



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTÉ MATHÉMATIQUES ET INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : Génie Logiciel

Par :

Malki Yasmine
Chekari kheira

Sur le thème

L'application des techniques de Data Mining sur les systèmes d'intelligence ambiante

Soutenu publiquement le 29 / 09 / 2020 à Tiaret devant le jury composé de :

Mr Boudaa Boudjemaa

MCB Université tiaret

Président

Mr Mokhtari Ahmed

MAA Université tiaret

Encadreur

Mr Djafri Laouni

MAA Université tiaret

Examinateur

2019-2020

Dédicace

A mes chers parents

Pour leurs immense soutien, leurs grand amour, leurs sacrifices et leurs prières...Vous méritez tout éloge, vous qui avez fait de moi ce que je suis, je ne peux pas le reconnaître en quelques lignes. J'espère être à la hauteur de vos attentes. Et je tiens à ce que vous sachiez que vous êtes les personnes les plus chères à mon cœur.

A mon mari Saadi Abdelkader qui m'a encouragé de terminer cette thèse .

A mes très chères frères Mohamed et Hamza , mon petite sœur Fatima, Sachez bien que je vous dois beaucoup pour votre compréhension, votre soutien et votre amour qui m'ont été très utiles pour aller de l'avant.

A ma belle-famille.

A mon binôme Kheira et toute sa famille

A tous mes amies, pour les beaux moments que nous avons passées ensemble.

A tous ceux qui m'aiment,

A tous ceux que j'aime,

A tous ceux qui m'ont aidé de près ou de loin.

YASMINE

Dédicace

Je dédie ce mémoire

A mes chers parents, pour leur patience, leur soutien et leurs encouragements ;

A mes très chers frères et mes belles sœurs et leurs petites familles ;

A mon mari Abdelkader avec tout mon amour ;

A toute ma famille ;

A mon binôme Yasmin et toute sa famille ;

A mes amis et à tous ceux que j'aime et qui m'aiment ;

A tous ceux qui m'ont aidée.

Kheira

Remerciements

Nous remercions DIEU tout puissant de nous avoir données la force, la santé, le courage et la patience de pouvoir accomplir ce travail.

Un grand merci à toutes nos familles surtout nos parents pour leur encouragement et leur suivi avec patience du déroulement de notre projet.

Nous tenons à remercier Mr MOKTARI Ahmed ,notre tuteur de projet de fin d'étude à l'Université IBN Khaldoun tiaret qui nous a suivi tout au long de cette période et nous 'a conseillé sur l'orientation que celui-ci devait prendre.

Nous tenons à remercier chacun des membres du jury pour nous avoir fait l'honneur d'examiner et d'évaluer notre travail et de l'enrichir par leurs propositions.

Résumé :

L'intelligence ambiante vise à offrir un espace quotidien « intelligent » d'accès à l'information ou à des services numériques permettant une utilisation adaptée, naturelle et conviviale. Un espace technologique adapté est capable de «comprendre» les caractéristiques des usagers, de l'environnement, de s'adapter contextuellement aux besoins, de répondre intelligemment aux demandes ou de réagir de façon appropriée. Cette « intelligence » est rendue possible par la convergence des technologies mêlant objets intelligents, réseaux de communication et interfaces multimodales pour fournir de nouveaux services aux utilisateurs.

L'intelligence ambiante peut être vue comme l'opposé de la réalité virtuelle. La réalité virtuelle met une personne à l'intérieur d'un monde créé par l'ordinateur. L'intelligence ambiante a pour but de permettre à l'ordinateur de «vivre» dans le monde des hommes et de s'y intégrer au point de disparaître.

Le but recherché dans ce travail est l'étude de comportements des habitants qui utilisent un système d'intelligence ambiante par l'utilisation des techniques de Data Mining pour l'extraction des connaissances sur leurs habitudes.

Mots-clés :

Intelligence ambiante, Data Mining ,processus ECD, Comportement des habitants .

Abstract:

Ambient intelligence aims to provide an "intelligent" daily space for accessing information or digital services allowing appropriate, natural and user-friendly use. A suitable technological space is capable of "understanding" the characteristics of users, of the environment, of contextually adapting to needs, of responding intelligently to requests or of reacting appropriately. This "intelligence" is made possible by the convergence of technologies combining intelligent objects, communication networks and multimodal interfaces to provide new services to users.

Ambient intelligence can be seen as the opposite of virtual reality. Virtual reality puts a person inside a world created by the computer. The purpose of ambient intelligence is to allow the computer to "live" in the world of men and to integrate into it to the point of disappearing.

The goal of this work is to study the behavior of inhabitants who use an ambient intelligence system through the use of Data Mining techniques to extract knowledge about their habits.

Keywords :

Ambient intelligence, Data Mining, KDD process, Behavior of inhabitants.

Table de matière

Résumé

Abstract

Introduction général 1

Chapitre 01 : L'intelligence ambiante

Introduction	3
1.1 Historique	4
1.2 Définition	4
1.3 Les technologie de l'AMI	8
1.3.1 L'informatique ubiquitaire	8
1.3.2 Communication omniprésente	9
1.3.3 Interface utilisateur intelligente	9
1.3.4 Bluetooth Low Energy	9
1.3.5 Identification par radio fréquence	9
1.3.6 Implant Microchip	10
1.3.7 Capteur	10
1.4 Les étapes principales de L'AMI	11
1.4.1 Détection	11
1.4.2 Raisonnement	11
1.4.3 Agissant	11
1.4.4 Interagir avec l'utilisateur	12
1.5 Application de L'AMI	12
1.6 Problèmes de L'AMI	13
1.7 Avantage de L'AMI	14
Conclusion	14

Chapitre02 : Data Mining

Introduction	15
2.1 Qu'est-ce que le Data Mining?	15
2.2 Définition de Data Mining	16
2.3 Processus ECD	16
2.3.1 Définition	17
2.3.2 Etapes de processus	18

2.3.3 Caractéristique d'un processus ECD	20
2.4 Data Mining sur quels types de données ?	21
2.5 Les objectifs au data mining	21
2.6 Tâches réalisées en Data Mining	22
2.6.1 Descriptives	22
2.6.2 Prédictives	23
2.7 Techniques du Data Mining	25
2.8 Domaines d'application du data mining	25
2.9 Avantages du Data Mining	26
2.10 Difficultés de Data Mining	26
Conclusion	27

Chapitre 03 : Application de data mining sur AMI

Introduction	28
3.1 Reconnaissance d'activité en temps réel dans une maison intelligente à l'aide de capteurs binaires	28
3.2.1 Pourquoi les capteurs binaires	29
3.1.2 Base de données	29
3.2 L'application de processus ECD	31
3.2.1 Sélectionner les données pertinentes	32
3.2.2 Prétraitement de la base de données	33
3.2.3 Transformation des données	33
3.2.4 Application de techniques de Data Mining	33
3.2.5 Évaluer et valider les résultats	35
Conclusion	35

Chapitre 04 : L'implémentation de notre application

Introduction	36
4.1 Ressource utilisées	36
4.1.1 Excel	36
4.1.2 Weka	37
4.1.3 NetBeans	38

4.1.4 Fichier WEKA. JAR	38
4.2 Mise en œuvre de la solution	39
4.2.1 Présentation de prétraitement de notre base de données	39
4.2.2 Transformation des données	40
4.3 Capture d'écran	40
Conclusion	45
Conclusion générale	46
Bibliographie	47
Webographie	47

Table des figures

Figure 01 :	architecture générique d’AmI	08
Figure 02 :	Technologies utilisées dans AMI	10
Figure 03 :	Les étapes principales de l’AMI	11
Figure 04 :	Les étapes de processus ECD	18
Figure 05 :	Graphe représentant les méthodes du Data Mining	22
Figure 06 :	Maison intelligente	29
Figure 07 :	Plan de localisation des capteurs	30
Figure 08 :	exemple de la base de données	31
Figure 09 :	Exemple de l’arbre de décision	34
Figure 10 :	Les principales classes du fichier weka.jar	39
Figure 11 :	La base de données après le prétraitement des données	39
Figure 12 :	La base de données après transformation des données	40
Figure 13 :	Interface Accueil	41
Figure 14 :	Interface de classification des activités par rapport aux temps d’un 1 jour ..	42
Figure 15 :	Interface de classification des activités par rapport aux temps de la BDD Aruba	42
Figure 16 :	Interface de visualisation l’arbre de décision pour 1 jour	43
Figure 17 :	Interface de visualisation l’arbre de décision pur la BDD Aruba	43

Liste des tableaux

Tableau 01 :	Les différentes définitions et caractéristiques de l’AmI	06
---------------------	--	----

Introduction générale

Introduction générale

L'intelligence ambiante offre un nouveau niveau d'interaction homme-machine dans lequel les gens sont entourés d'interfaces intelligentes et intuitives intégrées dans les objets quotidiens qui les entourent.

