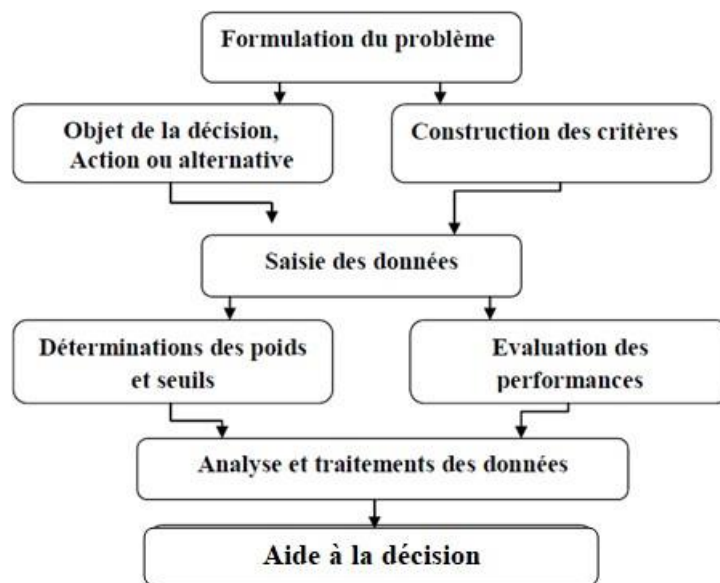


Figure 6 : Le processus de décision multicritères. [1]

## 1.5 Conclusion

Sous le terme « Aide à la décision » (AD), on désigne essentiellement les travaux qui prennent appui sur les modèles plus ou moins formalisés et qui ont pour objet d'améliorer, de faciliter ou d'accompagner le déroulement, la conduite d'un processus de décision ou encore la maîtrise ou la compréhension que peut en avoir une instance de décision.

L'AD, ainsi conçue, reprend et actualise le projet initial de la recherche opérationnelle : chercher à prendre appui sur la science pour éclairer les décisions afin de conduire les processus de décision dans les systèmes organisés. Dans le second chapitre, nous abordons le paradigme *multicritères en aide* à la décision en présentant les règles d'association.

# **CHAPITRE 2 : LES REGLES D'ASSOCIATION**

## 2.1 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, il est indéniable en revanche que son avènement s'est accompagné d'une forte accélération de la diffusion de logiciels spécialisés estampillés 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. [6]

## 2.2 La recherche d'associations

Elle consiste à trouver des corrélations ou relations entre attributs (items). La recherche de liens associatifs est utilisée intensivement dans les bases transactionnelles. Il s'agit de découvrir les items les plus fréquemment en relation. Un exemple, est la détermination des articles (cahier et stylo bleu,, ...) qui se vendent ensemble dans un supermarché.

Ainsi, pour réaliser l'une ou l'autre des tâches précitées, on utilise une ou plusieurs méthodes et elles sont nombreuses

Les règles d'association que j'ai utilisées pour identifier les symptômes et les médicaments pour chaque maladie sont des relations qui relient des items spécifiques, telles que des maladies et leurs traitements correspondants. Par exemple, si les données montrent qu'un grand nombre de patients atteints de diabète se voient prescrire de l'aspirine, cela indique une forte relation d'occurrence entre ces deux items. Cette relation est appelée association et permet de formuler la règle d'association suivante : Homme et diabète => aspirine.

Une telle règle d'association est dite classique ou traditionnelle, car les deux événements, le diagnostic du diabète et la prescription de l'aspirine, se produisent simultanément lors de la même consultation ou diagnostic.

En outre, lorsque les items apparaissent dans un ordre spécifique, par exemple si les patients diagnostiqués avec de l'asthme se voient prescrire du Lipitor dans les jours suivants, les règles d'association entre ces items ordonnés sont qualifiées d'associations séquentielles.

Pour illustrer, voici quelques exemples de règles d'association que j'ai travaillées :

- [" Homme et diabète", "aspirine", 0.25, 0.833333, 1.0]

- [" Femme et asthme", "Lipitor", 0.25, 0.833333, 1.0]

Ici, chaque règle est représentée sous forme de liste où :

- Le premier élément est le genre.
- Le deuxième élément est la maladie.
- Le troisième élément est le médicament associé.
- Le quatrième élément représente le support de la règle.
- Le cinquième élément est la confiance.
- Le sixième élément est le lift.

Ces règles d'association permettent de mieux comprendre les relations entre les symptômes des patients et les médicaments prescrits, ce qui peut aider à améliorer les traitements médicaux et la gestion des soins.

L'objectif des règles d'association, qu'elles soient appliquées aux à des données médicales, est de découvrir des associations fréquentes dans les données étudiées. Ces règles sont particulièrement utiles pour identifier des relations significatives entre des symptômes et des traitements dans le domaine médical.

### 2.2.1 Définition et concepts de base

La recherche d'associations vise à identifier des motifs récurrents, des corrélations ou des structures causales entre les ensembles d'éléments dans les bases de données transactionnelles. Une règle d'association prend généralement la forme de "si X alors Y", où X et Y sont des ensembles d'éléments.

- **Support** : Le support d'une règle est la proportion d'ensembles de données où l'ensemble d'éléments apparaît. Par exemple, si X est acheté dans 20% des transactions, alors le support de X est 0.20.
- **Confiance** : La confiance d'une règle est la proportion de transactions contenant X qui contiennent également Y. Si 70% des transactions contenant X contiennent également Y, alors la confiance de la règle  $X \Rightarrow Y$  est 0.70.
- **Lift** : Le lift mesure l'importance de la règle, c'est-à-dire combien la présence de X augmente la probabilité de la présence de Y. Un lift supérieur à 1 indique une corrélation positive.

### 2.2.2 Algorithmes de recherche d'associations

Plusieurs algorithmes ont été développés pour effectuer des recherches d'associations. Les plus populaires incluent :

- **Algorithme Apriori** : Introduit par Agrawal et Srikant, il fonctionne en identifiant d'abord les éléments fréquents individuels et en les étendant progressivement pour trouver des ensembles d'éléments plus larges. Il utilise une approche itérative basée sur des niveaux de longueur croissante d'ensembles d'éléments (k-itemsets).
- **Algorithme Eclat** : Il utilise une approche basée sur les intersections de tidsets (Transaction ID sets) pour déterminer les ensembles d'éléments fréquents. Cet algorithme est plus efficace en termes de mémoire et de temps pour les bases de données denses.
- **Algorithme FP-Growth** : Cet algorithme construit un arbre de préfixes fréquent (FP-tree) pour représenter les transactions de manière compacte et permet une extraction plus rapide des règles d'association sans nécessiter de génération de candidats explicite.

### 2.2.3 Processus de recherche d'associations

Le processus typique de recherche d'associations comprend les étapes suivantes :

1. **Préparation des Données** : Nettoyage et transformation des données en un format approprié pour l'analyse, généralement des transactions sous forme de listes d'éléments.
2. **Détection des Éléments Fréquents** : Utilisation d'algorithmes pour trouver tous les ensembles d'éléments fréquents dépassant un seuil de support minimal.
3. **Génération des Règles** : Création des règles d'association à partir des ensembles d'éléments fréquents, en calculant leur confiance et en appliquant un seuil de confiance minimal.
4. **Évaluation et Filtrage des Règles** : Évaluation des règles générées en utilisant des mesures comme le lift et l'intérêt, puis filtrage des règles non significatives.

#### ➤ Concepts de support et de confiance

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 [13].

Une règle d'association s'énonce comme suit :

**"Homme et Diabetes ⇒ Aspirin [support = 25%, confiance = 83.33%]."**

La règle d'association précédente se lit ainsi : 83.33% des patients diabétiques ont pris de l'aspirine, et 25% des transactions de la base de données correspondent à cette association.

La confiance est définie comme la probabilité d'obtenir dans une transaction le médicament (dans ce cas, de l'aspirine) sachant qu'un patient est diabétique.

Le support, quant à lui, est donné par la proportion des transactions de la base de données qui contiennent à la fois le diabète et l'aspirine.

#### Comment le calculer le support et la confiance :

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

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

Soit la règle A⇒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

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

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 [13].

### ➤ 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.

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,  $ij \not\subseteq A$ . 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 [5].

