

# REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 20 Août 1955- SKIKDA

Faculté des Sciences. Département Informatique



Mémoire présenté en vue de l'obtention du diplôme de

Master en Informatique.

Option : Système Informatique.

Thème

## **PREDICTION BASEE REGLES D'ASSOCIATIONS POUR L'EVALUATION DU NIVEAU DE STRESS CHEZ LES ÉTUDIANTS.**

**Réalisé par :**

BASSOUT Daya Eddine

TAMINE Abdelghafour

**Encadreur.**

Mr. MANSOUL A.

Année universitaire 2023/2024

## Table des matières

Remerciements .....	5
Dédicaces .....	6
Résumé.....	8
<b>Introduction Générale .....</b>	<b>9</b>
• <b>Problématique .....</b>	<b>10</b>
• <b>Objectifs.....</b>	<b>10</b>
• <b>L'organisation du mémoire.....</b>	<b>11</b>
<b>Chapitre 01 : La prédiction .....</b>	<b>13</b>
<b>1 Introduction.....</b>	<b>14</b>
<b>2 La prédiction .....</b>	<b>14</b>
<b>2.1 Fonctionnement.....</b>	<b>14</b>
<b>3 Les méthodes de prédiction.....</b>	<b>15</b>
<b>3.1 Les règles d'association .....</b>	<b>16</b>
<b>3.2 Les arbres de décision .....</b>	<b>16</b>
<b>3.3 Les réseaux neuronaux.....</b>	<b>17</b>
<b>4 Les systèmes de prédiction.....</b>	<b>17</b>
<b>5 État de l'art de la prédiction par les méthodes de FFD.....</b>	<b>18</b>
<b>5.1 La prédiction par les règles d'association .....</b>	<b>18</b>
<b>5.2 La prédiction par les réseaux de neurones .....</b>	<b>18</b>
<b>5.3 La prédiction par les arbres de décision .....</b>	<b>19</b>
<b>6 Conclusion .....</b>	<b>19</b>
<b>Chapitre 02 : La recherche d'association.....</b>	<b>20</b>
<b>1. Introduction.....</b>	<b>21</b>
<b>2. Recherche de règles d'association .....</b>	<b>21</b>
<b>2.1 Extraction des règles d'association (Support, Confiance, Lift) .....</b>	<b>21</b>
<b>2.1.1 Support d'une règle d'association .....</b>	<b>21</b>
<b>2.1.2 Confiance d'une règle d'association.....</b>	<b>22</b>
<b>2.1.3 Lift d'une règle d'association.....</b>	<b>22</b>
<b>2.2 Algorithme Apriori.....</b>	<b>22</b>
<b>2.2.1 Avantages et inconvénients de l'Algorithme Apriori.....</b>	<b>23</b>
<b>3 Conclusion .....</b>	<b>23</b>

<b>Chapitre 3 : Approche de la prédiction par les règles d'association.....</b>	<b>25</b>
1 Introduction.....	26
2 La maladie du stress .....	26
3 Les systèmes des prédictions.....	27
4 Les données utilisées .....	28
5 Le prétraitement des données.....	29
6 L'apprentissage du modèle .....	30
7 La prédiction .....	33
8 Conclusion .....	34
<b>Chapitre 4 : Implémentation et résultats.....</b>	<b>36</b>
1 Introduction.....	37
2 Environnement de développement matérielle et logicielle.....	37
3 Implémentation .....	39
3.1 Le prétraitement .....	40
3.2 L'Apprentissage.....	40
3.3 La prédiction .....	42
4 Expérimentation et discussion des résultats .....	42
4.1 Les métrique de performance utilisées (conf,Lift).....	42
4.2 Discussion des résultats .....	43
5 Conclusion .....	46
Conclusion générale.....	47
Bibliographie.....	49

## Liste des figures

Figure 1 Exemple de model d'arbre de décision .....	17
Figure 2 les données de la maladie traité.....	28
Figure3 schéma d'apprentissage du modèle.....	32
Figure4 Schéma de La prédiction du modèle .....	34
Figure 5 pc acer aspire 5.....	37
Figure 6 programmes de division des données.....	40
Figure 7 programme d'extraction des règle d'associons .....	41
Figure 8 programme d'extraction des règle d'associons de plus grand Lift .....	41
Figure 9 programmes de la prédiction du modèle .....	42
Figure 10 l'interface du modèle.....	43
Figure 11 exemples de niveau de stress 0.....	44
Figure 12 exemples de niveau de stress 1.....	44
Figure 13 Exemple de niveau de stress 2.....	45

## Remerciements

C'est avec un immense plaisir que je réserve ces quelques lignes en signe de gratitude et de reconnaissance à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce travail.

nous souhaite adresser, en premier lieu, mes remerciements les plus sincères à mon encadrant Dr .Mansoul.A pour sa disponibilité, sa patience et son précieux.

Suivi tout au long de la réalisation de ce travail.

nous tiens également à remercier les membres du jury d'avoir consacré une partie de leur temps à la lecture de ce mémoire et pour l'intérêt qu'ils ont porté à ce travail.

nous remercie en toutes les personnes qui ont contribué de près ou de loin à l'accomplissement de ce travail.

## Dédicaces

Je dédié ce travail

A ma mère, ma raison d'être, la lanterne qui éclaire mon chemin et m'illumine

D'affection et d'amour.

A mon père, en signe d'amour, et de gratitude pour tous les soutiens et les sacrifices

Dont il a fait preuve à mon égard

A mon frère islam et Aboubacar Sadik ma sœurs Anfel et ma petite sœur Rtaj.

a la reine de mon cœur ma muse pour toujours je dédie mon travail à vos jolies yeux

A mon binôme Abdelghafour,

A mes amis (es) Younes, Hachem, Raid. Nadji, Chouaib, Oussama, Badro, Raouf Aucun mot ne pourra décrire vos dévouements.

A Mr. Mansoul .A pour l'aide et le temps qu'il m'a consacré. A tous

Mes proches et mes amis, Je vous dédié ce travail en vous souhaitant un avenir éclatant

Et plein de bonnes promesses.

A tous les gens qui ont cru en moi et qui me donnent l'envie d'aller en avant, je vous

Remercie tous, vos encouragements et votre soutien me donnent la force de continuer.

**Bassout Daya Eddin**

## Dédicaces

J'ai l'honneur de dédier ce modeste travail à ceux qui m'ont encouragé et soutenu.

Ma très chère mère, qui me donne toujours l'espoir de vivre.

Mon très cher père qui était toujours à mes côtés à tous moments.

A mon binôme Daya Eddine,

A mes chères amies.

Aucun mot ne pourra décrire vos dévouements.

A tous mes enseignants durant mes années d'études avec lesquels j'ai beaucoup appris.

**Abdelghafour Tamine**

## Résumé

Avec la prolifération des maladies et l'émergence du stress, la prédiction des maladies a commencé à émerger comme un problème de recherche intéressant. Les méthodes de prévision reposent sur des méthodes d'exploration de données. Dans ce travail, nous présentons la conception d'un modèle de prédiction du niveau de stress chez les Person. Ce projet de recherche utilise des outils de traitement et d'analyse de données pour prédire les niveaux de stress. Nous avons utilisé une procédure de collecte et de prétraitement des données concernons des Person avec leur niveau de stress, en suit un apprentissage du modèle de la prédiction du stress a été réalisé .le modèle réalisé est à base de règle d'association.pour réaliser ce travaille nous avons utilisé plusieurs outille et bibliothèque de python. . Ces outils ont permis de développer un modèle prédictif efficace pour prédire les niveaux de stress. Ce travail démontre le potentiel de l'exploration de données pour améliorer la gestion du stress et offre des perspectives prometteuses pour des extensions futures.

**Mots clé :** prédiction, niveau de stress, règle d'association, Apriori, santé

## Abstract

With the proliferation of diseases and the emergence of stress, disease prediction has begun to emerge as an interesting research problem. Prediction methods rely on data mining techniques. In this work, we present the design of a model for predicting stress levels in individuals. This research project uses data processing and analysis tools to predict stress levels. We used a procedure for collecting and preprocessing data concerning individuals and their stress levels, followed by learning a stress prediction model. The model developed is based on association rules. To carry out this work, we used several Python tools and libraries. These tools allowed us to develop an effective predictive model for predicting stress levels. This work demonstrates the potential of data mining to improve stress management and offers promising perspectives for future extensions.

**Keywords:** prediction, stress level, association rules, Apriori, health

# Introduction générale

## **Introduction générale**

Le stress chez les étudiants est un problème croissant dans l'enseignement supérieur, affectant leur santé et leurs résultats académiques. Cette étude explore l'utilisation de la prédiction basée sur les règles d'associations comme outil innovant pour évaluer le niveau de stress des étudiants. En combinant l'analyse de motifs récurrents avec des techniques prédictives, cette approche vise à identifier précocement les facteurs de stress, permettant ainsi des interventions ciblées et efficaces pour améliorer le bien-être étudiant.

- **Problématique**

Le stress est une réaction naturelle de l'organisme face à des situations perçues comme menaçantes ou exigeantes. Cependant, un stress chronique non pris en charge peut avoir des répercussions néfastes sur la santé physique et mentale. Dans le contexte actuel où les exigences professionnelles et personnelles ne cessent d'augmenter, il devient crucial de trouver des moyens efficaces pour identifier et gérer le stress. Traditionnellement, l'évaluation des niveaux de stress repose sur des questionnaires auto-administrés ou des entretiens avec des professionnels de santé, des méthodes qui peuvent être subjectives et limitées par des biais de perception et de déclaration.