Elle prête discrètement attention aux actions des utilisateurs finaux dans certains environnements pour apprendre ce que font les gens, comment ils réagissent et leurs intérêts pour produire des données pour faire des suppositions intuitives et basées sur des données concernant ce que les utilisateurs veulent et l'action à effectuer. Celles-ci peuvent être aussi simples que d'allumer une lumière ou de vous assurer que votre maison est préchauffée le soir, car elle sait que vous arrivez à la maison tard le mardi soir.

Afin de réaliser ce que nous allons mentionner, on peut utiliser le Data Mining, qui est défini comme étant l'ensemble des méthodes et techniques destinées à l'exploration et l'analyse de bases de données informatiques (souvent grandes), en vue de détecter et extraire dans ces données des règles, des associations, des tendances inconnues ou cachées, des structures particulières restituant l'essentiel de l'information.

Contexte de travail

Le travail mené dans le cadre de ce mémoire relève du domaine de l'intelligence ambiante et du Data Mining appliqué aux données d'un habitant qui ont été recueillies au domicile d'une femme âgée bénévole pendant 7 mois.

Problématique

Dans ce travail, nous nous sommes intéressés aux activités et réactions des habitants qui utilisent un système d'intelligence ambiante qui constitue le cœur de notre problématique.

Une question principale a été posée au départ de nos recherches : Comment prédire le comportement ou les activités des utilisateurs de système d'intelligence ambiant ? Afin d'extraction des connaissances sur leurs habitudes.

Objectif

L'objectif de ce travail est d'améliorer les conditions du confort des habitants de système AmI par l'étude de leurs comportements .On utilise les techniques de Data Mining pour l'extraction des connaissances.

Plan du mémoire

Notre mémoire est organisé en quatre chapitres :

Premier chapitre : l'intelligence ambiante AmI

Ce chapitre présente des définitions et concepts sur l'AmI.

Deuxième chapitre : Data Mining

Ce chapitre présente les notions et les concepts de Data Mining et le processus Extraction de Connaissances dans les Bases de Données (ECD).

Troisième chapitre : Application de Data Mining sur AmI

Dans le troisième chapitre, nous détaillerons la conception de notre solution. Puis nous présenterons l'application de techniques de Data Mining sur un système de l'intelligence ambiante.

Quatrième chapitre : L'implémentation de notre application

Dans le quatrième chapitre nous présentons la mise en œuvre de notre solution. Premièrement nous décrivons les différents outils et technologies que nous allons utiliser. Puis nous présenterons la mise en œuvre de la solution. Enfin nous présenterons quelques captures d'écran de l'application.

Chapitre 01

Introduction

L'habitat est un lieu de grande importance pour tout et pour chacun, de sa nature il s'agit du lieu où l'on reste et on revient. Tous les individus, et en particulier les personnes âgées, passent beaucoup de leur temps à domicile, d'où l'influence considérable de la maison sur la qualité et la nature de la vie. L'amélioration du sentiment de confort et de sécurité dans l'habitat apparaît donc assez importante de point de vue social. Il n'y a pas longtemps, l'informatique a été appliquée à la création d'habitats intelligents afin d'améliorer les conditions de vie des gens lorsqu'ils sont à leur domicile et leurs offrir un contrôle distant fiable. Une telle maison est une résidence équipée de technologies d'informatique ambiante visant à assister l'habitant dans les situations diverses de la vie domestique. Les dites maisons intelligentes augmentent le confort de l'habitant à travers, des interfaces naturelles pour piloter l'éclairage, la température ou les différents appareils électroniques [1].

En outre, un autre but essentiel de l'application des technologies d'information aux habitats est la protection des individus. Cela est devenu possible par des systèmes capables d'anticiper et de prévoir des situations potentiellement dangereuses ou de réagir aux événements mettant en danger l'habitant. Les bénéficiaires de telles innovations peuvent être des individus autonomes mais également des personnes plus ou moins fragiles ayant une capacité de mouvement limitée. Les systèmes intelligents sont capables de rappeler entre autres aux habitants la prise de leurs médicaments, faciliter leur mise en communication avec l'extérieur ou même alerter les proches ou les services d'urgence [1].

Cela a donné naissance à une nouvelle informatique nommée « ambiante » ou « ubiquitaire ».

Dans ce chapitre, nous présentons l'historique de l'intelligence ambiante puis nous expliquons la définition de l'AmI, Les technologie et nous concluons son application et ses problèmes

1.1 Historique

Le terme « ubiquitous computing » (informatique ubiquitaire) a été introduit par Mark Weiser, chercheur à Xerox PARC3, pour désigner sa vision de l'ordinateur du 21^{ème} siècle. Weiser constate que les technologies ancrées dans nos activités quotidiennes sont celles qui savent s'y fondre, jusqu'à disparaître. Il illustre son propos avec l'exemple de l'écriture : omniprésente dans nos sociétés modernes, chacun l'utilise au quotidien sans même y prêter attention. Comparé à l'écriture, l'ordinateur actuel, bien que très répandu, est loin d'être aussi intégré dans le tissu de nos activités [2].

Selon Weiser, l'ordinateur d'aujourd'hui ne constitue qu'une étape vers l'informatique ubiquitaire dont l'ancrage dans le quotidien serait tel qu'il rendrait continuellement des services indispensables sans que personne ne remarque sa présence [2].

« Pervasive Computing » (informatique diffuse), apparu au milieu des années 90, est une initiative essentiellement industrielle marquée par l'implication d'IBM. L'accent est plutôt mis sur les aspects techniques avec notamment, le développement des supports matériels et logiciels nécessaires à la concrétisation de la vision de Weiser. Les termes « Ambient Intelligence » et « Disappearing Computer » apparaissent avec le 5^{ème} PRCD de la communauté européenne. Cette période est fortement marquée par l'initiative de Philips Research qui lance alors le projet « Vision of the Future » que concrétise un laboratoire d'expérimentations : le Philips Home-lab. Il s'agit de stimuler la créativité, d'explorer de nouvelles opportunités par la convergence de technologies, d'identifier la signification socioculturelle de ces inventions ; en somme, rendre les concepts tangibles, utiles et accessibles à tous [2].

1.2 Définitions

Actuellement, il n'existe pas une définition claire et précise car les avancées sont en perpétuelles évolutions dans les domaines de la recherche, la technologie, la société, etc. Les différentes définitions et caractéristiques de l'AmI sont résumées dans le tableau 1 [3]:

Chapitre 01 :

Définitions	Caractéristiques					
	S	R	A	T	U	I
L'intelligence ambiante s'insère dans une vision futuriste des télécommunications et d'informatisation de la vie. Notre environnement futur sera entouré par différents systèmes et applications qui seront portés par les technologies réseaux et l'informatique pour fournir une aide intelligente sans intrusion humaine.				X	X	X
l'intelligence ambiante est une mise en relation avec des systèmes de services, c'est-à-dire, avec des technologies pour automatiser des actions et avec des dispositifs pour personnaliser et adapter leurs comportements. Aujourd'hui on fabrique des objets de plus en plus petits qu'on peut intégrer dans les appareils. Ceci accroît leur utilisation et entraîne de nouvelles recherches. De plus, la recherche et le progrès en communication enrichissent le développement et l'utilisation de l'intelligence ambiante. Le succès de l'AmI dépend en grande partie du développement de la technologie de capteurs, d'actuateurs, etc. Et également des logiciels et de leur intelligence pour la prise de décision	X		X			
l'intelligence ambiante est une technologie qui consiste en la création de contextes intelligents. Ces contextes doivent s'adapter aux besoins et aux goûts des utilisateurs. Cette technologie doit aider à développer les tâches de tous les jours : au travail, à la maison, sur les lieux de loisirs où les utilisateurs interagissent avec normalité et sans efforts avec des systèmes composés de beaucoup de dispositifs quotidiens. De plus, la technologie est intégrée dans des objets communs et elle est invisible. Par conséquent, la technologie s'adapte aux utilisateurs et de manière autonome.	X		X	X	X	