Il existe deux algorithmes consécutifs et complémentaires :

## 1- Algorithm Apriori - Génération

### Apriori-gen (F).

Entrée : F : ensembles de mots fréquents de cardinal k

#### ⇒ Début

1-  $C \leftarrow \{c = f1 \cup f2 \text{ telsque } (f1, f2) \in F \times F, \text{card}(c) = k + 1\}$

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-  $k \leftarrow 1$
- 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

## 2.3 État de l'art en analyse multicritères par les méthodes de fouille des données

La fouille de données n'est ni un outil, ni une technique, c'est plutôt une science encapsulant des méthodes dont le but est d'extraire des connaissances à partir de données assez structurées. Chaque méthode vise un objectif, i.e. une tâche, qui n'est réalisable par une seule méthode, mais plus tôt par un ensemble de méthodes complémentaires, chacune choisie en fonction du domaine d'application, du type des données à analyser, et du type des connaissances que nous voulons extraire.

Ces méthodes de fouille de données ont été utilisées très tôt en aide à la décision, généralement pour supporter le processus décisionnel.

Les règles d'association sont utilisées dans divers domaines pour diverses applications :

- Marketing et Commerce de Détail : Identifier les produits fréquemment achetés ensemble pour améliorer les stratégies de vente croisée et de merchandising.
- Analyse de Marché : Découvrir les habitudes de consommation et les préférences des clients.
- Détection de Fraude : Identifier des modèles de transactions inhabituels qui peuvent indiquer une fraude.
- Recommandations Personnalisées : Utiliser les règles d'association pour recommander des produits ou des services aux clients en fonction de leurs comportements d'achat passés.

Comme Il n'existe pas de méthodologie bien spécifique de l'intégration entre la FDD et l'AD, les chercheurs ont utilisé telle ou telle méthode au besoin de l'objectif décisionnel. Ainsi on a utilisé les méthodes suivantes :

- **Les réseaux de neurones**

Les réseaux de neurones sont des outils très utilisés pour la classification, l'estimation, la prédiction et le groupement. Ils permettent de construire un modèle qui prédit la valeur d'une variable à partir d'autres variables connues appelées variables prédictives. Si la variable à prédire est discrète (qualitative) alors il s'agit d'une classification, si elle est continue (quantitative) il s'agit alors de régression. Les méthodes les plus utilisées sont : le Perceptron multicouches et les réseaux de Kohonen.

- **Les arbres de décision**

Un arbre de décision cible la classification i.e. la prédiction de variables discrètes. Un arbre de décision peut donc être perçu comme étant un ensemble de règles qui mènent à une classe. La construction de l'arbre se fait par un algorithme approprié. Les algorithmes les plus utilisés sont : ID3, CHAID, CART, QUEST, et C5

- **SVM (machines à vecteur de support)**

Les machines à vecteurs de support sont une classe d'algorithmes d'apprentissage initialement définis pour la discrimination i.e. la prévision d'une variable qualitative initialement binaire. Elles ont été ensuite généralisées à la prévision d'une variable quantitative. Donc les SVM peuvent être utilisés pour résoudre des problèmes de discrimination (i.e. décider à quelle classe appartient un échantillon), ou de régression (i.e. prédire la valeur numérique d'une variable).

- **K - plus proches voisins (K-ppv)**

La méthode des K plus proches voisins (K-ppv, nearest neighbor ou K-nn) est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimations. La méthode K-ppv est une méthode de raisonnement à partir de cas. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire. Elle décide de la classe à laquelle appartient un nouveau cas, en examinant les k cas qui lui sont les plus similaires ou proches

- **Induction de règles**

C'est une technique qui permet d'identifier des profils, associations ou structures entre les items ou objets qui sont fréquents dans les bases de données. Pour considérer et exprimer cette association sous forme d'une règle, il faut définir des quantités numériques qui vont servir à valider son intérêt, d'où : le support et la confiance.

Ainsi, les règles dont le support et la confiance sont assez élevés sont alors privilégiées. Les algorithmes les plus utilisés sont : Apriori, FP-Growth .

- **Les modèles de Markov cachés**

Les modèles de Markov cachés d'ordre 1 ou 2 (HMM1 et HMM2 pour Hidden Markov Models) sont utilisés pour la classification des différentes données temporelles ou spatiales., ils interviennent par exemple dans de nombreux algorithmes d'analyse de séquences biologiques que ce soit pour, la détection de gènes, l'inférence et la détection de motifs, ou encore la recherche de mots exceptionnels.

- **La régression linéaire**

C'est une technique qui vise la prédiction de la valeur d'une variable continue. Son objectif est de définir le meilleur modèle qui associe une variable quantitative « output » à plusieurs variables prédictives « input ».

## **2.4 Conclusion**

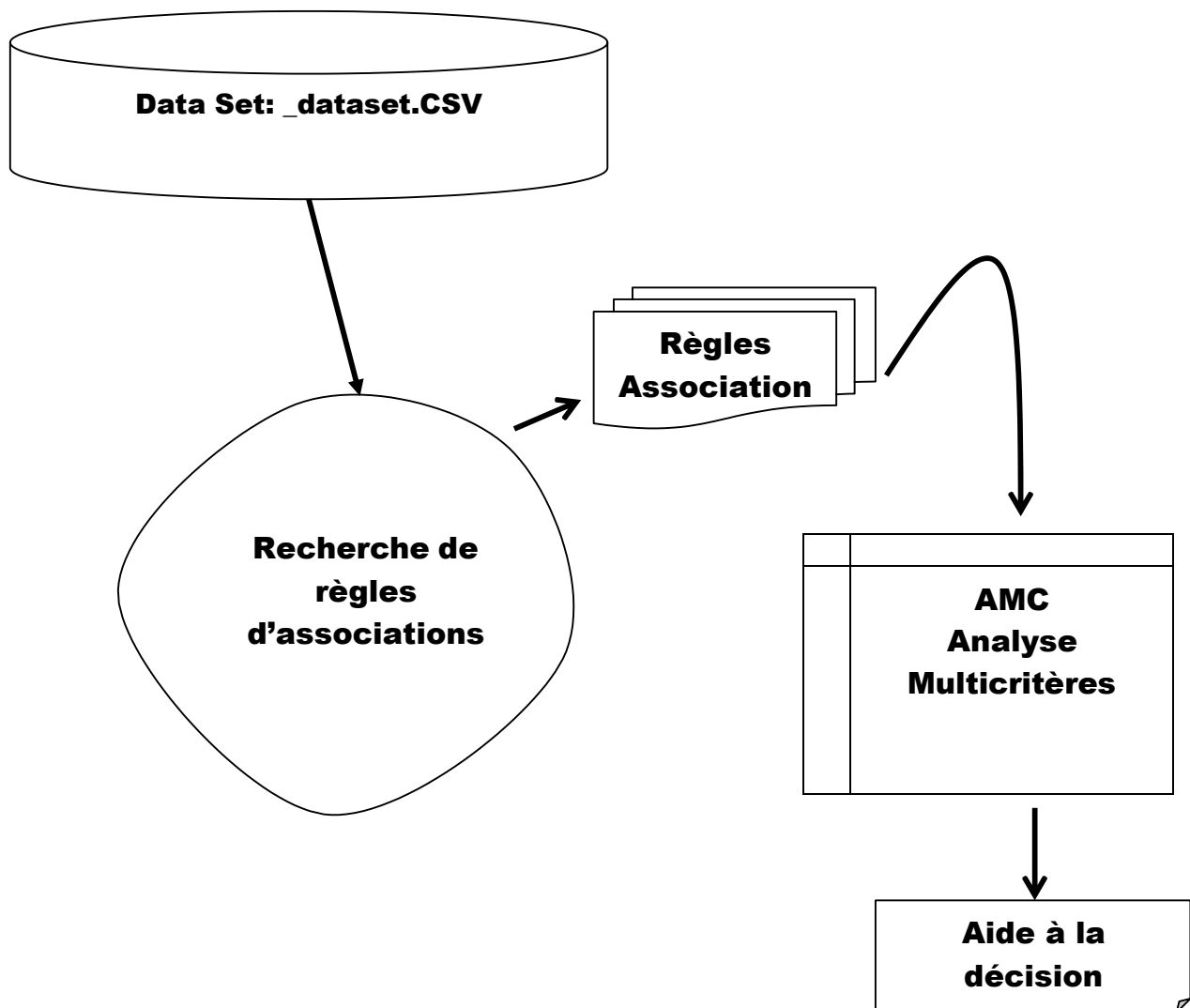
Les règles d'association offrent un cadre puissant pour la découverte de relations intéressantes dans les données. En combinant ces techniques avec l'analyse multicritères, il est possible de fournir des insights encore plus riches et plus pertinents. Cependant, cette approche présente des défis significatifs, notamment en termes de complexité computationnelle et d'interprétation des résultats, qui nécessitent des techniques avancées et des approches innovantes.