Avec l'essor des technologies numériques et l'accumulation massive de données, l'exploration de données (ou data mining) offre une opportunité inédite pour étudier et prédire les niveaux de stress de manière plus objective et précise. Cependant, plusieurs défis se posent : Comment exploiter efficacement les données disponibles pour prédire les niveaux de stress ? Quels algorithmes et techniques d'exploration de données sont les plus appropriés pour cette tâche ? Et comment garantir la confidentialité et l'éthique dans l'utilisation de ces données sensibles liées au stress des individus ?

- **Objectifs**

L'objectif principal de cette recherche est d'explorer l'utilisation de l'exploration de données pour prédire les niveaux de stress des individus. Pour atteindre cet objectif global, les objectifs spécifiques suivants seront poursuivis :

1. Identifier les sources de données pertinentes : Recenser et évaluer les types de données (physiologiques, comportementales, environnementales, etc.) pouvant être collectées et utilisées pour la prédiction du stress.

2. Développer un modèle prédictif : Appliquer des techniques d'exploration de données pour développer un modèle capable de prédire les niveaux de stress à partir des données recueillies, en explorant différents algorithmes d'apprentissage automatique et en comparant leur précision et leur fiabilité.

3. Évaluer l'efficacité du modèle : Tester et valider le modèle prédictif sur un ensemble de données indépendant afin d'évaluer ses performances et d'identifier ses éventuelles limites.

4. Proposer des applications pratiques : Identifier des applications concrètes du modèle prédictif dans divers contextes tels que la santé mentale, le milieu professionnel et les environnements scolaires.

En combinant l'analyse de données et les techniques d'exploration de données, cette recherche vise à offrir une nouvelle perspective sur la gestion du stress et à contribuer au développement d'interventions préventives et personnalisées.

## • L'organisation du mémoire

Ce document est divisé en quatre chapitres organisés comme suit :

### ✚ Introduction générale

Cette section introduit le sujet principal du document, à savoir l'utilisation de la fouille de données pour prédire le niveau de stress. Elle établit l'importance de la prédiction du stress et présente un aperçu des approches et méthodes qui seront discutées.

### ✚ Chapitre 1 : La prédiction

Ce chapitre introduit le concept de prédiction et ses applications dans divers domaines. Il présente les principales méthodes de prédiction, notamment les règles d'association, les arbres de décision et les réseaux neuronaux. Une attention particulière est portée aux systèmes de prédiction et à leur fonctionnement. Le chapitre se conclut par un état de l'art des différentes approches de prédiction, offrant ainsi une base solide pour comprendre les chapitres suivants.

### ✚ Chapitre 2 : La recherche d'association

Ce chapitre se concentre sur la recherche de règles d'association, une technique fondamentale en data mining. Il définit les concepts clés tels que le support, la confiance et le lift, essentiels pour comprendre et évaluer les règles d'association. L'algorithme Apriori, largement utilisé pour l'extraction de ces règles, est expliqué en détail, y compris ses avantages

et ses limites. Ce chapitre pose les bases théoriques nécessaires pour l'approche de prédiction développée dans la suite du mémoire.

### **Chapitre 3 : Approche de la prédiction par les règles d'association**

Dans ce chapitre, nous appliquons les concepts précédents à un cas concret : la prédiction de la maladie du stress. Il décrit le système de prédiction basé sur les règles d'association, en détaillant chaque étape du processus : la collecte et le prétraitement des données, l'apprentissage du modèle, et la phase de prédiction. Ce chapitre établit le lien entre la théorie et la pratique, préparant ainsi le terrain pour l'implémentation et l'analyse des résultats.

### **Chapitre 4 : Implémentation et Résultats**

Le dernier chapitre présente la mise en œuvre concrète de l'approche développée. Il décrit l'environnement de développement utilisé et détaille l'implémentation des différentes étapes du système de prédiction. Les expérimentations menées sont présentées, suivies d'une discussion approfondie des résultats obtenus. L'analyse des performances, basée sur les métriques de confiance et de lift, permet d'évaluer l'efficacité de l'approche proposée et d'identifier d'éventuelles pistes d'amélioration.

### **Conclusion générale**

Cette section synthétise les principaux résultats et conclusions du document, offrant une réflexion sur les implications des travaux réalisés. Elle propose également des perspectives pour des recherches futures et des améliorations potentielles des méthodes de prédiction du stress.

# Chapitre 01 : La prédiction

# 1 Introduction

La prédiction joue un rôle crucial en informatique et en science des données, en permettant d'anticiper les événements à venir en se basant sur les données disponibles. La fouille de données, ou data mining, constitue l'un des outils principaux utilisés dans ce domaine, en permettant d'extraire des modèles et des relations à partir des données pour les analyser et les utiliser dans le processus de prédiction.

Diverses méthodes et techniques sont utilisées dans la fouille de données pour la prédiction, notamment les "règles d'association", qui cherchent à identifier des relations entre des ensembles d'éléments dans les ensembles de données. Ces relations reposent sur des règles spécifiques qui peuvent être utilisées pour prédire des événements futurs.

## 2 La prédiction

- **Définition**

La prédiction est l'acte préventif ou anticipateur d'un événement, d'une tendance ou d'un résultat futur basé sur des données réelles, des modèles théoriques ou des observations passées. Les prédictions sont disponibles dans divers domaines tels que la médecine, la finance, l'astrologie, l'intelligence artificielle et autres. Il est possible d'utiliser des techniques statistiques, des modèles mathématiques, des algorithmes d'application automatisés ou des méthodes analytiques pour utiliser la même méthode qui est la plus vulnérable à ce produit.

### 2.1 Fonctionnement

Les data scientists utilisent des modèles prédictifs pour identifier des corrélations entre différents éléments dans des ensembles de données sélectionnés. Une fois la collecte de données terminée, un modèle statistique est formulé, entraîné et modifié pour générer des prédictions.

Le workflow pour la création des Framework d'analyse prédictive comporte cinq étapes de base:

- **Définir le problème** : une prédiction commence par une bonne problématique et un ensemble d'exigences. Par exemple, un modèle d'analyse prédictive peut-il détecter une fraude ? Déterminer les niveaux d'inventaire optimaux pour la période des fêtes de fin d'année ? Identifier les niveaux d'inondations potentiels en cas d'intempéries ? Un problème à résoudre clairement établi aidera à déterminer la méthode d'analyse prédictive à utiliser.
- **Collecter et organiser les données** : une entreprise peut avoir des décennies de données à exploiter, ou un afflux continu de données provenant des interactions client. Avant de pouvoir développer des modèles d'analyse prédictive, vous devez identifier les flux de données, puis organiser les ensembles de données dans un dépôt, par exemple un entrepôt de données comme BigQuery.
- **Prétraiter les données** : les données brutes n'ont qu'une utilité nominale en soi. Pour préparer les données qui seront utilisées dans les modèles d'analyse prédictive, elles doivent auparavant être nettoyées en vue de supprimer les anomalies, les points de données manquants ou les valeurs aberrantes qui peuvent être le résultat d'erreurs de saisie ou de mesure.
- **Développer des modèles prédictifs** : les data scientists disposent de divers outils et techniques pour développer des modèles prédictifs, en fonction du problème à résoudre et de la nature de l'ensemble de données. Le machine Learning, les modèles de régression et les arbres de décision comptent parmi les types de modèles prédictifs les plus courants.

**Valider et déployer les résultats** : vérifiez la justesse du modèle et ajustez-le en conséquence. Après obtention de résultats acceptables, mettez-les à la disposition des data dashboard via une application, un site web ou un tableau de bord de données. [1]

## 2 Les méthodes de prédiction

La prédiction en data Mining peut traiter soit des valeurs continues soit des valeurs discrètes, et elle est souvent approximative, associée à une probabilité ou à un ensemble de valeurs. Récemment, une méthode de prédiction a été proposée pour traiter indépendamment les cas continu et discret en se basant sur les règles d'association. Cette méthode permet de prédire, avec une confiance toujours égale à 1, soit un ensemble de valeurs dans le cas discret, soit un intervalle de valeurs dans le cas continu. Pour obtenir de telles règles, cette méthode

suppose que l'on dispose d'une hiérarchie décrivant des concepts généralisant les valeurs qui peuvent être prédites.

Les règles d'association ont attiré l'attention des chercheurs car elles permettent la découverte de règles intelligibles et exploitables dans un ensemble de données, tout en exprimant des associations entre les données. Cette technique est largement adoptée par les chercheurs pour effectuer leurs travaux en fouille de données, car elle permet la découverte de relations significatives entre les attributs des bases de données.

La recherche des règles d'associations est appliquée dans tous les domaines du Data Mining. Les règles d'association sont extraites des bases de données en appliquant des algorithmes d'extraction de motifs fréquents, tels que l'algorithme Apriori. Cette approche est utilisée dans le marketing et la planification commerciale, les réseaux de télécommunication, la recherche médicale et sur Internet.

### **3.1 Les règles d'association**

Une règle d'association est une application de la forme  $X \rightarrow Y$  dans laquelle X et Y sont des ensembles de terme.

Il existe deux mesures importantes, le support et la confiance, la robustesse d'une règle d'association est déterminée grâce à ces deux métriques [2]. Une règle d'association qui a un support faible va être observée rarement. La confiance mesure la pertinence de l'inférence dans une règle, par exemple plus grande est la mesure de confiance de la règle  $X \rightarrow Y$ , plus cette règle sera pertinente.