Chapitre 01 :

l'intelligence ambiante est une nouvelle technique qui va rendre notre environnement quotidien de plus en plus sensibles et réceptifs à notre présence.	X	X			X	
Elle est un futur potentiel dans lequel nous serons entourés par des objets intelligents et dans lequel l'environnement reconnaît la présence de personnes et d'y répondre d'une manière indétectable.	X	X				X
« L'intelligence ambiante » implique l'intelligence qui est tout autour de nous.						X
C'est la présence d'un environnement numérique qui est sensible, adaptatif, et réceptif à la présence de personnes.	X	X	X			
c'est une vision futuriste de la vie quotidienne. Elle contient l'hypothèse que la technologie intelligente devrait disparaître dans notre environnement pour apporter aux êtres humains une vie facile et amusante.				X		X
L'AmI est un nouveau domaine de recherche pour les systèmes distribués, non intrusifs, et des logiciels intelligents.					X	X
Dans un environnement de l'intelligence ambiante, les gens sont entourés par des réseaux intégrés de dispositifs intelligents capables de détecter leurs états, anticiper, et peut-être s'adapter à leurs besoins.	X		X		X	X
L'intelligence ambiante est un environnement numérique qui soutient les personnes dans leur vie quotidienne de manière non intrusive				X	X	
C'est un environnement numérique qui aide les gens dans leurs vies quotidiennes d'une manière proactive mais sensible.	X	X			X	
Milieu ayant la faculté de percevoir, de raisonner, d'agir et d'interagir afin de fournir des services améliorant la qualité de vie des êtres vivants et notamment des personnes.	X	X	X			X

Tableau 01 : Les différentes définitions et caractéristiques de l'AmI. Sensible(S), Réceptive (R), Adaptative (A), Transparente (T), Ubiquitaire (U) et Intelligente (I) [3].

Chapitre 01 :

D'après toutes les définitions présentées, on peut tirer les caractéristiques de l'intelligence suivantes :

- **L'ubiquité:** Un environnement dans lequel les ordinateurs et les réseaux sont « enfouis », « intégrés » et « omniprésents » dans le monde réel ;
- **La sensibilité :** La faculté du système de connaître et de percevoir la localisation des objets, des appareils et des personnes au moyen des capteurs pour établir le contexte ;
- **L'interaction naturelle :** L'interaction doit être intuitive et naturelle puisqu'on utilise des appareils utilisés quotidiennement ;
- **L'intelligence :** L'aptitude d'analyser le contexte perçu et l'ajustement dynamique aux utilisateurs et aux situations pour trouver une bonne réponse ;
- **La réceptivité :** L'aptitude à recevoir des informations ;
- **La transparence :** La capacité d'interaction entre l'homme et la machine de façon invisible, adéquate et personnelle ;
- **L'adaptabilité :** Les dispositifs peuvent changer en fonction des demandes des utilisateurs. En résumé l'AmI, peut être définie ainsi « une vision futuriste sur la vie quotidienne de l'homme, où notre environnement sera dominé par des systèmes hétérogènes et des applications capables de coopérer, d'interagir, d'anticiper, pour la résolution des problèmes pouvant créer des contraintes pour les utilisateurs, tout en s'adaptant aux changements environnementaux. ». Grâce à elle, l'environnement nous obéit désormais au doigt et à l'œil [3].

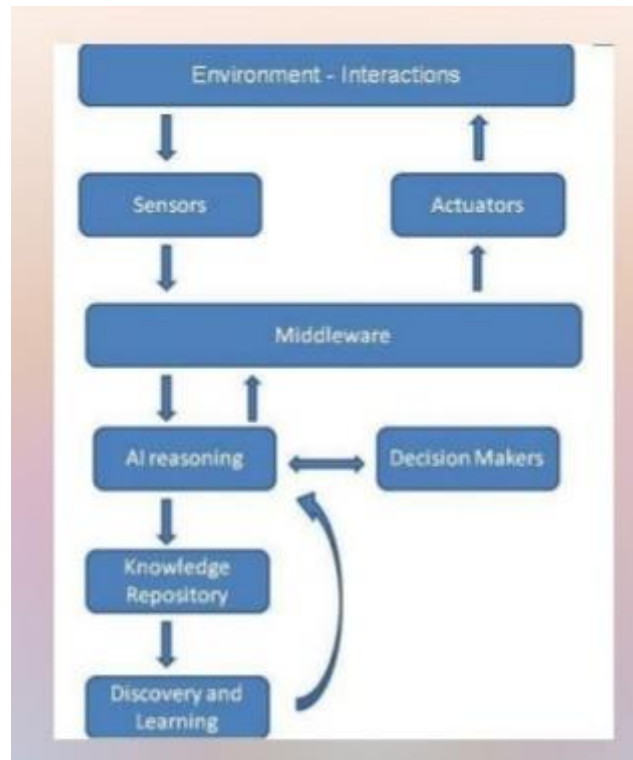


Figure 01 : architecture générique d'AmI [4].

1.3 Les technologies de L'AmI

L'intelligence ambiante s'appuie sur les concepts et les technologies récents :

1.3.1 L'informatique ubiquitaire

L'informatique ubiquitaire désigne le fait que l'informatique est omniprésente. C'est le modèle qui suit l'ordinateur personnel au niveau de l'interaction homme-machine dans lequel le traitement de l'information a été complètement intégré dans tous les objets des activités journalières. Par opposition au paradigme du PC, dans lequel un seul utilisateur engage consciemment un dispositif unique dans un but spécialisé, quelqu'un « qui utilise » l'informatique ubiquitaire engage beaucoup plus de dispositifs et de systèmes informatiques simultanément, au cours de ses activités ordinaires, et n'est pas nécessairement averti qu'il en est ainsi [5].

1.3.2 Communication omniprésente

Permettre aux objets de communiquer entre eux et avec l'utilisateur au moyen d'un réseau sans fil [6].

1.3.3 Interface utilisateur intelligente

Une interface utilisateur intelligente implique que le côté ordinateur ait une connaissance sophistiquée du domaine et / ou d'un modèle de l'utilisateur. Ceux-ci permettent à l'interface de mieux comprendre les besoins de l'utilisateur et de personnaliser ou guider l'interaction. Permettre à les habitants de l'environnement AmI de contrôler et d'interagir avec l'environnement de manière naturelle (voix, geste) et personnalisée (préférences, contexte) [6].

1.3.4 Bluetooth Low Energy

Bluetooth LowEnergy ou Bluetooth LE, commercialisé sous le nom de Bluetooth Smart, est une technologie de réseau personnel sans fil conçue et commercialisée par le groupe d'intérêt spécial Bluetooth visant des applications novatrices dans les secteurs de la santé, du fitness, des balises, de la sécurité et du divertissement à domicile. Par rapport au Bluetooth classique, Bluetooth Smart est conçu pour fournir une consommation d'énergie et un coût considérablement réduits tout en conservant une portée de communication similaire [7].

1.3.5 Identification par radio fréquence

L'identification par radiofréquence (RFID) est l'utilisation d'un système sans fil sans contact qui utilise des champs électromagnétiques radiofréquences pour transférer des données d'une étiquette attachée à un objet, à des fins d'identification et de suivi automatiques [7].

1.3.6 Implant Microchip (humain)

Un implant à micro puce humaine est un dispositif à circuit intégré ou un transpondeur RFID enfermé dans du verre de silicate et implanté dans le corps d'un être humain. Un implant sous-cutané contient généralement un numéro d'identification unique qui peut être lié à des informations contenues dans une base de données externe, telles que l'identification personnelle, les antécédents médicaux, les médicaments, les allergies et les informations de contact [7].

1.3.7 Capteur

Un capteur (également appelé détecteur) est un convertisseur qui mesure une quantité physique et la convertit en un signal qui peut être lu par un observateur ou par un instrument (aujourd'hui principalement électronique). Par exemple, un thermomètre à mercure dans le verre convertit la température mesurée en expansion et contraction d'un liquide qui peut être lu sur un tube de verre calibré. Un thermocouple convertit la température en une tension de sortie qui peut être lue par un voltmètre. Pour plus de précision, la plupart des capteurs sont étalonnés par rapport à des normes connues [8].