**CHAPITRE 3 : CONCEPTION : APPROCHE DE  
L'ANALYSE MULTICRITERES BASEE REGLES  
D'ASSOCIATION**

### 3.1 Introduction

Dans ce chapitre, nous aborderons l'approche de l'analyse multicritères par les règles d'association. Cette méthode offre un cadre puissant pour explorer les relations entre différentes variables dans un ensemble de données, permettant ainsi une prise de décision éclairée dans divers domaines.

### 3.2 Architecture générale du système



**Figure 7 :** Architecture générale du système

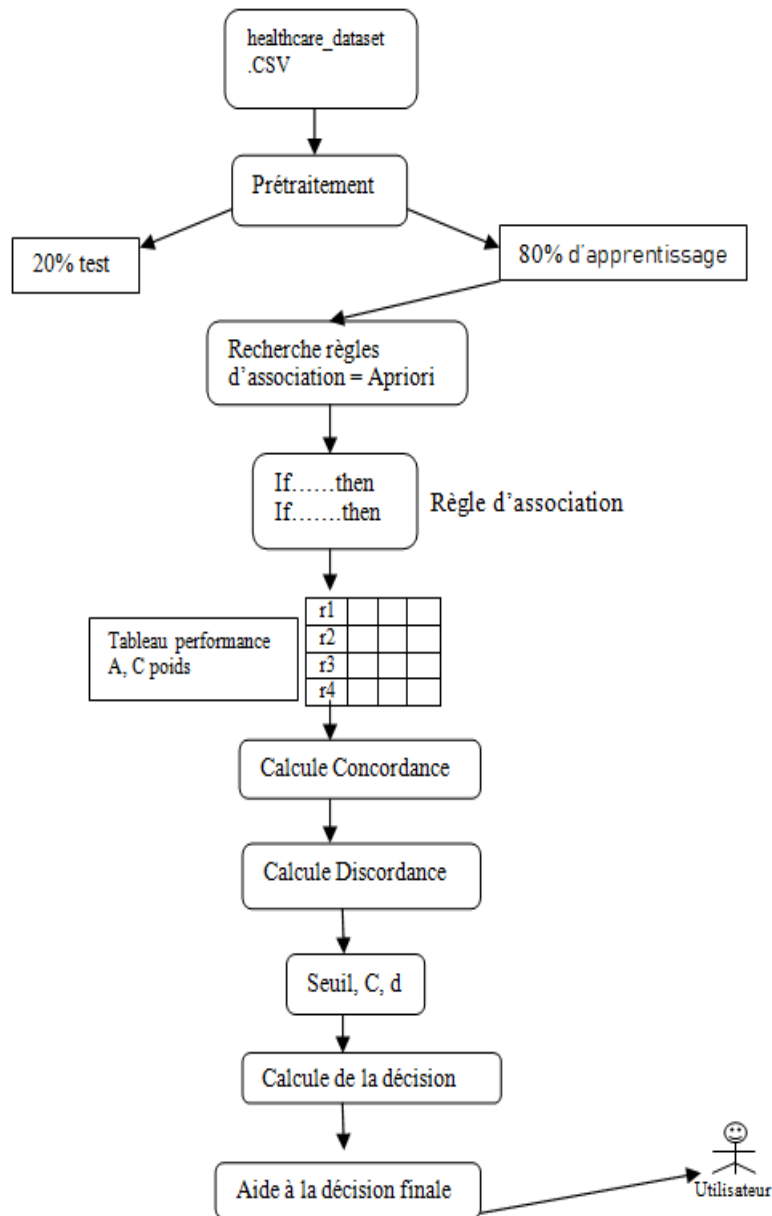


Figure 8 : Processus du système.

### 3.3 Le système d'analyse multicritères par les règles d'association

Le système d'analyse multicritères par les règles d'association permet d'extraire des connaissances à partir de grandes quantités de données en identifiant des relations ou des motifs fréquents sous forme de règles. Il se compose de plusieurs phases essentielles : le prétraitement des données, la phase d'apprentissage, et l'analyse multicritères.

### 3.3.1 Le dataset

Une attention particulière est portée à la base de données utilisée dans notre étude, qui est centrée sur les maladies. Nous explorons la structure de cette base de données et les différentes variables qu'elle contient, qui serviront de base à notre analyse.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Name, Age, Gender, Blood Type, Medical Condition, Date of Admission, Doctor, Hospital, Insurance Provider, Billing Amount, Room Number, Admission Type, Discharge Date, Medication, Test Results														
2	Tiffany Ramirez, 81, Female, O-, Diabetes, 2022-11-17, Patrick Parker, Wallace-Hamilton, Medicare, 37490.98336352819, 146, Elective, 2022-12-01, Aspirin, Inconclusive														
3	Ruben Burns, 35, Male, O+, Asthma, 2023-06-01, Diane Jackson, "Burke, Griffin and Cooper", UnitedHealthcare, 47304.06484547511, 404, Emergency, 2023-06-15, Lipitor, Normal														
4	Chad Byrd, 61, Male, B-, Obesity, 2019-01-09, Paul Baker, Walton LLC, Medicare, 36874.89699661277, 292, Emergency, 2019-02-08, Lipitor, Normal														
5	Antonio Frederick, 49, Male, B-, Asthma, 2020-05-02, Brian Chandler, Garcia Ltd, Medicare, 23303.322092196904, 480, Urgent, 2020-05-03, Penicillin, Abnormal														
6	Mrs. Brandy Flowers, 51, Male, O-, Arthritis, 2021-07-09, Dustin Griffin, "Jones, Brown and Murray", UnitedHealthcare, 18086.344183563877, 477, Urgent, 2021-08-02, Paracetamol, Normal														
7	Patrick Parker, 41, Male, AB+, Arthritis, 2020-08-20, Robin Green, Boyd PLC, Aetna, 22522.363384853266, 180, Urgent, 2020-08-23, Aspirin, Abnormal														
8	Charles Horton, 82, Male, AB+, Hypertension, 2021-03-22, Patricia Bishop, "Wheeler, Bryant and Johns", Cigna, 39593.4357611308, 161, Urgent, 2021-04-15, Lipitor, Abnormal														
9	Patty Norman, 55, Female, O-, Arthritis, 2019-05-16, Brian Kennedy, Brown Inc, Blue Cross, 13546.817249364824, 384, Elective, 2019-06-02, Aspirin, Normal														
10	Ryan Hayes, 33, Male, A-, Diabetes, 2020-12-17, Kristin Dunn, "Smith, Edwards and Obrien", Aetna, 24903.03726973869, 215, Elective, 2020-12-22, Aspirin, Abnormal														
11	Sharon Perez, 39, Female, O-, Asthma, 2022-12-15, Jessica Bailey, Brown-Golden, Blue Cross, 22788.23602637676, 310, Urgent, 2022-12-16, Aspirin, Normal														
12	Amy Roberts, 45, Male, B-, Cancer, 2021-04-13, Anthony Roberts, Little-Spencer, Aetna, 40325.071393983046, 306, Emergency, 2021-05-11, Penicillin, Abnormal														
13	Mrs. Caroline Farrell, 23, Female, O-, Hypertension, 2019-06-09, William Miller, Rose Inc, Medicare, 6185.903529552847, 126, Emergency, 2019-06-26, Paracetamol, Inconclusive														
14	Christina Williams, 85, Female, A+, Diabetes, 2021-11-29, Laura Roberts, "Malone, Thompson and Mejia", Aetna, 4835.945649525413, 444, Elective, 2021-12-14, Aspirin, Inconclusive														
15	William Page, 72, Female, A+, Diabetes, 2021-07-29, James Carney, Richardson-Powell, Cigna, 13669.377744164889, 492, Elective, 2021-08-14, Aspirin, Normal														
16	Michael Bradshaw, 65, Female, AB+, Cancer, 2021-06-05, Katherine Lowe, Castaneda-Hardy, Cigna, 10342.836123527373, 120, Emergency, 2021-06-25, Ibuprofen, Inconclusive														
17	Brian Dorsey, 32, Female, O+, Arthritis, 2021-08-07, Curtis Smith, Burch-White, Aetna, 27174.942906232496, 492, Emergency, 2021-08-14, Aspirin, Inconclusive														
18	Olivia Gonzalez, 64, Male, AB-, Diabetes, 2019-11-15, Clayton Mcknight, Cunningham and Sons, Aetna, 17394.994264453202, 315, Elective, 2019-12-04, Aspirin, Inconclusive														
19	Teresa Caldwell, 23, Male, A+, Arthritis, 2022-03-08, Debra Meyers, "Bell, Mcknight and Willis", Medicare, 45213.53762641964, 475, Elective, 2022-03-16, Ibuprofen, Inconclusive														
20	Desiree Williams MD, 66, Male, O+, Obesity, 2022-06-19, Megan Sanders, Pugh-Rogers, UnitedHealthcare, 4262.91157816964, 125, Elective, 2022-06-29, Aspirin, Inconclusive														
21	Sally Shaw, 80, Male, O-, Arthritis, 2019-07-10, Zachary Horton DDS, "Rush, Owens and Johnson", Blue Cross, 16609.311817449878, 366, Emergency, 2019-08-07, Ibuprofen, Inconclusive														
22	William Johnson, 55, Female, AB+, Arthritis, 2023-02-25, Kelly Thompson, Pearson LLC, Aetna, 32263.622156562455, 238, Emergency, 2023-03-27, Penicillin, Normal														
23	Steven Bennett, 79, Male, O-, Asthma, 2022-12-12, Michael Chang, Schultz-Powers, Blue Cross, 42610.70456262938, 364, Urgent, 2022-12-26, Penicillin, Abnormal														
24	Haley Li, 51, Male, B-, Obesity, 2022-10-09, Nicole Wood, Jordan Inc, Medicare, 16701.34713196211, 130, Emergency, 2022-11-01, Lipitor, Abnormal														
25	Angela Brown, 33, Female, B-, Diabetes, 2019-01-10, Angela Kim, Lewis-Nelson, UnitedHealthcare, 22331.280156199206, 120, Urgent, 2019-01-31, Aspirin, Abnormal														
26	Beverly Miller, 54, Male, A-, Asthma, 2022-08-05, Jodi Holland, Vaughn PLC, Cigna, 41319.500315655176, 293, Urgent, 2022-09-03, Paracetamol, Normal														
27	Daniel Dickson, 26, Female, B-, Obesity, 2021-05-27, Christina Flores, Mcknight-Mclean, UnitedHealthcare, 37766.52123735682, 292, Urgent, 2021-06-23, Paracetamol, Abnormal														
28	Kimberly Mason, 70, Female, B-, Obesity, 2021-07-12, Natalie Sullivan, Gilbert and Sons, Aetna, 35834.356603177686, 379, Elective, 2021-07-29, Paracetamol, Abnormal														
29	Francis Newman, 74, Female, AB+, Hypertension, 2021-05-25, Carolyn Baker MD, Hess-Lowe, Cigna, 12680.730521058496, 298, Emergency, 2021-06-08, Lipitor, Abnormal														
30	Ronnie Hughes, 56, Female, A-, Cancer, 2019-05-14, Nicole McClain, Hawkins Ltd, Aetna, 12935.413249863539, 392, Elective, 2019-06-11, Penicillin, Abnormal														
31	Shannon Fitzpatrick DVM, 35, Female, O+, Hypertension, 2022-12-02, Chloe Thomas, Miller-Grimes, UnitedHealthcare, 18393.421751093483, 162, Emergency, 2022-12-18, Lipitor, Inconclusive														
32	Tareca Gonzalez, 77, Female, B-, Arthritis, 2021-11-09, Brian Watson, Sanders Ltd, Medicare, 22356.082768250706, 456, Emergency, 2021-11-28, Lipitor, Abnormal														