### **3.2 Les arbres de décision**

Dans l'analyse prédictive, les arbres de décision sont un type de modèle prédictif. En raison de sa nature extrêmement simple, elle est catégorisée parmi les techniques d'apprentissage automatique en boîte blanche. Un arbre de décision permet aux utilisateurs de comprendre clairement la relation entre les entrées de données et les sorties. Lorsqu'on combine divers modèles d'arbre de décision, on obtient des modèles prédictifs d'analyse appelés forêts aléatoires. Même si les forêts aléatoires sont en général plus précises que les arbres de décisions seuls, il n'est pas toujours aisé de comprendre leurs sorties en fonction de leurs entrées.

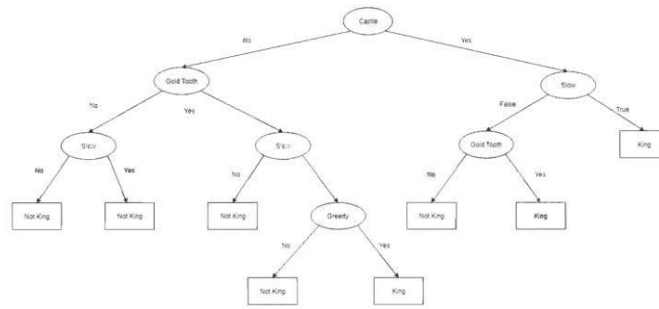


Figure 1 Exemple de model d'arbre de decision

[3]

### 3.3 Les réseaux de neuronaux

Avec le développement des puissances de calcul, les réseaux neuronaux artificiels sont devenus des outils parmi les plus répandus et évolués du Data mining. C'est une technique issue de la biologie. En effet, un réseau de neurones artificiel est un modèle de calcul dont la conception est très schématiquement inspirée des interconnexions du cerveau humain, ce qui permet au réseau de prendre des décisions s'appuyant davantage sur la perception que sur le raisonnement logique formel. De plus, les modèles dont ils se basent sont généralement optimisés par des méthodes d'apprentissage statistique, telles que les probabilités et les distributions, leur permettant de générer de vastes espaces fonctionnels souples.

Un réseau de neurones artificiel possède généralement d'un algorithme d'apprentissage, qui consiste à auto-modifier des paramètres réels en entrée du réseau, appelés coefficients ou poids synaptiques en fonction d'un jeu de données exemple présenté en entrée. Le but de cet entraînement est de permettre au réseau, d'apprendre à partir d'exemples d'apprentissage. Si l'entraînement est correctement réalisé, le réseau est capable de généraliser à partir de ce jeu, ce qui représente tout l'intérêt des réseaux de neuronaux artificiels.[4]

### 3 Les systèmes de prédiction

La prédiction est une tâche très attrayante à de nombreuses applications. Cette fonction peut être comparée à la classification ou à l'estimation à la différence que ses observations sont classées ou estimées sur la base de la prévision d'un comportement futur. En d'autres termes, les valeurs connues sont historiées et on cherche à prédire la valeur future d'un champ. On comprend facilement, que la seule manière de mesurer la qualité d'une prédiction est d'attendre! Des exemples d'applications de la prédiction sont :

- Economie : la prédiction de valeurs futures d'actions et d'indicateurs boursiers, la prédiction des prix futurs de matières brutes (pétrole, gaz, or, blé, ...etc.).

- Marketing : la prédiction au vu de leurs comportements passés des départs de clients.
- Industrie : la prédiction de pannes futures dans équipements industriels. [5]
- Diagnostic médical : la prédiction de l'efficacité d'un traitement médical.
- Météorologie : les prédictions météorologiques.

## **4 État de l'art de la prédiction par les méthode de FDD**

L'état de l'art de la prédiction implique d'analyser les événements passés pour anticiper les résultats futurs avec une plus grande précision. Ce domaine a connu une évolution significative avec les progrès en sciences des données, en apprentissage automatique et en intelligence artificielle. Voici un aperçu de quelques développements récents dans ce domaine:

### **5.1 La prédiction par les règle d'association**

- **Tao-Yuan Jen, Dominique Laurent, Gorgoumack Sambe**

Cette étude propose une méthode améliorée pour prédire les valeurs manquantes dans les entrepôts de données. Basée sur les règles d'association et utilisant une hiérarchie de concepts, elle génère des prédictions uniques avec une confiance maximale, contrairement aux approches précédentes qui fournissaient des ensembles ou intervalles de valeurs. Cette méthode s'applique aux données continues et discrètes, offrant une solution plus précise et fiable pour ce problème crucial en gestion de données.[6]

- **Lallich, Stéphane, and O.Teytaud**

L'article traite de l'amélioration de la recherche de règles d'association dans les données. Il examine les limites des méthodes actuelles, propose de nouvelles mesures d'évaluation, et suggère une approche statistique pour valider simultanément de nombreuses règles et mesures.

### **5.2 La prédiction par les réseaux de neurones**

- **Alexandre Aussem**

Cette thèse explore les réseaux de neurones dynamiques et récurrents (DRNN) avec des connexions modélisées par filtres autorégressifs. Elle présente un algorithme d'apprentissage efficace et démontre son potentiel pour la prédiction de séries temporelles complexes. L'étude unifie théoriquement prédiction, modélisation et contrôle adaptatif, et applique le DRNN à des

problèmes variés, de l'astrophysique à la gestion de réseaux ATM, démontrant sa polyvalence et son efficacité.[7]

- **Salhi, Chahrazed, Bénina Touaibia, and Z.Ayoub**

L'étude développe des modèles prédictifs de l'érosion des bassins versants en Algérie, utilisant des paramètres environnementaux. Le modèle de réseau neuronal, avec quatre variables d'entrée, s'est avéré le plus efficace, offrant des résultats précis.

### **5.3 La prédiction par les arbres de décision**

- **N. Sutton-Charani**

Cette étude améliore les prédictions probabilistes des arbres de décision, surtout pour les petits échantillons, en utilisant une approche bayésienne empirique. Elle corrige les probabilités des feuilles peu peuplées, génère des fonctions de croyance basées sur ces corrections, et évalue les résultats avec des indices adaptés aux prédictions incertaines. Cette méthode augmente significativement la fiabilité et la performance des prédictions, particulièrement pour les ensembles de données limités. [8]

- **Adam Salvail-Bérard**

L'article explore l'hybridation des algorithmes de forage de données, en se concentrant sur les arbres de décision. Il présente les concepts de base, les méthodes d'hybridation, et un exemple d'arbre de décision hybride amélioré.

## **5 Conclusion**

La prédiction est un outil puissant qui peut apporter de nombreux avantages, mais qui nécessite également une approche réfléchie et une compréhension approfondie des données et des méthodes utilisées. En intégrant la prédiction dans les processus décisionnels, il est possible d'améliorer la prise de décision et d'obtenir un avantage concurrentiel dans un large éventail de domaines. Malgré ces avantages, il est important de noter les limites et les défis de la prédiction. Les modèles de prédiction reposent sur des données historiques et peuvent ne pas toujours être précis dans des situations nouvelles ou inattendues. De plus, la qualité des prédictions dépend largement de la qualité des données utilisées, soulignant ainsi l'importance d'une collecte et d'une analyse de données rigoureuses.

# **Chapitre 02 : La recherche d'association**

# 1. Introduction

La recherche d'association est un domaine clé dans l'analyse des données, utilisé pour découvrir des relations ou des associations intéressantes et utiles entre des variables dans de grands ensembles de données. Elle est largement employée dans divers secteurs, tels que le marketing, la médecine, la biologie, et bien d'autres.

## 2. Recherche de règles d'association

Dans la prédiction des règles d'association, les items et itemsets jouent un rôle crucial. Un item représente un élément unique dans une transaction, tandis qu'un itemset est un ensemble d'items. L'analyse des itemsets fréquents permet de découvrir des associations significatives entre les items. Ces associations sont ensuite utilisées pour générer des règles d'association, qui expriment la probabilité qu'un item ou un ensemble d'items apparaisse dans une transaction, étant donné la présence d'autres items. Par exemple, dans l'analyse du panier de la ménagère, une règle pourrait indiquer que les clients achetant du pain et du lait sont susceptibles d'acheter également des œufs. Ces règles sont précieuses pour la prise de décision.

- **Définition**

On peut définir une règle d'association comme une règle de la forme dans laquelle B et H sont des ensembles de terme. Une telle règle peut être interprétée de la façon suivante : “Les documents qui possèdent les termes de B possèdent également les termes de H”. B et H sont appelés des motifs (ou itemsets) [9]

### 2.1 Extraction des règles d'association (Support, Confiance, Lift)

#### 2.1.1 Support d'une règle d'association

Le support d'une règle d'association  $A \rightarrow C$  définit le pourcentage de transaction qui contiennent A et C. Il représente le nombre de transaction qui contiennent A et C support(A  $\cup$  C) divisé par le nombre total des transaction:

$$\text{support}(A \rightarrow C) = \frac{\text{support}(A \cup C)}{|D|} . \text{support}(A \rightarrow C) \in [0,1] .$$

### 2.1.2 Confiance d'une règle d'association

La confiance est une mesure permettant d'évaluer la validité d'une règle d'association. La confiance d'une règle d'association  $A \rightarrow C$ , notée  $\text{conf}(A \rightarrow C)$  représente la proportion de documents qui contient A et qui contient aussi C. Elle est définie comme suit

$$\text{conf}(A \rightarrow C) = \frac{\text{support}(A \cup C)}{\text{support}(A)}, \text{conf}(A \rightarrow C) \in [0,1].$$

### 2.1.3 Lift d'une règle d'association

Lift ( $A \rightarrow B$ ) fait référence à l'augmentation du ratio de vente de B lorsque A est vendu.

Lift ( $A \rightarrow B$ ) peut être calculé en divisant la confiance ( $A \rightarrow B$ ) divisée par le support(B).