Figure 02 : Technologies utilisées dans AmI .

1.4 Les étapes principales de L'AmI

Les étapes principales de l'AmI sont montrées dans la figure 3 ci dessous :

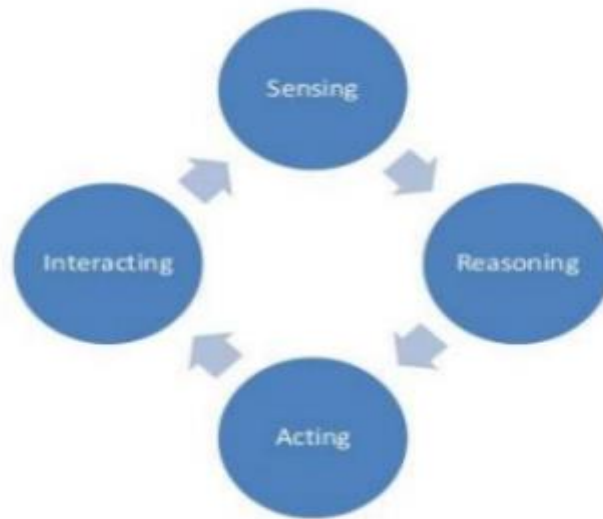


Figure 03 : Les étapes principales de l'AmI [6].

1.4.1 Détection(Sensing) :

- AmI est créé pour le monde réel, une utilisation efficace des capteurs est vitale.
- Les capteurs sont la clé qui relie la puissance de calcul disponible à l'application physique.

1.4.2 Raisonnement (Reasoning)

- Interpréter et reconnaître le contexte et l'activité.
- Pour rendre l'algorithme réactif, adaptatif et bénéfique pour les utilisateurs.
- La modélisation.
- Prédiction et reconnaissance d'activité.
 - Données du capteur.
 - Modèle d'activité.

1.4.3 Agissant (Acting)

- Les systèmes AmI lient le raisonnement au monde réel par la détection et l'action.
- Appareils intelligents et fonctionnels.
- Les robots.

1.4.4 Interagir avec les utilisateurs (Interacting with users) [6]

- Interfaces utilisateur traditionnelles: web, mobiles.
- Luminaires à domicile.
- Interface utilisateur naturelle (parole, geste, suivi des mouvements corporels, émotions, expressions faciales, attention).

1.5 Application de L'AmI

Les différentes applications où l'intelligence ambiante peut être utilisée :

- **Smart Homes**

L'intelligence ambiante permet à la maison elle-même de posséder de l'intelligence et de prendre des décisions concernant son état et ses interactions avec ses habitants. La spécification AmI peut inclure l'environnement significatif est la maison, y compris la cour arrière et une partie de la porte d'entrée car ces zones ont également des capteurs [7].

- **Applications liées à la santé**

Les hôpitaux peuvent accroître l'efficacité de leurs services en surveillant la santé et les progrès des patients en effectuant une analyse automatique des activités dans leurs chambres. Les services de santé dans les hôpitaux pourraient utiliser un lien vers le domicile des patients via un téléphone mobile vers des capteurs de santé, par exemple, lors de la récupération après une opération ou pendant le traitement à long terme d'une maladie, par ex. cancer [7].

- **Secteur des transports publics**

Les transports publics peuvent bénéficier de technologies supplémentaires, notamment des services par satellite, la localisation spatiale basée sur le GPS, l'identification des véhicules, le traitement d'images et d'autres technologies pour rendre les transports plus fluides et donc plus efficaces et plus sûrs. En fonction du temps nécessaire pour connecter deux emplacements avec des capteurs, le système peut également prédire la vitesse de chaque unité. Des exemples d'objets dans cet environnement sont les pistes et les stations. Les

intervenants sont des trains, des chauffeurs et des officiers du centre de commandement. Les capteurs sont utilisés à des fins d'identification sur la base des signaux d'identification envoyés par le train. D'autres signaux peuvent également être envoyés, par exemple l'état d'urgence. Les actionneurs seront des signaux coordonnant le flux des trains et des messages qui pourront être délivrés à chaque unité afin de réguler leur vitesse et le temps qu'ils doivent passer à un arrêt. Les contextes d'intérêt peuvent être des «retards» ou «train arrêté». Une règle d'interaction peut être «si la ligne est bloquée devant et qu'il y a des arrêts intermédiaires, décrivez la situation aux passagers» [7].

- **Services éducatifs**

Les établissements d'enseignement peuvent utiliser la technologie pour créer des salles de classe intelligentes où les modes d'apprentissage sont améliorés. Les objets dans une salle de classe ou un terrain de jeu sont des tables et d'autres éléments disponibles. Les interacteurs sont des étudiants et des enseignants. Les capteurs identifieront qui utilise quel kit scientifique et cela permettra à son tour de surveiller la durée pendant laquelle les étudiants sont impliqués dans une expérience particulière. Les actionneurs peuvent être des recommandations délivrées à des affichages personnalisés de type montre-bracelet. Les contextes d'intérêt peuvent être «l'élève utilise un seul kit d'expérimentation depuis trop longtemps» ou «l'élève ne s'est pas engagé dans une expérimentation active». Le premier contexte déclenchera une règle «si l'étudiant a interagi avec un seul kit pendant plus de 20 minutes, conseillez-lui d'essayer la prochaine expérience disponible» tandis que le second peut envoyer un message à un tuteur, par exemple «si l'étudiant a pas engagé pendant plus de 5 minutes avec une expérience, le tuteur doit encourager et guider l'élève » [7].

1.6 Problèmes de L'AmI

Parmi les problèmes rencontrés dans le domaine de l'AmI, on cite par exemple [6]:

- Confidentialité et sécurité.
- Interopérabilité.
- Fiabilité.
- Contrainte de ressources.

1.7 **Avantage de L'AmI**

L'AmI offre beaucoup d'avantage à la société, par exemple on a [9] :

- Réduire l'effort humain.
- Information, connectivité et loisirs accrus.
- Futur prêt ambiant.

Conclusion :

L'introduction d'AmI dans l'environnement domestique aura une impatience sur la vie personnelle de plusieurs manières. Le temps gagné permettra aux gens de passer plus de temps avec leur famille et leurs amis.

La commodité, l'argent, le gain de temps, la sécurité, la sûreté et les divertissements réduisent le stress menant à une meilleure qualité de vie globale. -interaction face à face entre personnes ou, au moins, à une interaction sélective restreinte principalement à la famille et aux amis.

Chapitre 02

Introduction

Les données brutes, malgré leur quantité qui augmente d'une façon exponentielle, n'ont presque aucune valeur, ce qui est le plus important en fait c'est les connaissances pour lesquelles nous sommes tous assoiffés et qui sont obtenus par la compréhension de ces données, mais plus on a de données plus ce processus devient difficile [10].

De nos jours, les changements de notre environnement sont dénotés par des capteurs qui sont devenus de plus en plus nombreux. Par conséquent, la compréhension de ces données est très importante. Et comme il est dit par Piatestky-Shapiro, « [...] as long as the world keeps producing data of all kinds [...] at an ever increasing rate, the demand for Data Mining will continue to grow ». D'où la fouille de données devient une nécessité [10].

L'expression de « Data Mining » est apparue en 1960, à cette époque, elle a un sens péjoratif. Le développement des moyens informatiques et du calcul a permis le stockage d'une grande masse d'informations. Le traitement et l'analyse de celle-ci nécessitant des méthodes sophistiquées, ce qui a conduit les chercheurs d'utiliser des méthodes d'autres domaines, tel que : les statistiques, les mathématiques, l'intelligence artificielle, etc. afin de mieux exploiter ces gros volumes de données. Jusqu'au milieu des années 90, le « Data Mining », trouve sa place dans le champ scientifique [11].

Dans cette chapitre, nous présentons la définition de Data Mining puis nous expliquons le processus de Data Mining, l'objectif de Data Mining et nous concluons ses tache et ses techniques, pour identifier les technique appropriées pour l'utiliser dans la résolution des problèmes trouvés dans la premier chapitre.

2.1. Qu'est-ce que le Data Mining?

Le Data Mining est un nouveau champ situé au croisement de la statistique et des technologies de l'information (bases de données, intelligence artificielle, apprentissage etc.) dont le but est de découvrir des structures dans de vastes ensembles de données [12].