Figure 9 : Dataset [14]

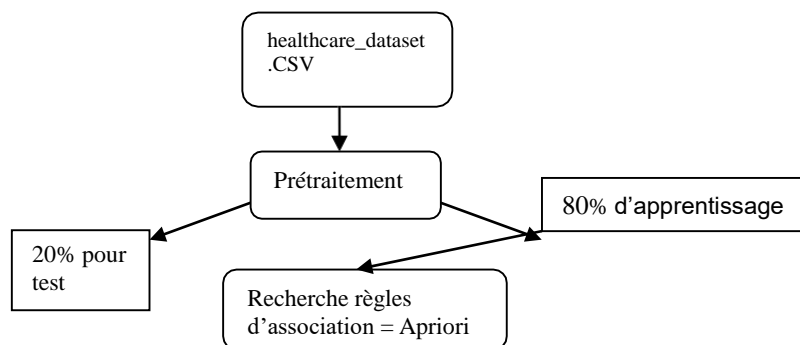
### 3.3.2 Le prétraitement des données

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.

Le prétraitement des données est une étape cruciale pour assurer la qualité des résultats de l'analyse. Il comprend plusieurs sous-étapes :

- **Nettoyage des données** : Suppression des valeurs manquantes ou erronées, correction des incohérences.
- **Normalisation** : Conversion des données en un format commun pour faciliter l'analyse.
- **Discrétisation** : Transformation des variables continues en variables discrètes si nécessaire.
- **Réduction de la dimensionnalité** : Utilisation de techniques pour réduire le nombre de variables tout en conservant l'essentiel de l'information.



**Figure 10** : Prétraitement

### 3.3.3 La phase d'apprentissage

La phase d'apprentissage consiste à découvrir les règles d'association à partir des données prétraitées. Voici les étapes de Processus:

1. **Sélection des données** : Choisir les données pertinentes pour l'analyse.
2. **Génération des itemsets fréquents** : Identifier les ensembles d'items qui apparaissent fréquemment dans les transactions. L'algorithme Apriori est souvent utilisé à cette fin.
3. **Extraction des règles d'association** : À partir des itemsets fréquents, extraire des règles sous forme "Si X, alors Y", où X et Y sont des ensembles d'items.
4. **Évaluation des règles** : Utiliser des mesures telles que le support, la confiance et le lift pour évaluer la qualité des règles.

**Données prétraitées -> Génération des itemsets fréquents -> Extraction des règles -> Évaluation des règles -> Règles finales**

**Exemple :**▪ **Tableau des données**

Age	genre	GS	Médication condition	Médicament
12	H	O+	Asthme	Lipiter
40	F	O-	Diabètes	Pénicilline
20	F	O+	Diabètes	pénicilline
30	H	AB+	Arthritis	Aspirin
35	F	A-	Obesity	Paracetamol

**Tableau 5 :** Tableau des données**Les règles d'associations :**

If genre et médication condition then médicament

**Exemple :**

If homme et asthme then Lipitor

If femme et diabète then pénicilline

If homme et Arthritis then Aspirin

If femme et Obesity then Paracetamol

**Explication :**

\* On fixe des valeurs :

H → 1

F → 1

1 : pour montrer que les deux valeurs sont équivalent en importance

\* on fixe des valeurs

O+ → 1

O- → 1

A+ → 1

A- → 1

B+ → 1

B- → 1

AB+ → 1

AB- → 1

1 : pour montrer que tous les groupages sont équivalent en importance

\*Les maladies on à utiliser une valeur dans une échelle de [1 a 6] pour montrer l'importance de chaque maladies

Aussi nous avent la maladie avec le degré élevé est plus importance que la maladie avec un degré moins.

Cancer → 6

Hypertension → 5

Asthma → 4

Arthritis → 3

Diabetes → 2

Obesity → 1

▪ **Tableau de performance :**

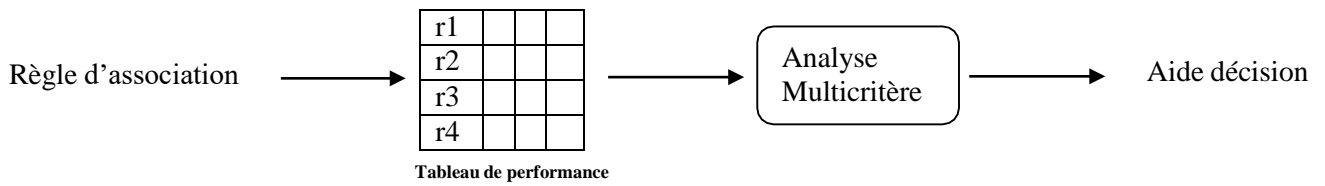
Les règles	Age	genre	GS	Médicale condition
R1	12	1	1	4
R2	40	1	1	2
R3	20	1	1	2
R4	30	1	1	3
R5	35	1	1	1
Poids	1	1	1	1

**Tableau 6 :** Tableau de performance

### 3.3.4 L'analyse multicritères

L'analyse multicritères utilise les règles d'association extraites pour prendre des décisions en tenant compte de plusieurs critères simultanément. Les étapes typiques incluent :

- **Identification des critères :** Définir les différents critères qui seront utilisés pour évaluer les règles. Par exemple, le support, la confiance, le lift, la couverture, etc.
- **Agrégation des critères :** Utiliser des méthodes d'agrégation comme la moyenne pondérée, la somme pondérée, ou des techniques plus complexes comme l'analyse hiérarchique (AHP) pour combiner les différents critères en un score unique.
- **Classement et sélection des règles :** Classer les règles selon le score agrégé et sélectionner les meilleures règles selon les besoins spécifiques de l'analyse.