Mathématiquement, cela peut être représenté comme suit :

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

Pour en revenir à notre problème Burger et Ketchup, le Lift ( $\text{Burger} \rightarrow \text{Ketchup}$ ) peut être calculé comme suit :

$$\text{Lift}(\text{Burger} \rightarrow \text{Ketchup}) = (\text{Confiance}$$

$$(\text{Burger} \rightarrow \text{Ketchup}) / (\text{Support}(\text{Ketchup})) \text{Lift}(\text{Burger} \rightarrow \text{Ketchup}) =$$

$$33.3/10 = 3.33$$

Lift nous dit essentiellement que la probabilité d'acheter un hamburger et du ketchup ensemble est 3,33 fois plus élevée que la probabilité d'acheter simplement le ketchup. Un Lift de 1 signifie qu'il n'y a aucune association entre les produits A et B. Un Lift supérieur à 1 signifie que les produits A et B sont plus susceptibles d'être achetés ensemble. Enfin, un lift inférieur à 1 fait référence au cas où il est peu probable que deux produits soient achetés ensemble. [10]

## 2.2 Algorithme Apriori

L'algorithme Apriori créé par Agrawal et Srikant en 1994, procède en deux temps.

Il est basé sur le principe lié à l'approche de support et de confiance.

L'algorithme parcourt le treillis des itemsets pour rechercher les itemsets fréquents et en déduire les règles d'association dont la confiance dépasse le seuil de confiance  $\text{minconf}$ . Le treillis des itemsets permet d'utiliser plus efficacement cet algorithme d'extraction en admettant les propriétés suivantes :

- Tout sous-ensemble d'un itemset fréquent est fréquent.
- Tout sur-ensemble d'un itemset non fréquent est non fréquent.

### Exemple de treillis des itemsets.

Le nombre d'itemsets fréquents qui peuvent être générés de  $n$  items est de  $2^n - 1$ , la génération des itemsets fréquents est de complexité exponentielle, il est alors essentiel de trouver la méthode de recherche la plus optimale.

L'algorithme Apriori se fait comme suit :

- Générer les Règles candidates.
- Calculer le support pour chaque règle candidate.
- Appairer les règles dont on a calculé le support avec le support choisi.
- On rejette les candidats dont le support est inférieur au support minimal.
- On termine en sortie avec toutes les règles dont le support est supérieur au support minimal. [11]

### 2.2.1 Avantages et inconvénients de l'Algorithme Apriori

- **Avantages:** Il existe une multitude d'avantages dans l'utilisation de l'algorithme Apriori.

On en énumère quelques-uns:

- La découverte rapide de règles d'association pertinentes entre objets.
- La facilité d'interprétation des résultats lors de l'extraction des règles d'association, malgré le nombre important de ces dernières. [12] [13]
- **Inconvénients :** Les inconvénients auxquels on fait face lors d'une utilisation de l'algorithme Apriori sont les suivants:
  - Les algorithmes d'extraction liés à l'approche support / confiance génèrent un grand nombre de règles d'association.
  - Un nombre important de configurations d'items ne peuvent pas engendrer de règles d'association.
  - La recherche de règles d'association impose un temps considérable qui peut s'avérer désavantageux si l'on fait face à une énorme base de données.

## 3 Conclusion

La recherche d'association est une technique d'analyse de données qui identifie les relations récurrentes entre les éléments de grands groupes. Largement utilisé, il révèle les motivations cachées et oriente les décisions commerciales. Son efficacité repose sur des algorithmes comme Apriori, mais nécessite une interprétation spécialisée des résultats.

# **Chapitre 3 : Approche de la prédiction par les règles d'association**

# 1 Introduction

L'extraction des règles d'association est l'une des tâches fondamentales de la fouille de données. Elle vise à découvrir des relations intéressantes cachées dans de grandes bases de données. Ces relations entre les items (ensembles d'attributs) sont présentées sous forme d'implications, appelées règles d'association. Depuis la première application, largement connue sous la dénomination de panier de la ménagère, plusieurs nouveaux domaines d'application ont été identifiés comme la bio-informatique, le diagnostic médical, la détection d'intrusion, la fouille du web et l'analyse des données scientifiques. L'extraction de règles d'association a connu de nombreux développements théoriques et algorithmiques. Parmi ces algorithmes, Apriori et FP-Growth sont les plus connus. Tous ces algorithmes partagent une même méthodologie à deux étapes. La première est consacrée à la recherche de tous les items Ets fréquents, et la seconde consiste à générer les règles avec une grande confiance en combinant ces items Ets fréquents. On peut citer aussi l'approche logique GUHA. Les contraintes font souvent partie de la spécification de plusieurs problèmes de fouille de données. Cette observation a conduit à un nouveau domaine de recherche actif et multidisciplinaire, et permettant une fertilisation croisée entre la fouille de données et l'intelligence artificielle (IA). Deux modèles de représentation et de résolution d'IA ont été utilisés pour modéliser et résoudre plusieurs problèmes de fouille de données, à savoir la programmation par contrainte (PPC) et la satisfiabilité propositionnelle (SAT). Parmi ces problèmes, nous pouvons citer l'extraction de motifs et le clustering.

## 2 La maladie du stress

Le stress désigne une tension nerveuse, contrainte de l'organisme face à une situation redoutée (événement soudain, traumatisme, sensation forte, bruit, surmenage). "Le stress est une notion récente, défini pour la première fois au 20e siècle, Pour le pionnier des études sur le stress, le stress se définit comme "une agression de l'organisme par un agent physique, psychique, émotionnel entraînant un déséquilibre qui doit être compensé par un travail d'adaptation".

Comme l'explique le docteur Alexandre Hubert, "le stress est une réaction psychologique (et émotionnelle) et physique (et physiologique) à un agent stressant. Le stress peut se traduire par un sentiment d'impuissance et d'incapacité de l'individu à surmonter la situation redoutée. . "Et de préciser : "L'état d'alerte du sujet déclenche une cascade hormonale avec la libération d'adrénaline et de cortisol qui mobilise l'énergie du sujet. Des

manifestations physiques sont alors éprouvées : mains moites, tremblements, augmentation du rythme cardiaque..."[14]

### **3 Les systèmes des prédictions**

En règle générale, il existe deux types de modèles d'analyse prédictive : les modèles de classification et de régression. Les modèles de classification tentent de placer des objets de données (tels que des clients ou des résultats potentiels) dans une catégorie ou une autre. Par exemple, si un marchand dispose de nombreuses données sur différents types de clients, il peut essayer de prédire quels types de clients seront réceptifs aux e-mails marketing. Les modèles de régression tentent de prédire des données continues, telles que le revenu que va générer un client au cours de sa relation avec l'entreprise. L'analyse prédictive est généralement réalisée selon trois grands types de techniques :

- **Analyse de régression**

La régression est une technique d'analyse statistique qui permet d'estimer les relations entre les variables. La régression est utile pour déterminer les tendances dans des ensembles de données volumineux afin d'établir une corrélation entre les entrées. Il est préférable de l'employer sur des données continues suivant une distribution connue. La régression est souvent utilisée pour déterminer comment une ou plusieurs variables indépendantes affectent une autre, par exemple l'impact d'une augmentation de prix sur la vente d'un produit.

- **Arbres de décision**

Les arbres de décision sont des modèles de classification qui placent les données dans différentes catégories en fonction de variables distinctes. Cette méthode est surtout utile pour tenter de comprendre les décisions d'un individu. Le modèle ressemble à un arbre, où chaque branche représente un choix potentiel, et la feuille de la branche représente le résultat de la décision. Les arbres de décision sont généralement faciles à comprendre et fonctionnent bien lorsqu'il manque plusieurs variables dans un ensemble de données.

- **Réseaux de neurones**

Les réseaux de neurones sont des méthodes de machine Learning utiles pour l'analyse prédictive lors de la modélisation de relations très complexes. En résumé, il s'agit de puissants moteurs de reconnaissance de formes. Les réseaux de neurones sont surtout utilisés pour déterminer les relations non linéaires dans les ensembles de données, notamment lorsqu'il

n'existe aucune formule mathématique connue pour analyser les données. Les réseaux de neurones peuvent être utilisés pour valider les résultats des arbres de décision et des modèles de régression. [1]

## 4 Les données utilisées

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	anxiety_level	self_esteem	mental_health	depression	headache	blood_pressure	sleep_quality	breathing_problems	noise_level	living_conditions	safety	basic_needs	academic_performance	study_load	teacher_student_ratio	future_care	social_support	peer_pressure	extracurricular	bullying	stress_level
2	14	20	0	11	2	1	2	4	2	3	3	2	3	2	3	3	2	3	3	2	1
3	15	8	1	15	5	3	1	4	3	1	2	2	1	4	1	5	1	4	5	5	2
4	12	18	1	14	2	1	2	2	2	2	3	2	2	3	3	2	2	3	2	2	1
5	16	12	1	15	4	3	1	3	4	2	2	2	2	4	1	4	1	4	4	5	2
6	16	28	0	7	2	3	5	1	3	2	4	3	4	3	1	2	1	5	0	5	1
7	20	13	1	21	3	3	1	4	3	2	2	1	2	5	2	5	1	4	4	5	2
8	4	26	0	6	1	2	4	1	1	4	4	4	5	1	4	1	3	2	2	1	0
9	17	3	1	22	4	3	1	5	3	1	1	1	1	3	2	4	1	4	4	5	2
10	13	22	1	12	3	1	2	4	3	3	3	3	3	3	2	3	3	3	2	2	1
11	6	8	0	27	4	3	1	2	0	5	2	2	2	2	1	5	1	5	3	4	1
12	17	12	1	25	4	3	1	3	4	2	1	1	1	3	1	4	1	4	4	5	2
13	17	15	1	22	3	3	1	5	5	2	1	1	1	3	1	4	1	5	5	4	2
14	5	28	0	8	1	2	4	2	2	3	5	5	5	2	4	1	3	1	1	1	0
15	9	23	1	24	4	3	1	0	1	2	4	3	1	2	3	3	0	1	0	1	2
16	2	28	0	3	1	2	4	2	1	3	4	4	4	2	5	1	3	1	2	1	0
17	11	21	0	14	3	1	2	4	2	2	2	2	3	3	3	3	2	3	2	2	1
18	6	28	0	1	1	2	4	2	1	4	5	4	5	1	5	1	3	2	2	1	0
19	7	25	0	3	1	2	4	2	2	4	5	4	4	2	5	1	3	1	1	1	0
20	11	23	0	12	3	1	2	2	3	2	3	3	2	3	2	2	3	3	2	3	1