2.2. Définition de Data Mining

Le data mining, ou fouille de données, est l'ensemble des techniques et des méthodes intelligentes permettant l'extraction des connaissances, auparavant inconnues, à partir d'une grande masse de données. Il s'agit de recherche aux profonds de ces données visant à découvrir des informations cachées, afin de rendre ces quantités de données sous forme des modèles (une connaissance qui concerne la totalité des données et on peut l'appliquer à chaque nouvelle donnée) ou patterns (une connaissance qui concerne une partie des données et on ne peut pas l'appliquer à des nouvelles données). Ainsi on trouve dans la littérature plusieurs définitions de la fouille de données, parmi elles [11]:

- C'est un processus de recherche d'information dans des bases de données gigantesque
- Exploration est analyse des données par des moyens automatique et semi-automatique pour la découverte de modèles de données ou la découverte des connaissances.
- Un processus itératif et interactif de découverte des modèles valides, nouveaux, utiles, est compréhensible dans une grande base de données.
- Est l'étape cœur dans le processus d'Extraction de Connaissances dans les Bases de Données (ECD).

Le terme Data Mining est plus populaire que le terme ECD (extraction des connaissances à partir des données). Beaucoup, font la confusion entre le concept de Data Mining et celui de l'ECD, ils les considèrent comme synonymes. Le Data Mining est l'élément essentiel dans le processus d'extraction des connaissances à partir des données [11].

2.3. Processus ECD

L'extraction de Connaissances dans les Bases de Données (E.C.D.) est une discipline récente, à l'intersection des domaines des bases de données, de l'intelligence artificielle, de la statistique, des interfaces homme / machine et de la visualisation. A partir de données collectées par des experts, il s'agit de proposer des connaissances nouvelles qui enrichissent les interprétations du champ d'application, tout en fournissant des méthodes automatiques qui exploitent cette information [13].

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

2.3.1 Définition

Frawley et al. Indiquer que la découverte des connaissances est l'extraction d'informations non triviale implicite, auparavant inconnues, potentiellement utiles à partir des données. Afin d'obtenir ces informations, nous essayons de créer des modèles, parmi ceux erronés et d'autres utiles. Pour choisir les modèles intéressants, il faut l'examiner dans un programme. Les modèles intéressants sont appelés connaissances, et le résultat du programme est appelé découverte de connaissance [11].

Fayyad et al. ont défini l'ECD comme «le processus non trivial d'identification, à partir de données, de patterns valides, nouveaux, utiles et compréhensibles». D'après cette définition, on conclut que l'ECD est un processus qui comporte plusieurs étapes débutant par parcourir toute la base de données ou une partie de celle-là afin de décrire des modèles ou patterns ou découvrir des relations entre les données. Le but est de rendre les informations plus lisibles et compréhensibles, pour découvrir des nouvelles connaissances qui sont utiles à la prise de décisions [11].

Une autre définition dans annonce que : « l'ECD vise à transformer des données (volumineuses, multiformes, stockées sous différents formats sur des supports pouvant être distribués) en connaissances. Ces connaissances peuvent s'exprimer sous forme d'un concept général qui enrichit le champ sémantique de l'utilisateur par rapport à une question qui le préoccupe. Elles peuvent prendre la forme d'un rapport ou d'un graphique. Elles peuvent s'exprimer comme un modèle mathématique ou logique pour la prise de décision. Les modèles explicites quelle que soit leur forme, peuvent alimenter un système à base de connaissances ou un système expert ». Cette définition est plus détaillée que la précédente et où elle introduit un nouveau concept qui est celui de l'utilisation des connaissances extraites, par un système de gestion des connaissances [11].

2.3.2 Etapes de processus

L'ECD Peut être vu comme un processus en cinq étapes [11]:

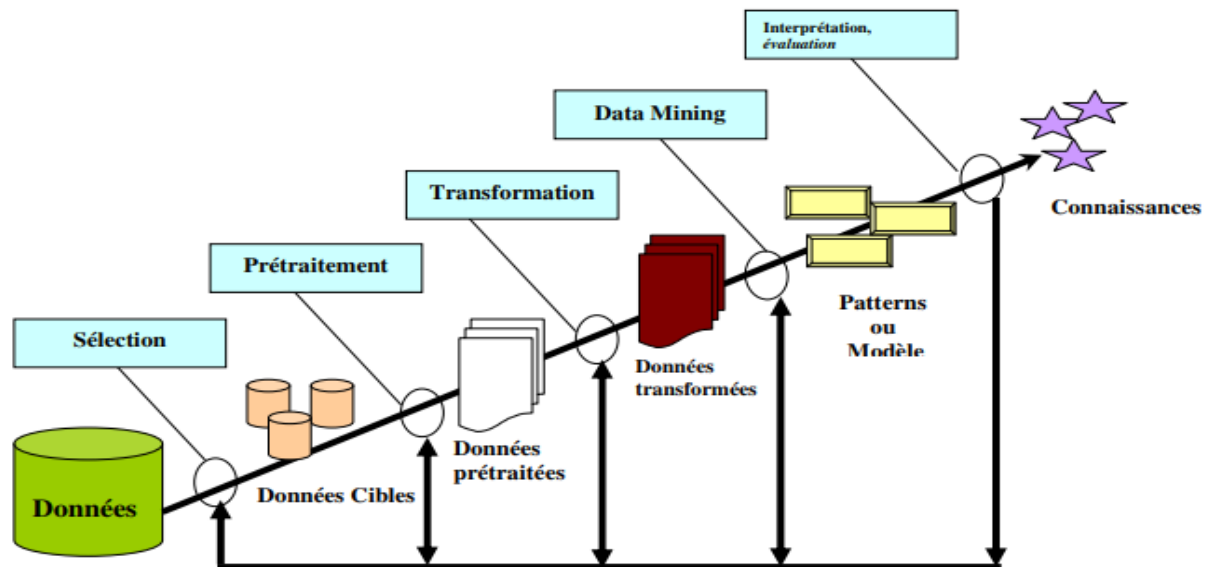


Figure 04: Les étapes de processus ECD [11].

a. La sélection des données

L'objectif de l'extraction de connaissance est de déduire des nouvelles connaissances valides et utiles. Cet aspect est très important, mais on ne peut pas appliquer le processus d'ECD sur toutes les données qu'on a, donc le besoin est exprimé par l'utilisateur. Ce dernier fait la sélection des données selon l'objectif visé. Par exemple, si on veut extraire ou trouver des relations entre les produits vendus par une entreprise commerciale, il est inutile de consulter les données du personnel qui travaillent dans l'entreprise, ou encore de télécharger des pages Web qui parlent de Marketing; il s'agit d'explorer les données transactionnelles archivées, concernant les achats des clients.

Cette étape concerne donc le filtrage des données qui comprend deux opérations nécessaires :

- ✓ La réduction de la dimensionnalité des données :

L'élimination d'attributs sans intérêt, ou ayant beaucoup de valeurs erronées ou manquantes.

- ✓ La réduction de la taille des données :

L'application des techniques du Data Mining est très coûteuse en terme de temps CPU et d'espace mémoire, c'est pour cela que l'on ne peut pas l'appliquer sur la totalité des données. La réduction des données peut également être faite par des techniques statistiques d'échantillonnage.

b. Le prétraitement des données

Le rôle de cette étape est de préparer les données afin qu'ils soient de meilleures qualités afin d'arriver à des résultats de qualité.

Le prétraitement des données concerne, entre autres, le nettoyage des données, c'est-à-dire l'élimination du bruit, ainsi que le traitement des valeurs manquantes, ou erronées. Il faudrait alors définir les méthodes à utiliser pour le remplacement de ces valeurs. De nombreuses solutions existent pour ce problème. On peut, par exemple, remplacer les valeurs manquantes par la valeur la plus fréquente de l'attribut en question, ou l'on estime ces valeurs à partir des enregistrements complets à travers la régression ou les réseaux de neurones. Par exemple dans une valeur est considérée comme erronée si elle s'écarte de la moyenne de deux fois l'écart type.

c. La transformation de données

Cette étape consiste à préparer les données brutes, et à les convertir en données appropriées. La transformation se fait par attribut, c'est-à-dire toutes les valeurs d'un attribut doivent être transformées en un format unique. Formellement, un attribut A est transformé en A' qui serait utilisable par la tâche de la fouille de données choisie.