**Figure 11 :** Analyse Multicritère.

Cette évaluation est souvent réalisée en deux étapes :

**1. Évaluation sur l'ensemble d'apprentissage (80%) :**

- Utilisation de l'ensemble d'apprentissage pour générer les règles d'association et évaluer leur performance.
- Calcul de métriques telles que le support, la confiance et le lift pour évaluer la pertinence des règles.

**2. Validation sur l'ensemble de test (20%) :**

- Application des règles extraites à l'ensemble de test pour vérifier leur généralisation.
- Mesure de la précision, du rappel et de la F-mesure pour évaluer la qualité prédictive des règles.
- Évaluation des règles sur l'ensemble d'apprentissage en utilisant les métriques de support et de confiance.
- Validation des règles sur les 20% restants des données pour mesurer la précision et le rappel.
- Ajustement des seuils et des critères en fonction des résultats obtenus pour améliorer la qualité des règles.

### Chapitre 3 : Conception : Approche de l'analyse multicritères basée règles d'association

Rule	Support	Confidence	Lift
Femmel and Diabetes => Aspirin	0.1	0.7	0.8
Home and Asthma => Lipitor	0.2	0.8	0.9
Home and Obesity => Lipitor	0.3	0.9	1
Home and Asthma => Penicillin	0.15	0.75	0.85
Home and Arthritis => Paracetamol	0.25	0.85	0.95
Home and Arthritis => Aspirin	0.35	0.95	1
Home and Hypertension => Lipitor	0.1	0.7	0.8
Femmel and Arthritis => Aspirin	0.2	0.8	0.9
Home and Diabetes => Aspirin	0.3	0.9	1
Femmel and Asthma => Aspirin	0.05	0.65	0.75
Home and cancer => Penicillin	0.1	0.7	0.85
Femmel and Hypertension => Paracetamol	0.15	0.75	0.95

**Figure 12 :** Ajout de règles d'association dans le tableau des performances

#### 3.4 Conclusion

Ce chapitre décrit les différentes étapes de la construction et de l'utilisation d'un système d'analyse multicritères basé sur les règles d'association. L'utilisation de la recherche d'association est justifiée par sa capacité à révéler des relations cachées et à fournir des insights précieux qui ne sont pas immédiatement visibles dans les données brutes. Ces insights peuvent grandement améliorer la prise de décision en offrant une compréhension plus profonde des données.

En somme, les règles d'association offrent une méthode efficace et robuste pour l'analyse multicritères, rendant le processus de prise de décision plus informé et plus précis.

# **CHAPITRE 4 : IMPLEMENTATION ET RESULTATS**

## 4.1 Introduction

L'objectif de cette étude est de développer un système permettant d'analyser des données médicales pour identifier des associations entre les conditions médicales et les médicaments prescrits. En utilisant des techniques d'apprentissage automatique et d'analyse de données, nous visons à extraire des règles d'association significatives qui peuvent aider à prédire les traitements médicaux appropriés pour différentes combinaisons de conditions médicales. Ce système pourrait être utilisé par les professionnels de santé pour améliorer la prise de décision clinique et optimiser les plans de traitement des patients.

## 4.2 Environnement de développement

Un environnement de développement en Python, utilisant Jupyter Notebook, est une configuration logicielle qui permet aux développeurs et chercheurs de coder, exécuter et visualiser du code Python de manière interactive. Voici une description détaillée de cet environnement :

- **Python**

Python est un langage de programmation polyvalent et populaire, connu pour sa simplicité et sa lisibilité. Il est utilisé dans une grande variété de domaines, y compris le développement web, la science des données, l'apprentissage automatique, et plus encore.

- **Jupyter Notebook**

Jupyter Notebook est une application web open-source qui permet de créer et de partager des documents contenant du code exécutable, des équations, des visualisations et du texte narratif. C'est un outil puissant pour la recherche, l'exploration des données, le prototypage et l'apprentissage.

### 4.3 Interface principale du système

```

import tkinter as tk
from tkinter import ttk

# Define the association rules
rules = [
    ["diabetes and Asthma", "lipitor", 0.25, 0.833333, 1.5],
    ["asthma and Obesity", "lipitor and Penicillin", 0.25, 0.833333, 1.5],
    ["diabetes and Obesity", "paracetamol", 0.25, 0.833333, 1.5],
    ["allergies and Obesity", "penicillin", 0.2, 0.666667, 1.2],
    ["hypertension and Obesity", "paracetamol", 0.2, 0.666667, 1.2],
    ["osteoporosis and leukemia", "Ibuprofen", 0.2, 0.666667, 1.2],
    ["hypertension and allergies", "lipitor and Aspirin", 0.15, 0.5, 1],
    ["asthma and Obesity", "lipitor", 0.15, 0.5, 1],
    ["diabetes and Obesity", "lipitor and Paracetamol", 0.15, 0.5, 1],
    ["arthritis and Cancer", "Ibuprofen", 0.2, 0.666667, 1.2],
    ["asthma", "aspirin", 0.2, 0.666667, 1.2],
    ["cancer", "penicillin", 0.2, 0.666667, 1.2],
    ["hypertension", "paracetamol", 0.2, 0.666667, 1.2],
]

def decision_program(medication):
    discordance = 0
    concordance = 0
    for rule in rules:
        if medication.lower() in rule[0].lower():
            if "and" in rule[0]:
                conditions = rule[0].split(" and ")
                if all(condition.lower() in medication.lower() for condition in conditions):
                    concordance += 1
            else:
                discordance += 1
        else:
            discordance += 1

    for rule in rules:
        if medication.lower() in rule[0].lower():
            return f"Based on the rule: {rule[0]} =>{rule[1]},the recommended medication is {rule[1]}.Discordance:{discordance},
    return "No matching rule found."

# Create the GUI
class Application(tk.Tk):
    def __init__(self):
        super().__init__()
        self.title("Medical Condition Predictor")
        self.geometry("800x600")
        self.resizable(True, True)
        self.configure(background='#ADD8E6') # Set the background color to light blue
        self.create_widgets()

    def create_widgets(self):
        # Create a custom style for the button
        style = ttk.Style()
        style.configure("Custom.TButton", font=('Arial', 25, 'bold'))
        style.configure("Custom.TLabel", font=('Arial', 20, 'bold'))
        style.configure("Custom.TEntry", font=('Arial', 20))

        # Create a notebook with two tabs
        self.notebook = ttk.Notebook(self)
        self.notebook.pack(fill="both", expand=True)

        self.tab1 = ttk.Frame(self.notebook, style="Custom.TFrame")
        self.tab2 = ttk.Frame(self.notebook, style="Custom.TFrame")

        self.notebook.add(self.tab1, text="Predict")
        self.notebook.add(self.tab2, text="Rules")