Figure 2 les données de la maladie traité [15]

Cet ensemble de données contient environ 20 fonctionnalités qui créent le plus d'impact sur le stress d'un étudiant. Les caractéristiques sont sélectionnées scientifiquement en tenant compte de 5 facteurs majeurs : les facteurs psychologiques, physiologiques, sociaux, environnementaux et académiques. Certains d'entre eux sont:

Facteurs psychologiques => « niveau d'anxiété », « estime de soi », « historique\_de\_santé\_mentale », « dépression »,

Facteurs physiologiques => 'maux de tête', 'tension artérielle', 'qualité du sommeil', 'problème respiratoire',

Facteurs environnementaux => « niveau de bruit », « conditions de vie », « sécurité », « besoins fondamentaux »,

Facteurs académiques => « performances académiques », « charge d'études », « relation\_enseignant\_étudiant », « préoccupations\_de\_carrière\_future »,

Facteur social => « soutien social », « pression des pairs », « activités parascolaires », « intimidation » [15]

Variable	Min	25%	50%	75%	Max
Niveau d'anxiété	0	6	11	16	21
Estime de soi	0	11	19	26	30
Historique de santé mentale	0	0	0	1	1
Dépression	0	6	12	19	27
Maux de tête	0	1	3	3	5
Tension artérielle	1	1	2	3	3
Qualité du sommeil	0	1	3	4	5
Problème respiratoire	0	2	3	4	5
Niveau de bruit	0	2	3	3	5
Conditions de vie	0	2	2	3	5
Sécurité	0	2	2	4	5
Besoins fondamentaux	0	2	3	4	5
Performances académiques	0	2	2	4	5
Charge d'études	0	2	2	3	5
Relation enseignant-étudiant	0	2	2	4	5
Préoccupations de carrière future	0	1	2	4	5
Soutien social	0	1	2	3	3
Pression des pairs	0	2	2	4	5
Activités parascolaires	0	2	3	4	5
Intimidation	0	1	3	4	5

**Tableau 1 : Pourcentage de Stress**

## 5 Le prétraitement des données

- **Partitionnement des données**

Pour traiter les données, nous avons sélectionné les 7 variables suivantes: ['anxiety\_level', 'depression', 'headache', 'blood\_pressure', 'sleep\_quality', 'breathing\_problem', 'stress\_level']. Nous avons ensuite divisé le jeu de données en deux parties: 80% des données ont été utilisées pour l'entraînement du modèle, et les 20% restants ont été réservés pour tester et évaluer les performances du modèle.

- **extraire et testé les règles d'association et le model**

Après avoir divisé les données, nous prenons les 80% pour l'entraînement et nous appliquons l'algorithme Apriori afin d'en extraire les règles d'association. Nous sélectionnons alors les bonnes règles pour la prédiction, c'est-à-dire celles ayant une valeur de "lift" élevée. Ensuite, nous testons le modèle sur les 20% de données restantes. Si le modèle est performant,

nous devrions obtenir un bon taux de prédictions correctes. La qualité du modèle est jugée sur sa capacité à bien prédire les cas de l'ensemble de test qui n'a pas servi à l'entraînement. Un taux de prédictions correctes élevé sur ces nouvelles données indique que le modèle a bien appris les motifs sous-jacents et qu'il sera capable de bien généraliser sur d'autres exemples.

## 6 L'apprentissage du modèle

Une explication détaillée de ce schéma d'apprentissage du modèle :

**1. Données initiales :** Le processus commence par l'accès à une base de données stress (Dataset stress)[15]. Ces données sont la matière première pour l'apprentissage du modèle.

**2. Prétraitement :** Les données subissent une étape de prétraitement pour les préparer à l'analyse.

**3. Séparation des données :** Les données sont divisées en deux ensembles :

- 80% pour l'entraînement (Train)
- 20% pour les tests

**4. Recherche des R- ASS:** Cette étape implique l'analyse des relations ou motifs dans les données.

**5. Apprentissage :** Le processus d'apprentissage consiste à calculer des règles basées sur les motifs identifiés.

**6. Vérification des règles :** Les règles générées sont vérifiées avec un critère appelé "Lift".

**7. Création du modèle :** Deux types de modèles sont créés :

- Modèle 01 =  $\Sigma$  règle de moins Lift
- Modèle 02 =  $\Sigma$  règle de grand Lift

**8. Test du modèle :** Le modèle est testé sur les 20% de données réservées aux tests.

**9. Évaluation des résultats :** Les résultats du test sont évalués. Deux issues sont possibles :

- Pourcentage faible : le modèle n'est pas satisfaisant
- Pourcentage élevé : le modèle est considéré comme fiable

**10. Finalisation :** Si le modèle est jugé fiable, il est validé pour être utilisé en prédiction.

Ce schéma représente un processus typique d'apprentissage automatique, comprenant la préparation des données, l'entraînement du modèle, sa validation et son utilisation pour des prédictions. Il met l'accent sur l'importance de la qualité des données, de la validation rigoureuse et de l'itération pour obtenir un modèle fiable.

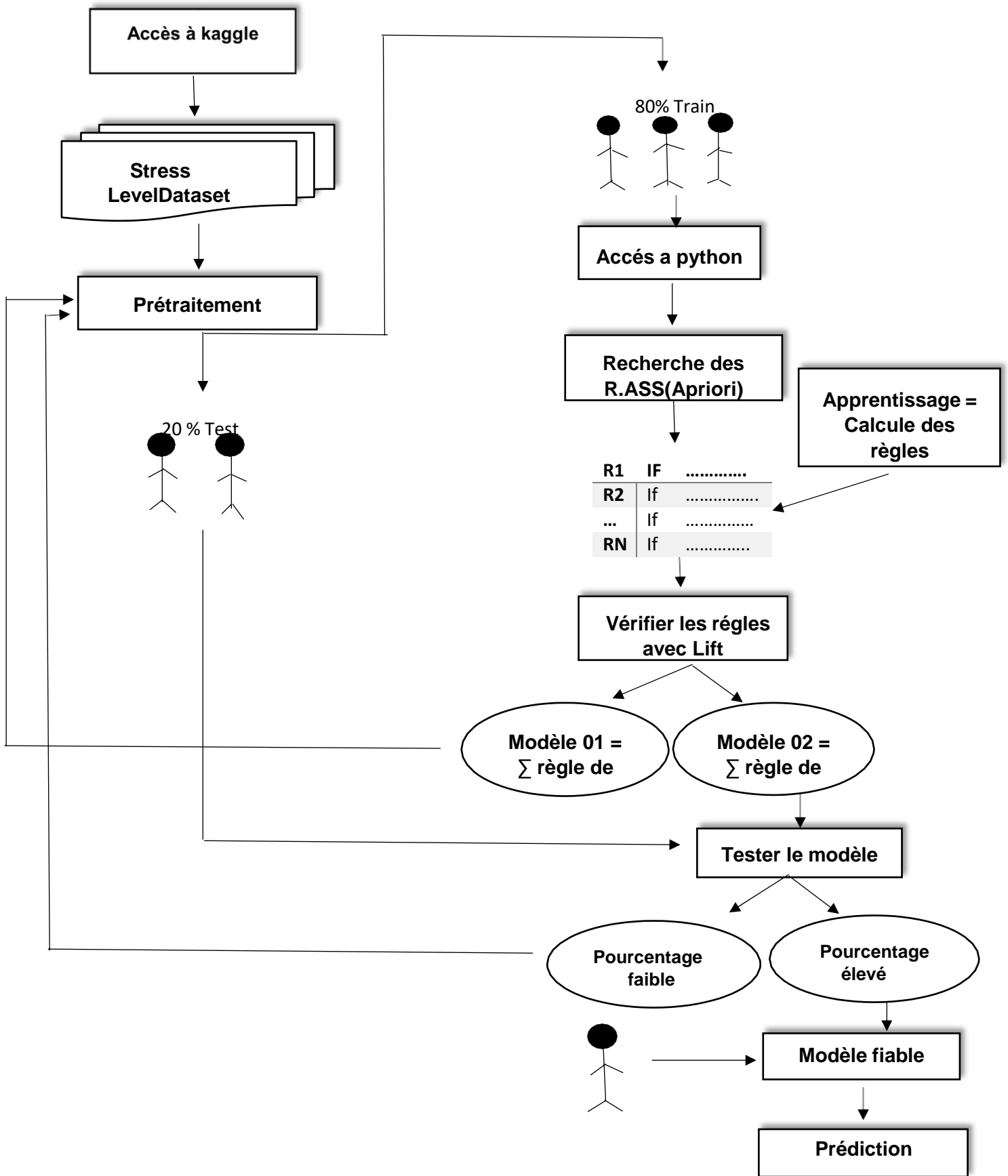


Figure3 schéma d'apprentissage du modèle

## 7 La prédiction

- L'explication détaillée de ce schéma :

**1. Traitement des données :** Le processus commence par le traitement des données. Il existe plusieurs ensembles de données. Cette étape est cruciale car elle prépare les données initiales à l'analyse et à la modélisation.