Un exemple connu de transformation est la discrétisation de variable continue. Il s'agit de transformer un attribut continu en divisant son domaine en intervalles finis. Ainsi, le domaine de l'attribut transformé devient un ensemble de valeurs discrètes. Il y a beaucoup de méthodes de discrétisation dans la littérature.

L'agrégation de données est un autre type de transformation. L'agrégat d'un attribut est la transformation de ce dernier par une règle ou équation. Imaginons, que l'on veut analyser les salaires annuels des employés, et que l'on dispose seulement des salaires mensuels. Un nouvel attribut agrégat serait le salaire multiplié par douze.

d. Fouille de données (Data Mining)

Dans cette étape, des méthodes intelligentes sont utilisées afin d'extraire des modèles ou patterns. Cette étape est aussi désignée comme l'étape cœur du processus d'ECD. Il est clair que les étapes qui précèdent la fouille de données sont très importantes, car la qualité des modèles ou patterns extraites, ainsi que leur coût d'extraction sont liés directement à ces étapes.

e. Evaluation et interprétation des connaissances

Les modèles ou patterns extraits ne sont pas dans la plupart du temps exploitables. En effet, il est difficile d'avoir directement des connaissances valides et utiles, à ce point là. Il existe, cependant, des méthodes d'évaluation des modèles extraits. Ces méthodes peuvent aussi aider à corriger les modèles, et à les ajuster aux données.

Les connaissances obtenues devraient être interprétables, nouvelles, valides et utiles aux utilisateurs. Ces derniers peuvent les utiliser directement, ou les incorporer dans un système de gestion de connaissances. Enfin, cette étape identifie les modèles intéressants qui représentent les connaissances ne se basant pas seulement sur des mesures d'intérêt ou des résultats affichées mais aussi sur l'avis de l'expert.

Ces principales étapes de processus (ECD) visent à partir de données volumineuses l'extraction des connaissances, qui peuvent être exprimés sous forme d'un concept général qui enrichit le champ sémantique de l'utilisateur par rapport à une question qui le préoccupe. Elles peuvent prendre la forme d'un rapport ou d'un graphique. Elles peuvent aussi s'exprimer comme des modèles mathématiques ou logiques. Toutes ces formes des connaissances ont comme but, commun et le plus intéressant, l'aide dans la prise de décision.

2.3.3 Caractéristique d'un processus ECD

Selon Frawley et al. l'ECD expose les principales caractéristiques suivantes [11]:

- **Langage de haut niveau**

La découverte des connaissances est représentée par un langage qui ne doit pas nécessairement être utilisé par les humains, mais son expression doit être compréhensible.

- **Précision**

La mesure de certitude implique de savoir que les modèles découvraient qu'ils sont représenté correctement le contenu d'une base de données ou non.

- **La notion d'intérêt**

la découverte des connaissances est considérée comme intéressante si elle remplit les normes prédéfinies en désignant un modèle intéressant si elle est nouvelle, potentiellement utile et non trivial.

- **Efficacité**

Le temps d'exécution de l'algorithme est acceptable quelque soit le type de traitement.

2.4. Data Mining sur quels types de données ?

Le Data Mining n'est pas spécifique à un type de médias ou de données. Il est applicable à n'importe quel type d'information. Le Data Mining est utilisé et étudié pour les Bases de Données incluant les Bases de Données relationnelles et les Bases de Données Orientées-Objets, les Data Warehouses, les Bases de Données transactionnelles, les supports de données non structurés et semi-structurés comme le World Wide Web, les Bases de Données avancés comme les Bases de Données spatiales, les Bases de Données multimédia, les Bases de données de séries temporelles et les Bases de Données textuelles et même fichiers plats [10].

2.5. Les objectifs du data mining

Nous évoquons dans cette section, trois intérêts du data mining [13] :

- **Expliquer**

Le Data Mining pourrait tenter d'expliquer un événement ou un incident indiscernable. Par la consultation des informations contenues dans l'entrepôt de données de l'organisation.

- **Confirmer**

Le data Mining aidera à confirmer un comportement ou une hypothèse. Dans le cas où le décisionnaire aurait un doute concernant une hypothèse, le data mining pourra tenter de confirmer cette hypothèse en la vérifiant en appliquant des méthodes statistiques ou d'intelligence artificielle.

- **Explorer**

Enfin, le Data Mining peut explorer les données pour découvrir un lien «inconnu». Quand le décisionnaire n'as pas d'hypothèse ou d'idée sur un fait précis, il peut demander au système de proposer des associations ou des corrélations qui pourront aboutir à une explication.

2.6. Les méthodes du Data Mining

Les méthodes de Data Mining, sont appliquées sur les bases de données pour but de construire des modèles d'une base de données afin de décrire le comportement actuel et de prédire les comportements futurs.

On peut classer ces modèles, comme des modèles de calculs ou des modèles logiques: ces modèles sont utilisés pour différentes tâches et pour des traitements différents, menant à la découverte de connaissances, ces tâches sont [11]:

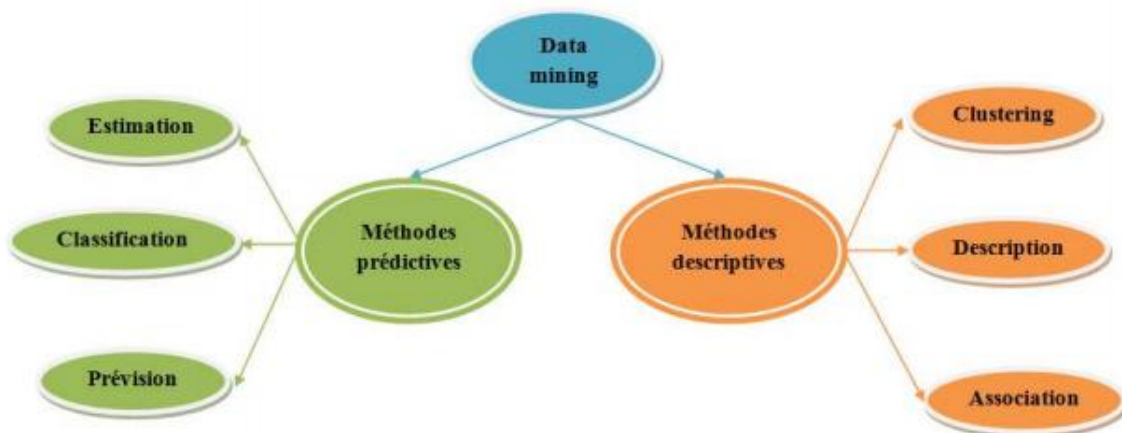


Figure 05 : Graphe représentant les méthodes du Data Mining [14].

2.6.1 Descriptives

Consiste à trouver les caractéristiques générales relatives aux données fouillées [15].

- **Clustering**

Le clustering est une méthode statistique d'analyse de données qui a pour but de regrouper un ensemble de données en différents groupes homogènes. Chaque sous-ensemble regroupe des éléments ayant des caractéristiques communes qui correspondent à des critères de proximité [16].

Le but des algorithmes de clustering est donc de minimiser la distance intra-classe (grappes d'éléments homogènes) et de maximiser la distance inter-classe afin d'obtenir des sous-ensembles le plus distincts possible [16].

La mesure des distances est un élément prépondérant pour la qualité de l'algorithme de clustering. Cette classe d'algorithmes ne sera pas traitée dans le présent document [16].

- **La description**

C'est souvent l'une des premières tâches demandées à un outil de data Mining. On lui demande de décrire les données d'une base complexe. Cela engendre souvent une exploitation supplémentaire en vue de fournir des explications [17].

Il existe plusieurs manières afin de représenter les données à visualiser: L'histogramme, Les statistiques élémentaires, Nuage des points [14].

- **Associations :**

L'association est probablement la technique du fouille des données la plus connue et la plus simple. On fait une corrélation simple entre deux éléments ou plus, souvent du même type, pour identifier les modèles. Par exemple, lors du suivi des habitudes d'achat des gens, on peut identifier qu'un client achète toujours de la crème lorsqu'il achète des fraises, ce qui suggère que, la prochaine fois qu'ils achètent des fraises, il pourrait aussi acheter de la crème [14].

2.6.2 Prédictives

Consiste à utiliser certaines variables pour prédire les valeurs futures inconnues de la même variable ou d'autres variables [15].