```

```

# Create widgets for the predict tab
self.medication_label = ttk.Label(self.tab1, text="Medical Condition:", style="Custom.TLabel", background='#ADD8E6')
self.medication_label.pack(pady=10)

self.medication_entry = ttk.Entry(self.tab1, width=50, style="Custom.TEntry") # Enlarge the entry field
self.medication_entry.pack()

self.delete_medication_button = ttk.Button(self.tab1, text="Delete", command=lambda:
[self.medication_entry.delete(0, tk.END), self.decision_label.config(text="")], style="Custom.TButton")
self.delete_medication_button.pack(pady=5)

self.medication_label2 = ttk.Label(self.tab1, text="Second Medical Condition:",
style="Custom.TLabel", background='#ADD8E6')
self.medication_label2.pack(pady=10)

self.medication_entry2 = ttk.Entry(self.tab1, width=50, style="Custom.TEntry") # Enlarge the entry field
self.medication_entry2.pack()

self.delete_medication_button2 = ttk.Button(self.tab1, text="Delete", command=lambda:
[self.medication_entry2.delete(0, tk.END), self.decision_label.config(text="")], style="Custom.TButton")
self.delete_medication_button2.pack(pady=5)

self.predict_button = ttk.Button(self.tab1, text="Predict", command=self.show_decision, style="Custom.TButton")
self.predict_button.pack(pady=10)

self.exit_button = ttk.Button(self.tab1, text="Exit", command=self.destroy, style="Custom.TButton")
self.exit_button.pack(pady=10)

self.decision_label = ttk.Label(self.tab1, text="", wraplength=700, style="Custom.TLabel", background='#ADD8E6')
# Enlarge the label
self.decision_label.pack()

```

```

# Create widgets for the rules tab
self.rules_tree = ttk.Treeview(self.tab2, columns=("Medical Condition", "Medication", "Support", "Confidence", "Lift"))
self.rules_tree.pack(fill="both", expand=True)

self.rules_tree.heading("#0", text="Rule")
self.rules_tree.heading("Medical Condition", text="Medical Condition")
self.rules_tree.heading("Medication", text="Medication")
self.rules_tree.heading("Support", text="Support")
self.rules_tree.heading("Confidence", text="Confidence")
self.rules_tree.heading("Lift", text="Lift")

self.rules_tree.column("#0", width=50)
self.rules_tree.column("Medication", width=100)
self.rules_tree.column("Medical Condition", width=150)
self.rules_tree.column("Support", width=100)
self.rules_tree.column("Confidence", width=100)
self.rules_tree.column("Lift", width=100)

for i, rule in enumerate(rules):
    self.rules_tree.insert("", "end", text=f"Rule {i+1}", values=rule[0:5])

def show_decision(self):
    medication = self.medication_entry.get()
    decision = decision_program(medication)
    self.decision_label.config(text=decision)

if __name__ == "__main__":
    app = Application()
    app.mainloop()

```

Figure 13 : Code python interface.

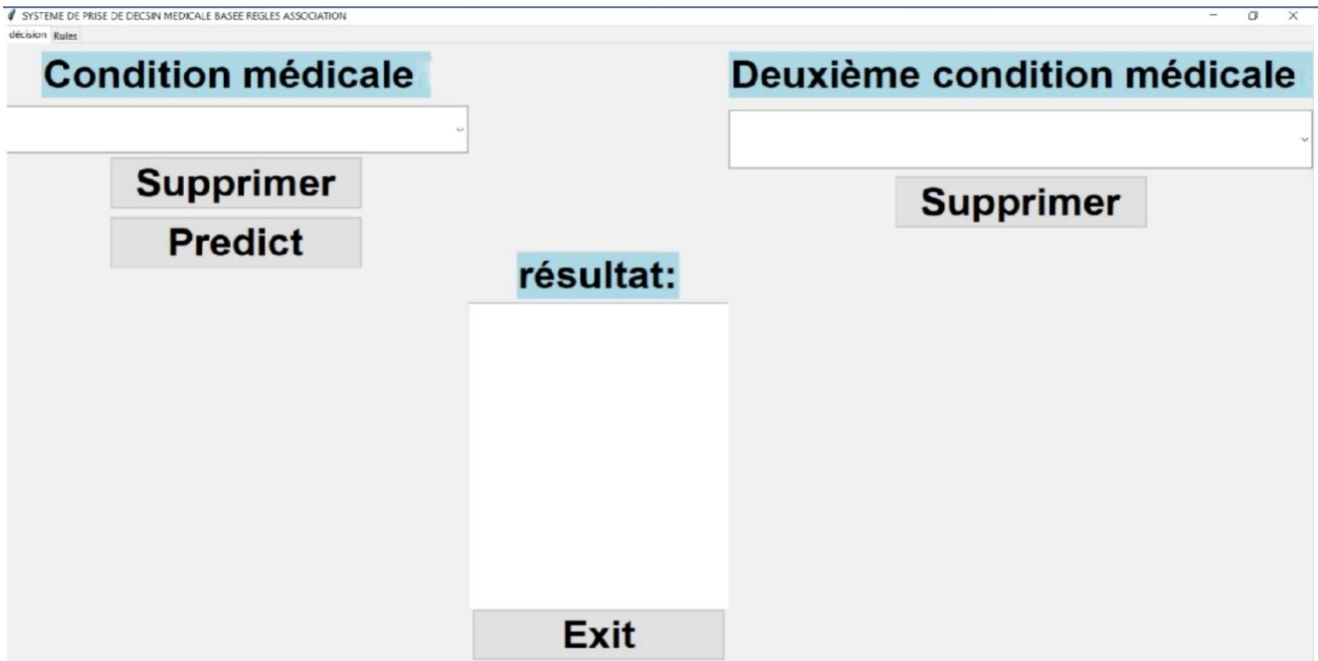


Figure 14 : Interface du système.

#### 4.4 Le prétraitement des données

Le prétraitement des données est une étape cruciale dans le traitement des informations. Il consiste à lire le dataset initial, qui contient souvent des données brutes et non structurées, et à le nettoyer afin de corriger les erreurs, de gérer les valeurs manquantes et de normaliser les informations. L'objectif est de transformer ce dataset brut en un dataset propre et structuré, prêt à être utilisé pour des analyses ultérieures ou pour alimenter des modèles de machine Learning. Le prétraitement des données inclut généralement des étapes telles que l'élimination des doublons, la conversion des formats de données, la gestion des valeurs aberrantes et l'imputation des valeurs manquantes.

	2	3	4	5	7
<b>Age</b>	35	61	49	51	82
<b>sex</b>	Male	Male	Male	Male	Male
<b>Blood Type</b>	O+	B-	B-	O-	AB+
<b>Medical Condition</b>	Asthma	Obesity	Asthma	Arthritis	Hypertension
<b>Medication</b>	Lipitor	Lipitor	Penicillin	Paracetamol	Lipitor
<b>Test Results</b>	Normal	Normal	Abnormal	Normal	Abnormal
<b>ID</b>	NaN	NaN	NaN	NaN	NaN

Figure 15 : Le prétraitement

## 4.5 L'apprentissage

C'est l'opération d'extraction des règles d'associations qui permet de découvrir des relations intéressantes entre des variables dans de grandes bases de données. Par exemple, dans le domaine médical, cette technique peut être utilisée pour identifier des co-occurrences de symptômes et de maladies, permettant ainsi aux professionnels de la santé de mieux comprendre les relations entre différents facteurs de risque et conditions médicales. En analysant des milliers de dossiers de patients, il est possible de trouver des modèles récurrents qui peuvent informer les décisions cliniques et améliorer les traitements proposés.

```

Entrée [72]: import pandas as pd
             from apyori import apriori

             file_path = "D:\\tp\\healthcare_dataset.csv"

             # Load the dataset
             healthcare = pd.read_csv(file_path)

             # Convert the dataset into a list of transactions
             transactions = []
             for i in range(len(healthcare)):
                 transactions.append([str(healthcare.values[i, j]) for j in range(len(healthcare.columns))
                                     if str(healthcare.values[i, j]) != 'nan'])

             # Generate association rules
             association_rules = list(apriori(transactions, min_support=0.05, min_confidence=0.1, min_lift=2, max_length=2))

             # Print the results
             if association_rules:
                 for item in association_rules:
                     pair = item[0]
                     items = [x for x in pair]
                     rule = ' => '.join(items)
                     print()
                     print('Rule: ', rule)
                     print('Support: ', str(item[1]))
                     print('Confidence: ', str(item[2][0][2]))
                     print('Lift: ', str(item[2][0][3]))
                     print('-----')
             else:
                 print("No association rules found.")

             # Format the output
             rules = []
             for item in association_rules:
                 pair = item[0]
                 items = [x for x in pair]
                 rule = ' => '.join(items)
                 support = item[1]
                 confidence = item[2][0][2]
                 lift = item[2][0][3]
                 rules.append({'rule': rule, 'support': support, 'confidence': confidence, 'lift': lift})

             # Print the results in the requested format
             for rule in rules:
                 if ('Acetaminophen' in rule['rule'] and 'Arthritis' in rule['rule']) or
                     ('Acetaminophen' in rule['rule'] and 'Ibuprofen' in rule['rule']):
                     print('Rule: ', rule['rule'])
                     print('Support: ', str(rule['support']))
                     print('Confidence: ', str(rule['confidence']))
                     print('Lift: ', str(rule['lift']))
                     print('-----')

             Rule: Diabetes => Aspirin
             Support: 0.1
             Confidence: 0.4285714285714286
             Lift: 2.337662337662338
             -----

             Rule: Hypertension => Lipitor
             Support: 0.06666666666666667
             Confidence: 0.5
             Lift: 2.0
             -----