**2. Prédiction :** Les données traitées sont ensuite utilisées à des fins de prédiction. Cette étape consiste à appliquer un modèle prédictif aux données préparées pour générer des résultats ou des prédictions.

**3. Modèle  $\Sigma$  Règle A.ss :** Ce bloc représente le cœur du système – un modèle basé sur la somme ( $\Sigma$ ) des règles A.ss. « A.ss » signifie « association ». Ce modèle combine différentes règles ou critères pour faire ses prédictions.

**4. Interaction humaine :** La figure humaine associée au modèle suggère une interaction ou une supervision humaine. Cela inclut l'ajustement d'un modèle, l'interprétation des résultats ou la prise de décisions basées sur les prédictions du modèle.

**5. Niveau de stress :** Le résultat final du processus est de déterminer le niveau de stress. Cela suggère que le modèle est spécifiquement conçu pour évaluer ou prédire les niveaux de stress, dans un contexte médical, psychologique ou organisationnel.

En résumé, ce diagramme illustre un processus de modélisation prédictive axé sur l'évaluation des contraintes. Il commence par traiter les données, utilise ces données pour faire des prédictions via un modèle basé sur des règles, implique une interaction humaine pour modifier ou interpréter et produit une estimation du niveau de stress. Ce type de système peut être utilisé dans divers domaines comme la santé au travail, la psychologie clinique ou la gestion des ressources humaines pour anticiper et gérer le stress.

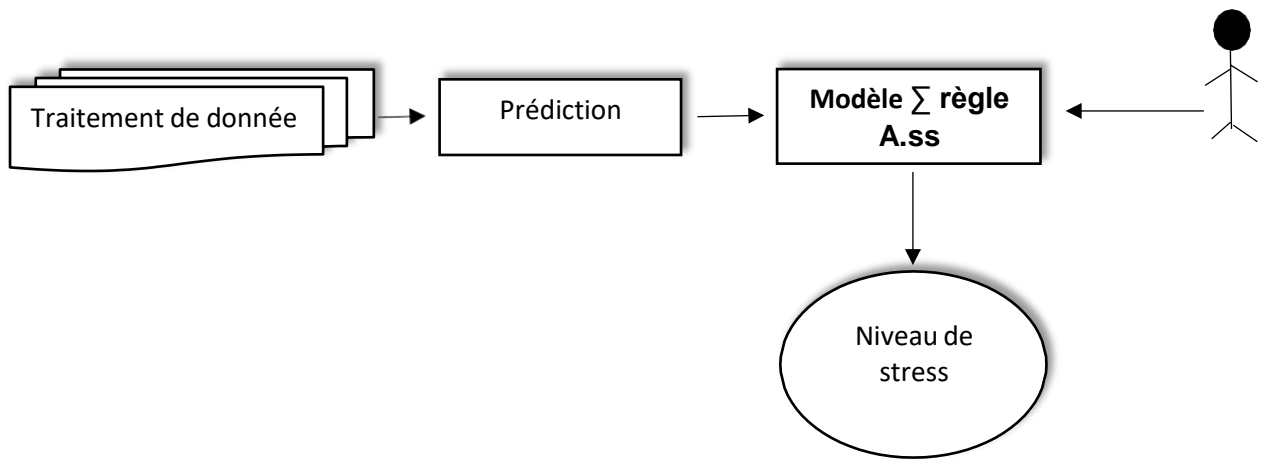


Figure4 Schéma de La prédiction du modèle

## 8 Conclusion

En conclusion, l'approche de la prédiction par les règles d'association est une méthode précieuse dans le domaine de l'exploration de données. Elle permet de découvrir des relations intéressantes et des modèles cachés entre les variables dans de grands ensembles de données transactionnelles. Grâce à cette approche, il est possible de détecter des tendances, des comportements d'achat, des associations entre produits, des corrélations entre symptômes médicaux, et bien plus encore.

L'utilisation des règles d'association offre de nombreux avantages, notamment la capacité à explorer les données sans supervision explicite, à générer des recommandations personnalisées, à optimiser les processus commerciaux, et à prendre des décisions éclairées. De plus, cette approche peut être mise en œuvre de manière efficace grâce à des algorithmes tels qu'Apriori et FP-Growth, qui permettent de traiter efficacement de grands ensembles de données.

Cependant, l'approche de la prédiction par les règles d'association présente également des défis, notamment la gestion de grands volumes de données, la sélection des paramètres appropriés pour les algorithmes, et l'interprétation des règles découvertes. De plus, il est important de noter que cette approche peut ne pas être adaptée à tous les types de données ou à tous les scénarios, et qu'elle doit être utilisée avec discernement en fonction du contexte spécifique de chaque application.

En résumé, l'approche de la prédiction par les règles d'association constitue un outil précieux dans l'arsenal des techniques d'exploration de données. Elle offre une méthode efficace pour découvrir des connaissances utiles à partir de données transactionnelles, et elle continue d'être largement utilisée dans de nombreux domaines pour améliorer la prise de décision et la génération de connaissances.

# **Chapitre 4 : Implémentation et Résultats**

# 1 Introduction

Ce chapitre se divise en deux parties : la première partie est consacrée à l'implémentation de du modèle proposé. Là où les différents choix d'implémentation sont présentés comme l'environnement de développement, les bibliothèques utilisées et les résultats de paramétrage effectué. La deuxième partie montre les différents tests réalisés, les résultats obtenus avec les discussions et les comparaisons avec des travaux reliés.

## 2 Environnement de développement matérielle et logicielle

- **Environnement matérielle**

Pour réaliser l'application, et pour le tester, on a utilisé les machines suivantes:



*Figure 5 pc acer aspire 5*

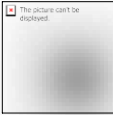
- **Processors:** 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.42 GHz
- **Mémoire installée :** (RAM) : 8.00 Go
- **Type de système :** système d'exploitation 64 bits.
- **Système d' exploitation:** Windows 11 pro
- **Environnement logicielle**
- **Langage de programmation**



**Anaconda :** Anaconda est une distribution scientifique de Python. Permet d'écrire et d'exécuter le code Python de son choix par le biais du navigateur. Offert par anaconda Enterprise (gratuit), utilisé Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. [16]



**Python :** Python est un langage de programmation interprété, orienté objet et de haut niveau, créé par Guido van Rossum et publié pour la première fois en 1991. Son nom est inspiré par la passion de van Rossum pour la série télévisée Monty Python. Python se distingue par sa syntaxe claire et concise, qui favorise la lisibilité et la productivité du code. Il supporte plusieurs paradigmes de programmation, notamment la programmation impérative, fonctionnelle et orientée objet. Python est souvent utilisé pour le développement web, le traitement des données, l'automatisation de tâches système, l'intelligence artificielle et bien d'autres domaines. Sa vaste bibliothèque standard et sa communauté active en font l'un des langages les plus populaires au monde. [17]



**Microsoft Word :** Word est un traitement de texte qui vous permet de créer différents types de documents tels que des lettres, des papiers, des dépliants et des fax. [18]



**Microsoft Excel:** Excel is a spreadsheet program that allows you to store, organize, and analyze information. While you may think Excel is only used by certain people to process complicated data, anyone can learn how to take advantage of the program's powerful features. Whether you're keeping a budget, organizing a training log, or creating an invoice, Excel makes it easy to work with different types of data. [19]

- **Bibliothèques utilisées**



**Pandas:** pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. [20]



**Tkinter :** Tkinter est la bibliothèque la plus couramment utilisée pour développer une GUI (Graphical User Interface) en Python. Il s'agit d'une interface Python standard pour la boîte à outils Tk GUI fournie avec Python. Comme Tk et Tkinter sont disponibles sur la plupart des plates-formes Unix ainsi que sur le système Windows, développer des applications GUI avec Tkinter devient le plus rapide et le plus simple.

[21]



**Pillow:** Python Pillow est construit sur PIL (Python Image Library) et est un fork du même PIL qui a été abandonné depuis 2011. Pillow prend en charge plusieurs formats de fichiers image, notamment BMP, PNG, JPEG et TIFF. La bibliothèque encourage l'ajout de la prise en charge de formats plus récents dans la bibliothèque en créant de nouveaux décodeurs de fichiers. [22]



**Apriori en Python :** algorithme d'exploration de données qui trouve des motifs fréquents et génère des règles d'association dans de grands ensembles de données. Utilisé pour l'analyse de transactions et les recommandations.

### 3 Implémentation

Cette partie est réservée aux détails de l'environnement de développement et le langage de programmation utilisés pour la réalisation de notre système. Nous avons présenté aussi la

base d'apprentissage et de test utilisés, les détails de l'implémentation de l'architecture proposée et l'interface graphique de ce système.

### 3.1 Le prétraitement

Ce code charge un fichier Excel, sélectionne certaines colonnes spécifiques, puis divise les données en deux sous-ensembles: 80% pour l'entraînement et 20% pour les tests. Il affiche le contenu des deux sous-ensembles ainsi que leur nombre de lignes respectif.