- **L'estimation**

Elle consiste à estimer la valeur d'un champ à partir des caractéristiques d'un objet. Le champ à estimer est un champ à valeurs continues. L'estimation peut être utilisée dans un but de classification. Il suffit d'attribuer une classe particulière pour un intervalle de valeurs du champ estimé. Des exemples de tâche d'estimation sont [18] :

- Estimer les revenus d'un client.
- Estimer les risques.

- **La classification**

« La classification consiste à examiner des caractéristique d'un élément nouvellement présenté afin de l'affecter à une classe d'un ensemble prédéfini ». Donc la classification consiste à examiner des caractéristiques d'un objet et lui attribuer une classe, la classe est un champ particulier à valeurs discrètes [17].

La classification peut s'appliquer en utilisant différentes algorithmes de classification, parmi eux:

- ✓ Réseaux de neurones artificiels (ANN).
- ✓ Les arbres de décision.

Les arbres de décisions sont des outils d'aide à la décision qui permettent selon des variables discriminantes de répartir une population d'individus en groupes homogènes en fonction d'un objectif connu. Les arbres de décision sont des outils puissants et populaires pour la classification et la prédiction. Un arbre de décision permet à partir des données connues sur le problème de donner des prédictions par réduction, niveau par niveau, du domaine des solutions [10].

Avantages et inconvénients

Les arbres de décisions ont certains avantages [19], on cite par exemple :

- Les arbres de décision sont capables de produire des règles compréhensibles.
- Les arbres de décision effectuent la classification sans exiger beaucoup de calcul.
- Les arbres de décision sont en mesure de manipuler à la fois les variables continues et catégorielles.

Par contre, parmi leurs inconvénients [19] :

- Manque de performance dans le cas de plusieurs classes; les arbres deviennent très complexes et ne sont pas nécessairement optimaux.
- Demande beaucoup de temps de calcul lors de la construction (le choix du meilleur partitionnement) et l'élagage (la comparaison de sous-arbres).
- Moins bonnes performances concernant les prédictions portant sur des valeurs numériques.

• La prévision

En utilisant des techniques du Data Mining, la prévision implique l'analyse des tendances, la classification et l'association. L'analyse des instances ou événements passés rend possible de prédire un autre événement [14].

Par exemple: Supposons que le responsable du marketing doit prédire combien de clients vont dépenser pendant une vente dans sa compagnie. Dans cet exemple, on prédit une valeur numérique. Par conséquent, la tâche d'analyse des données est un exemple de prédiction

numérique. Dans ce cas, un modèle ou un prédicteur sera construit et qui prédit un point de valeur continue ou ordonné [14].

2.7. Algorithmes du Data Mining

Les techniques et algorithmes de Data Mining diffèrent en fonction des besoins de l'utilisateur, entre autres la tâche à effectuer. Chacune des tâches citées ci-dessus regroupe une multitude d'algorithmes pour construire le modèle auquel elle est associée.

Selon, les dix algorithmes les plus populaires dans le domaine de Data Mining sont, dans l'ordre: C4.5, k-Means, SVM, Apriori, EM, PageRank, AdaBoost, k-NN, NaiveBayes et CART [19].

2.8. Domaines d'application du Data Mining

La technologie de Data Mining a une grande importance économique grâce aux possibilités qu'elle offre pour optimiser la gestion des ressources (humaines et matérielles). Les domaines d'application actuels du Data Mining sont les suivants [10]:

Le secteur bancaire , la bio-informatique et la biotechnologie ,le marketing direct et le collecte de fonds ,la détection de fraude ,la gestion de données scientifiques ,le secteur des assurances ,la télécommunication ,la médecine et la pharmacie ,le commerce au détail ,le e-commerce et le World Wide Web ,le marché boursier et l'investissement ,l'analyse de chaîne d'approvisionnement ,le système d'intelligence ambiante.

Notre choix dans ce mémoire est l'application de Data Mining dans le système d'intelligence ambiante.

2.9. Avantages du Data Mining

Le Data Mining a attiré beaucoup d'attentions dans l'industrie de l'information et dans la société en raison de la grande disponibilité d'énormes quantités de données et du besoin de transformer ces données en informations et connaissances utiles. Les informations et les connaissances acquises peuvent être utilisées pour des applications allant de l'analyse du marché, la détection de fraude, et la rétention des clients, au contrôle de la production et l'exploration scientifique [14].

La fouille de données peut être considérée comme le résultat de l'évolution naturelle de la technologie de l'information.

2.10. Difficultés de Data Mining

La mise en œuvre de Data Mining rencontre trois difficultés principales [19]:

- **Qualité des données**

Les statistiques ont montré que 60% à 70% du temps de travail dans un projet de Data Mining est consacré au prétraitement des données (sélection, correction, transcodage, chargement...), ce qui montre que le temps de préparation est un inconvénient majeur qui influe sur le temps global du projet.

- **Choix des algorithmes et de l'itinéraire du travail**

Pour pouvoir répondre aux questions qui se posent, les algorithmes doivent être choisis en fonction du problème traité. Il faut que l'expert en datamining soit aussi un animateur et possède des qualités que l'on trouve rarement ensemble chez la même personne : rigueur dans la méthode, ouverture et chaleur humaine dans la communication.

- **Evaluation des résultats**

Avant de procéder au déploiement final du modèle, il est important de l'évaluer plus complètement et de passer en revue toutes les différentes étapes exécutées pour construire ce modèle. Ceci permettra d'être certain qu'il permet d'atteindre les objectifs fixés. Lors de la définition du problème, un objectif principal est de déterminer s'il y a un aspect important du problème à résoudre qui n'a pas été suffisamment considéré. A la fin de cette phase, une décision sur l'utilisation des résultats fournis par les outils de Data Mining devrait être prise.

Conclusion

Le Data Mining est l'extraction d'informations prédictives cachées dans de grandes bases de données. C'est un domaine nouveau et puissant qui donne la possibilité aux hommes de se concentrer sur les informations les plus importantes. Les outils du Data Mining peuvent prédire les futures tendances et actions, permettant de prendre les bonnes décisions.

Nous avons exposé, à travers ce chapitre, les principales notions évoquées sur le Data Mining qui n'est qu'une étape de traitement au sein du processus ECD. Ce dernier est né de la nécessité d'englober et de compléter le Data Mining pour que le traitement soit le plus automatisé possible.

Nous avons vu les techniques de Data Mining. Nous avons présenté aussi, les principaux algorithmes de Data Mining couramment utilisés dans le but d'effectuer l'une des principales tâches de fouille de données (le clustering, la classification, la recherche d'associations)...

Chapitre 03

Introduction

Nous allons passer dans ce chapitre à la description de notre conception qui consiste à l'application de Data Mining sur un système d'intelligence ambiant.

Avant de commencer, nous allons présenter la reconnaissance d'activité en temps réel dans une maison intelligente à l'aide de capteurs binaires. Et nous allons aborder l'application de processus ECD. Et nous conclurons une étude des techniques de Data Mining utilisées.

Rappelant que notre problématique traitée dans ce projet de fin d'études est la prédiction du comportement des habitants.

3.1 Reconnaissance d'activité en temps réel dans une maison intelligente à l'aide de capteurs binaires

Aujourd'hui, on peut observer un fort intérêt pour le marché des maisons intelligentes de la part de grandes entreprises, telles que Google, Samsung, Amazon et autres. On s'attend à ce que le marché des maisons intelligentes dépasse les 40 milliards de dollars d'ici 2020. Cependant, malgré la croissance du marché, les maisons intelligentes présentent un certain nombre de limites, notamment l'incapacité d'analyser le comportement des résidents. Pour le moment, il n'existe pas de solutions et d'approches capables de résoudre complètement ce problème [20].

Les applications les plus importantes de la reconnaissance d'activité en temps réel dans un environnement domestique intelligent sont la domotique et les soins de santé. La capacité d'identifier l'activité actuelle nous permet de créer des règles comme «éteindre les lumières en cas d'activité de sommeil», etc. Dans le cas de la santé, il peut être utilisé par les membres de la famille ou les soignants de personnes âgées ou atteintes d'Alzheimer pour surveiller leurs activités quotidiennes afin de détecter des anomalies [20].

Dans cette section, il sera décrit comment nous pouvons analyser l'activité des résidents en utilisant les données du flux d'activations de capteurs binaires [20].