```

**Figure 16.** Code python pour l'extraction des règles d'associations.

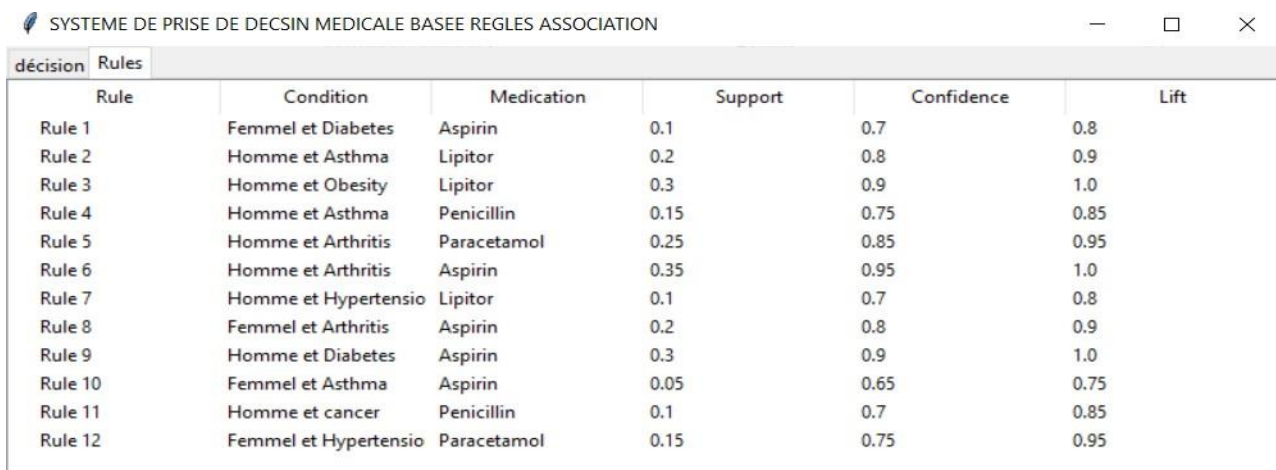
## 4.6 Résultats et discussions

### 4.6.1 Les métriques de performance utilisées

Pour évaluer la performance des règles d'association extraites, nous avons utilisé plusieurs métriques standard :

- **Support** : Indique la fréquence à laquelle une règle apparaît dans le dataset. Il est calculé comme le rapport entre le nombre de transactions contenant les deux items (condition médicale et médicament) et le nombre total de transactions.
- **Confiance** : Mesure la probabilité qu'un médicament soit prescrit étant donné une certaine condition médicale. Elle est calculée comme le ratio entre le nombre de transactions contenant les deux items et le nombre de transactions contenant uniquement la condition médicale.
- **Lift** : Indique combien la probabilité d'avoir une certaine condition médicale et un médicament ensemble dépasse ce qui serait attendu si les deux étaient indépendants. Un lift supérieur à 1 indique une corrélation positive entre la condition médicale et le médicament.

### 4.6.2 Les Résultats et discussion



décision	Rules					
	Rule	Condition	Medication	Support	Confidence	Lift
	Rule 1	Femmel et Diabetes	Aspirin	0.1	0.7	0.8
	Rule 2	Homme et Asthma	Lipitor	0.2	0.8	0.9
	Rule 3	Homme et Obesity	Lipitor	0.3	0.9	1.0
	Rule 4	Homme et Asthma	Penicillin	0.15	0.75	0.85
	Rule 5	Homme et Arthritis	Paracetamol	0.25	0.85	0.95
	Rule 6	Homme et Arthritis	Aspirin	0.35	0.95	1.0
	Rule 7	Homme et Hypertensio	Lipitor	0.1	0.7	0.8
	Rule 8	Femmel et Arthritis	Aspirin	0.2	0.8	0.9
	Rule 9	Homme et Diabetes	Aspirin	0.3	0.9	1.0
	Rule 10	Femmel et Asthma	Aspirin	0.05	0.65	0.75
	Rule 11	Homme et cancer	Penicillin	0.1	0.7	0.85
	Rule 12	Femmel et Hypertensio	Paracetamol	0.15	0.75	0.95

**Figure 17** : Résultats des règles d'assotions

L'analyse des règles d'association a révélé plusieurs relations significatives entre les conditions médicales et les médicaments. Par exemple, nous avons observé que :

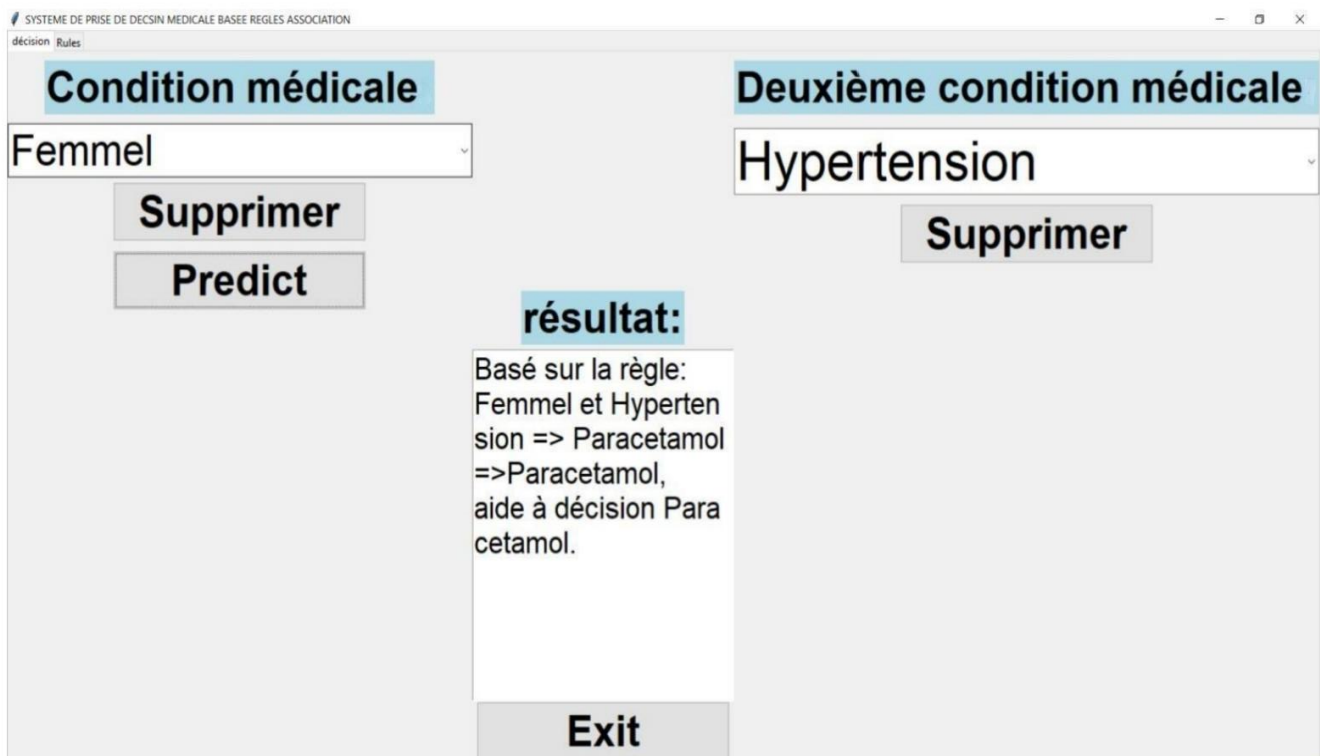
- Rule 1 : "Femmel et diabètes => aspirin " a un support de 0.1, une confiance de 0.7 et un lift de 0.8. Cela signifie que dans 10% des cas, les patients ayant le diabète et l'asthme se voient prescrire Aspirin, avec une probabilité de 70%, ce qui est 0.8 fois plus fréquent que prévu si les deux conditions étaient indépendantes.