En résumé, il prépare les données d'un fichier Excel pour une tâche d'apprentissage automatique en séparant les données en ensembles d'entraînement et de test.

```
import pandas as pd

# Load the Excel file
file_path = 'AAA.xlsx'
data = pd.read_excel(file_path)

# Select the specified columns
columns_to_use = ['anxiety_level', 'depression', 'headache', 'blood_pressure', 'sleep_quality', 'breathing_problem', 'stress_level']
data = data[columns_to_use]

# Split the data into 80% and 20% subsets
split_point = int(0.8 * len(data))
data_80 = data[:split_point]
data_20 = data[split_point:]

# Display the first subset and its count
print("First 80% of the data:")
print(data_80)
print(f"Number of rows in the first 80%: {len(data_80)}")

# Display the second subset and its count
print("\nSecond 20% of the data:")
print(data_20)
print(f"Number of rows in the second 20%: {len(data_20)}")
```

Figure 6 programme de division des données

### 3.2 L'Apprentissage

- **Extraction des règles d'association (conf,sup)**

Ce programme lit le fichier Excel nommé « AAA.xlsx » et extrait les 880 lignes sélectionnées représentant 80 % des données et les colonnes sélectionnées pour créer la liste des enregistrements. Il utilise ensuite la bibliothèque "Apriori" pour trouver des règles d'association dans ces données en fonction de certains paramètres (support minimum, confiance minimum, lift minimum, longueur de règle minimum). Les règles d'association sont stockées dans la variable "association\_results". En bref, ce code effectue une analyse de corrélation sur un ensemble de données provenant d'un fichier Excel.

```

import pandas as pd
from apyori import apriori

file_path = 'AAA.xlsx'

columns_to_use = ['anxiety_level', 'depression', 'headache', 'blood_pressure', 'sleep_quality', 'breathing_problem', 'stress_level']

df = pd.read_excel(file_path, usecols=columns_to_use, nrows=880)

records = []
for i in range(len(df)):
    record = []
    for column_name in columns_to_use:
        record.append(str(df.loc[i, column_name]))
    records.append(record)

association_rules = apriori(records, min_support=0.01, min_confidence=0.1, min_lift=2, min_length=2)
association_results = list(association_rules)

```

Figure 7 programme d'extraction des règle d'associations

### ▪ Extraire des règles avec le Lift élevé

Ce code analyse des règles d'association dans un ensemble de données. Il commence par filtrer les règles dont le "lift" dépasse 2,5, les considérant comme pertinentes. Si de telles règles existent, le programme affiche leur nombre total, puis détaille chacune d'entre elles. Pour chaque règle, il présente son numéro, son contenu (les éléments impliqués), son support, sa confiance et son lift.

Ces informations sont affichées de manière structurée pour une lecture facile. En l'absence de règles pertinentes, le code affiche simplement un message indiquant qu'aucune règle ne correspond aux critères de prédiction.

Ce processus permet d'identifier et de mettre en évidence les associations les plus significatives dans les données, facilitant ainsi l'interprétation et l'utilisation des résultats pour des analyses ultérieures ou des prises de décision.

```

relevant_rules = [rule for rule in association_results if rule[2][0][3] > 2.5]

if any(relevant_rules):
    print("Nombre de règles pertinentes pour la prédiction :", len(relevant_rules))
    for idx, item in enumerate(relevant_rules, 1):
        pair = item[0]
        items = [columns_to_use[int(x)] for x in pair if int(x) < len(columns_to_use)]
        print(f"numéro de règle {idx}")
        print("Règle: " + " -> ".join(items))
        print("support: " + str(item[1]))
        print("confidence: " + str(item[2][0][2]))
        print("lift: " + str(item[2][0][3]))
        print("=====")
else:
    print("Il n'y a pas de règles liées à la prédiction")

```

Figure 8 programme d'extraction des règle d'associations de plus grand Lift

### 3.3 La prédiction

Cette fonction prédit le niveau de stress selon les données d'entrée (anxiety, dépression, maux de tête, qualité de sommeil, problèmes respiratoires, tension artérielle). Elle renvoie 0 si tout est normal, 1 pour un stress modéré, 2 pour un stress élevé selon des seuils prédéfinis sur les valeurs d'entrée.

```
def predict_stress_level(input_data):
    if (input_data['anxiety_level'] == 0 or input_data['depression'] == 0 or input_data['headache'] == 0 or
        input_data['sleep_quality'] == 0 or input_data['breathing_problem'] == 0):
        return 0
    elif input_data['blood_pressure'] == 1 and input_data['headache'] == 0:
        return 0
    elif (input_data['anxiety_level'] > 10 or input_data['depression'] > 13 or input_data['headache'] > 2 or
        input_data['blood_pressure'] > 3 or input_data['sleep_quality'] > 3 or input_data['breathing_problem'] > 3):
        return 2
    elif (input_data['anxiety_level'] > 5 or input_data['depression'] > 7 or input_data['headache'] > 1 or
        input_data['blood_pressure'] > 1 or input_data['sleep_quality'] > 1 or input_data['breathing_problem'] > 1):
        return 1
    else:
        return 0
```

Figure 9 programmes de la prédiction du modèle

## 4 Expérimentation et discussion des résultats

On va présenter dans cette partie : la configuration expérimentale, les critères de performance, Résultats et comparaison

### 4.1 Les métrique de performance utilisées (conf,Lift)

Pour construire le modèle de règles d'association, nous avons utilisé la fonction primitive pour extraire des règles de nos données en utilisant à la fois le support et la confiance. Nous avons appliqué cette fonction à 80 % de nos données pour extraire un total de 138 règles. En raison du grand nombre de règles, il n'a pas été possible de construire un modèle de prédiction.

Nous avons donc sélectionné des règles avec un effet de levier supérieur à 2,5 en utilisant Apriori, ce qui nous a donné 12 règles. Pour valider ces règles, nous avons utilisé les 20 % restants des données et avons trouvé un taux de précision de 91,82 %. Cela suggère que les corrélations sont fortes, ce qui nous permet de les utiliser pour prédire efficacement les niveaux de stress.

Précision des prédictions :s 91.82%

## 4.2 Discussion des Résultats

Dans ce programme, créez une interface utilisateur avec des entrées de données aléatoires spécifiques pour chaque point de données, où : Anxiety\_level 0 à 21, Dépression de 0 à 27, Headache de 0 à 5, Blood\_prossure de 1 à 3, Sleep\_quality de 0 à 5, Breathing\_probeme de 0 à 5.

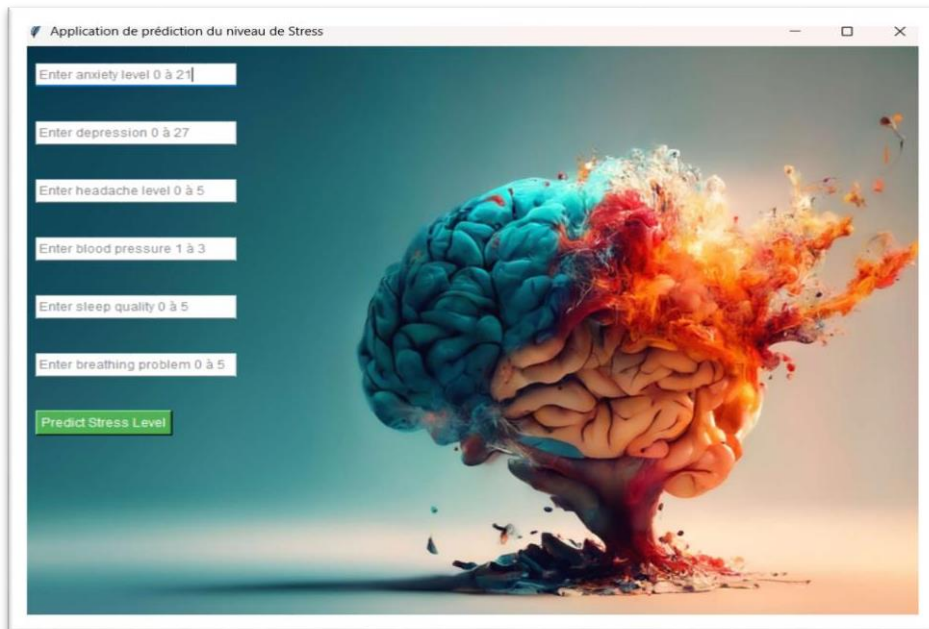


Figure 10 l'interface du modèle

Il s'agit de créer une application de prédiction où les niveaux de stress seront compris entre 0 et 2. Pour nous améliorer, nous avons inclus différentes images en fonction du niveau de stress attendu pour mieux illustrer l'état de la personne.