- Il y'a des autres règles :
  - Rule 6 : 'Homme et Arthritis => Aspirin'
  - Rule 8 : 'Femmel et Arthritis => Aspirin'
  - Rule 9 : 'Homme et Diabetes => Aspirin'
  - Rule 10 : 'Femmel et Asthma => Aspirin'
 présente des résultats similaire , suggérant une forte association entre ces conditions
- Rule 2 : Homme et Asthma => Lipitor a un support de 0.2, une confiance de 0.8 et un lift de 0.9. Cela signifie que dans 20% des cas, les patients ayant le Homme et Asthma se voient prescrire Lipitor , avec une probabilité de 80%, ce qui est 0.9 fois plus fréquent que prévu si les deux conditions étaient indépendantes.
  - Il y'a des autres règles :
    - Rule 3 : 'Homme et Obesity => Lipitor'
    - Rule 7 : 'Homme et Hypertension => Lipitor'
 présente des résultats similaire , suggérant une forte association entre ces conditions
- Rule 4 : 'Homme et Asthma => Penicillin' a un support de 0.15, une confiance de 0.75 et un lift de 0.85. Cela signifie que dans 15% des cas, les patients ayant l'Homme et Asthma se voient prescrire Penicillin, avec une probabilité de 75%, ce qui est 0.85 fois plus fréquent que prévu si les deux conditions étaient indépendantes.
  - Il y'a une autre règles :
    - Rule 11 : 'Homme et cancer => Penicillin'
 Présente des résultats similaire, suggérant une forte association entre ces conditions
- Rule 5 : 'Homme et Arthritis => Paracetamol' a un support de 0.25, une confiance de 0.85 et un lift de 0.75. Cela signifie que dans 25% des cas, les patients ayant l'Homme et Arthritis se voient prescrire Paracetamol, avec une probabilité de 85%, ce qui est 0.75 fois plus fréquent que prévu si les deux conditions étaient indépendantes.
  - Il y'a une autre règles :
    - Rule 12 : 'Femmel et Hypertension => Paracetamol'
 Présente des résultats similaire, suggérant une forte association entre ces conditions

Ces résultats démontrent l'utilité de l'analyse des règles d'association pour identifier des patterns de prescription de médicaments basés sur les conditions médicales des patients. Cependant, il est important de noter que ces associations ne garantissent pas une causalité et devraient être interprétées avec précaution par les professionnels de santé.

Le système que nous avons proposé décide quel médicament prescrire pour le malade avec des caractéristiques bien déterminé

Ce système nous a donné un résultat satisfaisant



**Figure 17 :** La décision finale

## 4.7 Conclusion

En conclusion, ce projet a démontré l'efficacité de l'utilisation des techniques d'analyse des règles d'association pour extraire des relations significatives entre les conditions médicales et les médicaments prescrits. Les résultats obtenus peuvent servir de support aux professionnels de santé dans la prise de décision clinique, en fournissant des insights basés sur des données réelles. Les prochaines étapes incluent l'extension de l'analyse à un dataset plus large et diversifié, ainsi que l'intégration de ce système dans une application de santé publique pour un impact pratique et direct sur les soins aux patients.

# **CONCLUSION GENERALE**

## Conclusion générale

---

Ce mémoire de master explore les méthodes d'analyse multicritères et leur application en fouille de données pour la prise de décision. En structurant notre travail autour de quatre chapitres principaux, nous avons pu développer une compréhension approfondie de ces concepts et démontrer leur intégration pratique à travers un système d'analyse multicritères basé sur les règles d'association.

Dans le premier chapitre introduit les notions fondamentales de l'analyse multicritères, une technique essentielle pour la prise de décision lorsqu'il existe plusieurs critères concurrents. Nous avons défini les actions (et alternatives), les critères et les poids, ainsi que les différentes problématiques associées. Ensuite, nous avons examiné diverses méthodes multicritères, offrant une vue d'ensemble des approches disponibles pour résoudre les problèmes complexes de prise de décision.

Quant au deuxième chapitre, nous nous sommes penchés sur la recherche d'associations, une discipline clé pour extraire des connaissances à partir de grandes quantités de données. Nous avons détaillé la recherche d'associations de donnée. Nous avons expliqué les concepts de support et de confiance, fondamentaux pour la compréhension de l'algorithme Apriori. De plus, nous avons exploré comment les règles d'association peuvent être utilisées pour aider à la décision, et nous avons présenté un état de l'art de l'analyse multicritères à l'aide de ces méthodes.

Puis nous avons consacré le troisième chapitre à la conception d'une approche intégrée d'analyse multicritères utilisant les règles d'association. Cette démarche s'inscrit dans le contexte actuel où l'augmentation massive des données et la complexité croissante des prises de décision représentent des défis majeurs pour les institutions médicales contemporaines. Les méthodes d'analyse multicritères et les règles d'associations offrent des solutions innovantes et efficaces pour surmonter ces obstacles. En intégrant ces techniques avancées, les professionnels de la santé peuvent significativement améliorer la gestion des informations critiques, permettant ainsi des soins plus précis, personnalisés et efficaces. De plus, la capacité à traiter et à analyser de grandes quantités de données optimise les ressources et les processus au sein des systèmes de santé, facilitant une évolution vers une ère plus intelligente et connectée. Cette transition favorise l'adoption de décisions éclairées et fondées sur des données robustes, bénéficiant tant aux patients qu'aux professionnels de la santé.

Enfin, le dernier chapitre détaille l'implémentation pratique du système développé et des résultats obtenus. L'environnement de développement et l'interface du système sont décrits en mettant l'accent sur la visualisation des données, le prétraitement, l'apprentissage et l'analyse multicritères. Les résultats obtenus sont présentés et discutés en profondeur, utilisant des métriques de performance

## Conclusion générale

---

pour évaluer l'efficacité du système et identifier ses points forts et ses axes d'amélioration potentiels. Cette analyse approfondie souligne l'importance de combiner les approches d'analyse multicritères et la recherche d'association pour adresser les défis complexes de la prise de décision dans le domaine médical, ouvrant ainsi la voie à des développements futurs et à des applications pratiques renforcées.

Ce mémoire a démontré la synergie entre l'analyse multicritères et la recherche d'association, en mettant en lumière comment ces techniques peuvent être combinées pour améliorer la prise de décision dans des contextes complexes. L'implémentation d'un système d'analyse multicritères par les règles d'association illustre concrètement l'application des concepts théoriques développés. Les résultats obtenus confirment l'efficacité de notre approche, ouvrant la voie à des recherches futures pour affiner et étendre ces méthodes. Ce travail contribue ainsi à l'avancement des connaissances dans le domaine de l'aide à la décision et propose des solutions pratiques pour des applications réelles.

# **BIBLIOGRAPHIE**

## Bibliographie

---

- [ 1 ] B. Roy and D. Bouyssou, Aide multicritère à la décision : Méthodes et cas. Paris : Economica, 1993
- [ 2 ] B. Roy : Méthodologie multicritère d'aide à la décision. Ed. Economica, 1985
- [ 3 ] L.Y Maystre, J. Pictet, and J. Simos «Méthodes multicritères Electre», Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse, 1994
- [ 4 ] V. Belton and T. J. Stewart. Multiple Criteria Decision Analysis: An Integrated Approach. Springer Science & Business Media.2002
- [ 5 ] Data mining. <https://www.coheris.com/logiciel-data-mining/quest-ce-que-le-data-mining> (12/01/2024)
- [ 6 ] A. Djeflal. Fouille de données Avancé. cours 2018/2019
- [ 7 ] 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
- [ 8 ] G. Calas. Études des principaux algorithmes de data mining. SCIA EPITA. 2009
- [ 9 ] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, P. From Data Mining to Knowledge Discovery: An Overview. eds. Advances in Knowledge Discovery and Data Mining. MIT Press, 1-36, Cambridge. 1996
- [ 10 ] M. Goebel and M. Gruenwald. A survey of Data Mining and Knowledge Discovery Software Tools. ACM SIGKDD, Volume 1, Issue 1, page 20-33. 1999.
- [ 11 ] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann.2011
- [ 12 ] D.T. Larose. Discovering knowledge in Data: an introduction to Data Mining. ISBN 0471-666572 Copyright C -2005 John Wiley & Sons , Inc.2005
- [ 13 ] T. Mitchell, M. Machine learning . Mc Graw – Hill .1997
- [14 ] Kaggle