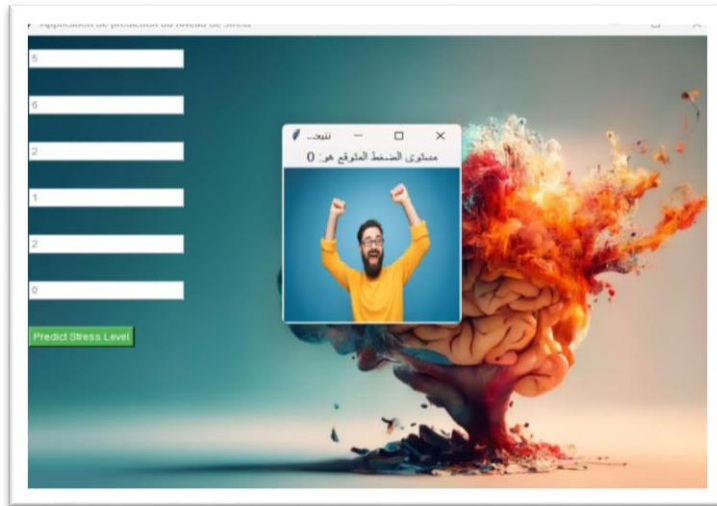


Figure 11 exemples de niveau de stress 0

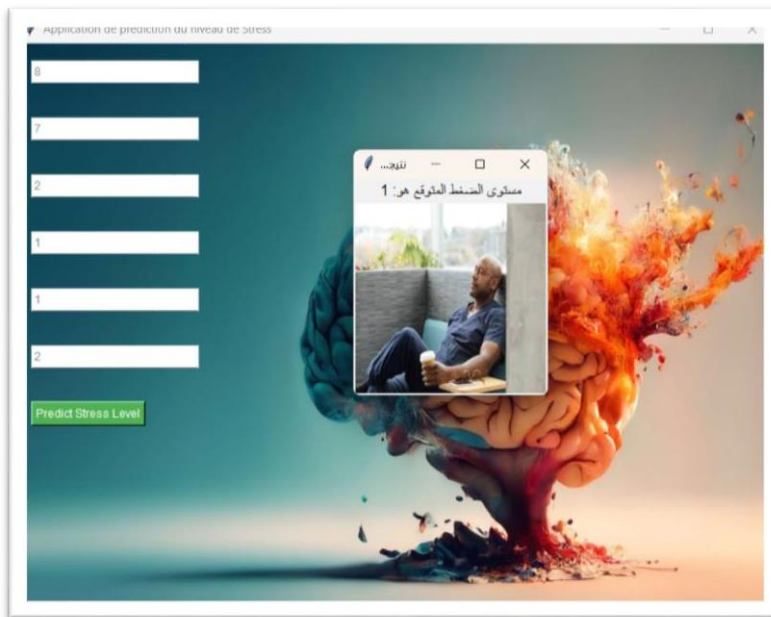


Figure12 exemples de niveau de stress 1

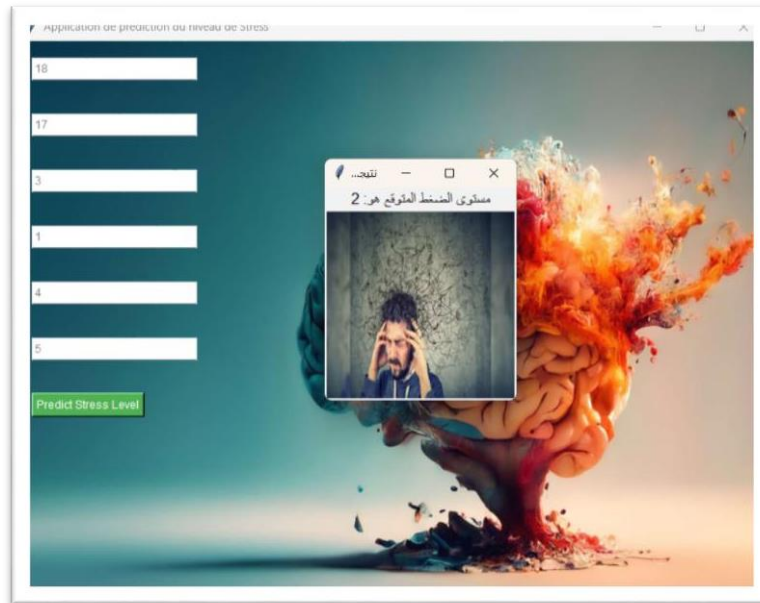


Figure 1 3 Exemple de niveau de stress 2

- **Discussion**

Le modèle que nous avons développé se distingue par ses performances exceptionnelles en matière de prévision et fonctionne avec une grande efficacité. En utilisant des algorithmes avancés tels que l'algorithme Apriori pour l'analyse des données, le modèle a réussi à extraire des règles d'association qui aident à fournir des prévisions précises sur les niveaux de stress.

Nous avons effectué des tests intensifs sur le modèle en utilisant divers ensembles de données, et les résultats ont démontré l'exactitude et l'efficacité du modèle dans la prédiction des différents niveaux de stress. Grâce à une analyse approfondie de multiples facteurs tels que l'anxiété, la dépression, la qualité du sommeil, et autres, le modèle peut fournir des résultats fiables en un temps record.

Non seulement le modèle est précis, mais il est également rapide dans le traitement des données et la fourniture de prévisions. Il a été conçu pour être un outil utile et facile à utiliser, ce qui en fait une solution idéale pour une variété d'applications nécessitant une évaluation précise des niveaux de stress.

En résumé, notre modèle atteint un équilibre parfait entre précision et efficacité, en faisant un outil précieux dans le domaine de la prévision et de la gestion du stress.

## 5 Conclusion

Ce chapitre a été divisé en deux parties : implémentation, résultats et comparaisons. Dans la première partie, nous avons présenté l'environnement de travail, le langage de programmation, les bases d'apprentissage et de test, les détails de l'implémentation de la phase d'apprentissage, ainsi que l'interface graphique. Dans la deuxième partie, nous avons présenté la configuration expérimentale, les résultats et la comparaison avec d'autres travaux. D'après les résultats, le modèle proposé la prédiction des règles d'association donne les meilleurs résultats.

# Conclusion général

## Conclusion général

Cette recherche a exploré l'utilisation de l'exploration de données pour la prédiction des niveaux de stress, aboutissant au développement d'un système prédictif performant. À travers l'analyse de données variées et l'application d'algorithmes d'apprentissage automatique, nous avons pu démontrer que notre modèle est capable de prédire efficacement les niveaux de stress des individus.

Les résultats obtenus montrent que le système développé est performant, offrant une précision notable dans la prédiction du stress. En utilisant des indicateurs physiologiques, le modèle fournit des évaluations fiables et continues des niveaux de stress, dépassant ainsi les méthodes traditionnelles reposant principalement sur des questionnaires auto-administrés.

Cette avancée technologique a des implications pratiques significatives. Les entreprises peuvent intégrer ce système dans leurs programmes de bien-être pour surveiller et gérer le stress des employés. Les institutions éducatives peuvent l'utiliser pour identifier les étudiants en détresse. Les professionnels de santé peuvent l'adopter comme outil complémentaire dans le suivi de leurs patients.

En conclusion, notre système prédictif de stress, basé sur l'exploration de données, constitue une avancée prometteuse pour la gestion du stress. En combinant technologie et analyse de données, nous ouvrons la voie à des interventions plus précoces et personnalisées, contribuant ainsi à une meilleure qualité de vie. Cette recherche pose les bases d'un avenir où la technologie joue un rôle central dans la promotion du bien-être mental.

## Bibliographie

- [1] La prédiction : <https://cloud.google.com/learn/what-is-predictive-analytics?hl=fr#section-7> consulté le 04/06/2024
- [2] Blanchard, Julien, et al. "Mesure de la qualité des règles d'association par l'intensité d'implication entropique." *Revue des nouvelles technologies de l'information RNTI E 1* (2004).
- [3] Huang, A. : Mesures de similarité pour le regroupement de documents texte. Dans : Actes de la sixième conférence I des étudiants en recherche en informatique de Nouvelle-Zélande (NZCSRSC2008), Christchurch, Nouvelle-Zélande, pp. 49-56 (2008)
- [4] L.Kanza (Extraction de motifs séquentiels)Mémoire de fin d'étude pour l'obtention du master UNIVERSITE MOHAMED BOUDIAF-M'SILA, (2015/2016).
- [5] N. Y. Chen, D. Zhu ( Data mining for Industrial design and Diagnosis.In proceedings of the 1st International) Conference on Multisensor Multisource Information Fusion, Las Vegas, 1998.
- [6] T.Yuan, Jen, D. Laurent, and G. Sambe. "Utilisation de règles d'association pour la prédiction de valeurs manquantes." *EDA*. (2009).
- [7] A.Aussem,. Théorie et applications des réseaux de neurones récurrents et dynamiques à la prédiction, à la modélisation et au contrôle adaptif des processus dynamiques. Diss. Paris 5, 1995.
- [8] S.Charani, Nicolas. "Correction bayésienne de prédictions issues d'arbres de décision et évaluation crédibiliste." *LFA 2019-28ème Rencontres Francophones sur la Logique Floue et ses Applications*. (2019).
- [9] H. CHERFI : « Etude et réalisation des d'un système d'extraction de connaissance à partir de textes », (2006)
- [10] ift, confiance, suporte : <https://stackabuse.com/association-rule-mining-via-apriori-algorithm-in-python/>
- [11] A.NOUASRIA : « EXTRACTION D'ASSOCIATIONS LEXICALES FORTES DANS LES COMMENTAIRES », L'UNIVERSITÉ DU QUÉBEC À TROIS RIVIÈRES, JUIN 2016.
- [12] A.Abdelghani : "Extraction de relations d'association maximales dans les textes : Représentation graphique", Université du Québec à Trois-Rivières, (2012).

- [13] C.Pagé : "Bases de règles multi-niveaux", Université du Québec à Montréal, Février 2008.
- [14] la maladie de stress : <https://www.santemagazine.fr/sante/fiche-maladie/stress-177599#D%C3%A9finition-%3A-qu%27est-ce-que-le-stress-%3F> consulté le 19/06/2024
- [15] StressLevelDataset: <https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis> consulté le 20/06/2024
- [16] anaconda : <https://www.anaconda.com/about-us> consulté le 23/06/2024
- [17] python : <https://www.python.org/doc/essays/blurb/> consulté le 23/06/2024
- [18] Microsoft Word: <https://edu.gcfglobal.org/en/word2010/getting-started-with-word/1/> consulté le 23/06/2024
- [19] Microsoft Excel : <https://edu.gcfglobal.org/en/excel/getting-started-with-excel/1/> consulté le 23/06/2024
- [20] Pandas: <https://pandas.pydata.org/> consulté le 23/06/2024
- [21] Tikenter : <https://www.geeksforgeeks.org/python-tkinter-tutorial/> consulté le 23/06/2024
- [22] Pillow : <https://www.geeksforgeeks.org/python-pillow-tutorial/> consulté le 23/06/2024