Figure 06 : Maison intelligente [20].

3.1.1 Pourquoi les capteurs binaires

Il existe plusieurs façons de collecter des données sur l'environnement dans les recherches existantes. L'un d'eux utilise des détecteurs de mouvement infrarouges ou des interrupteurs à lames placés sur les portes. En outre, certaines recherches utilisent des données de surveillance par caméras vidéo. Les capteurs binaires sont moins chers, discrets et les modèles nécessaires à la reconnaissance d'activité, en général, sont moins complexes. De plus, les recherches sur la reconnaissance d'activités basées sur la vidéo doivent également surmonter certains défis, tels que les variations d'éclairage, l'occlusion et les changements d'arrière-plan [20].

3.1.2 Dataset

L'ensemble de données du CASAS Smart Home Project (Center for Advanced Studies in Adaptive Systems de l'Université de l'État de Washington) a été utilisé. Nous utiliserons l'ensemble de données nommé Aruba. Les données ont été recueillies au domicile d'une femme âgée bénévole pendant 7 mois. Plan de la maison et emplacement du capteur que vous pouvez voir ci-dessous [20]:

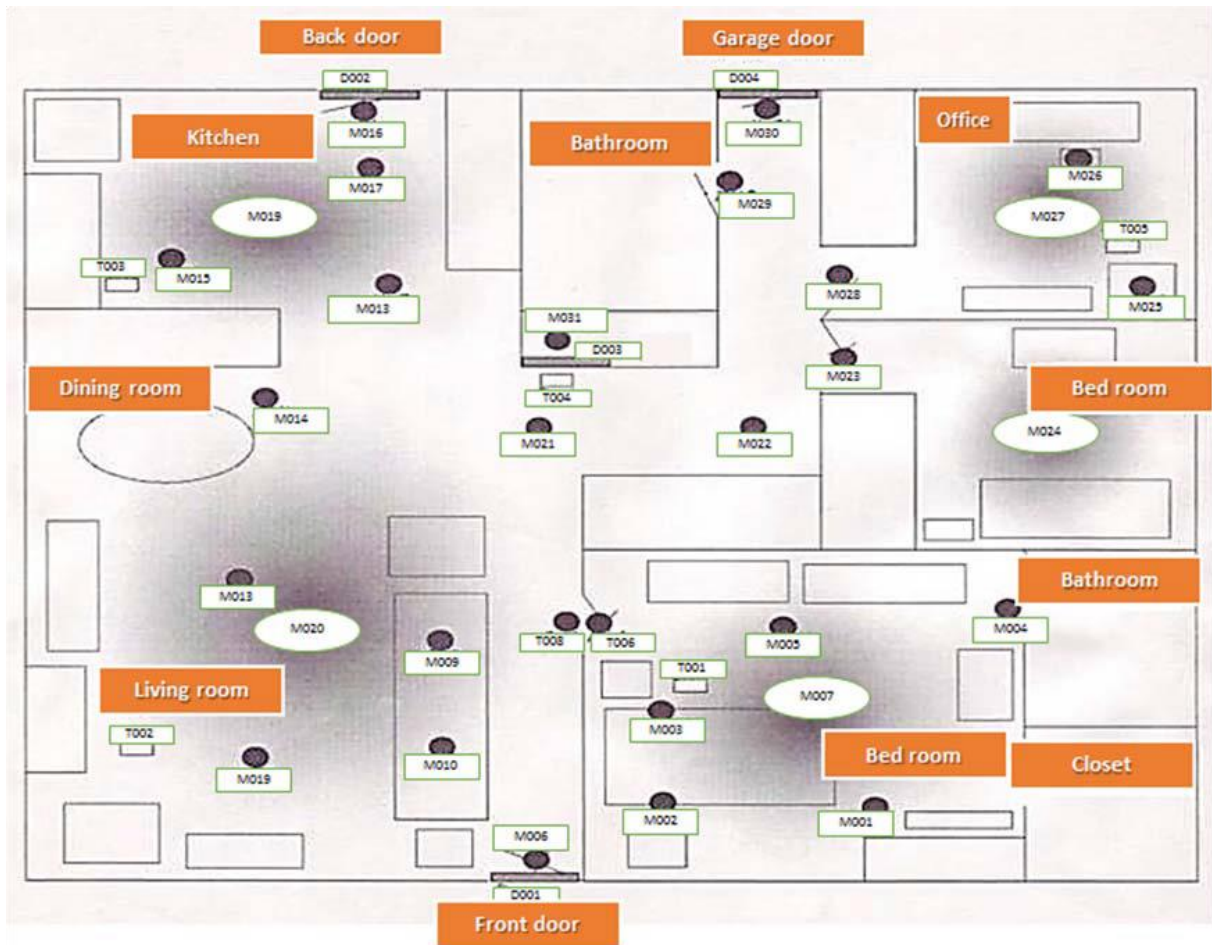


Figure 07: Plan de localisation des capteurs [20].

Il existe trois types de capteurs présentés dans la recherche actuelle: les capteurs de mouvement (ces ID de capteur commencent par «M»), les capteurs de fermeture de porte (ces ID de capteur commencent par «D») et les capteurs de température (ces ID de capteur commencent par «T»). Dans la recherche, les activations des capteurs ont été collectées au fil du temps. Un petit exemple de l'ensemble de données que vous pouvez voir ci-dessous [20]:

2010-11-04	05:40:51.303739	M004	ON	Bed_to_Toilet	begin
2010-11-04	05:40:52.342105	M005	OFF		
2010-11-04	05:40:57.176409	M007	OFF		
2010-11-04	05:40:57.941486	M004	OFF		
2010-11-04	05:43:24.021475	M004	ON		
2010-11-04	05:43:26.273181	M004	OFF		
2010-11-04	05:43:26.345503	M007	ON		
2010-11-04	05:43:26.793102	M004	ON		
2010-11-04	05:43:27.195347	M007	OFF		
2010-11-04	05:43:27.787437	M007	ON		
2010-11-04	05:43:29.711796	M005	ON		
2010-11-04	05:43:30.279021	M004	OFF	Bed_to_Toilet	end
2010-11-04	05:43:45.7324	M003	ON	Sleeping	begin
2010-11-04	05:43:52.044085	M003	OFF		
2010-11-04	05:43:53.185335	M002	ON		
2010-11-04	05:43:53.253809	M003	ON		
2010-11-04	05:43:59.493281	M002	OFF		
2010-11-04	05:44:04.048766	M003	OFF		
2010-11-04	05:44:06.14204	M003	ON		
2010-11-04	05:44:11.229146	M003	OFF		

Figure 08 : exemple de la base de données [20].

Comme vous pouvez le voir, chaque ligne contient des informations sur la date, l'heure et l'identifiant unique. Les deux dernières colonnes contiennent des informations sur le nom de l'activité et son état (début ou fin). Il y a 11 activités uniques dans le jeu de données actuel: Meal_Preparation (1606), Relax (2910), Eating (257), Work (171), Sleeping (401), Wash_Dishes (65), Bed_to_Toilet (157), Enter_Home (431), Leave_Home (431), Ménage (33), Resperate (6). Le nombre entre parenthèses correspond au nombre de fois où l'activité apparaît dans les données. Les activités ont été annotées manuellement par les chercheurs[20]

3.2 L'application de processus ECD

Dans cette section, nous présentons l'enchaînement des phases de processus d'extraction proposé dans le chapitre2, adapté au cas étudié : la prédiction du comportement des habitants.

❖ Analyse du problème

• Comportement des habitants

Dans cette section, nous allons présenter la formation du problème de prévision (entrées, sorties et objectif) et l'algorithme de résolution de Data Mining utilisé. L'intérêt de cette tâche de Data Mining est de analyser le comportement des habitants.

✓ Description du problème

Dans le domaine de l'intelligence ambiante, une attention croissante est consacrée aux préoccupations écologiques au sein de nos sociétés, notamment à la consommation énergétique du bâtiment. L'intelligence ambiante devrait ainsi permettre d'assurer le confort de l'utilisateur tout en réduisant considérablement le coût énergétique par une gestion habile des appareils connectés.

Ceci peut être fait en étudiant le comportement quotidien des utilisateurs du système ambiant par les techniques de Data Mining pour prédire leurs habitudes.

Ce dernier est caractérisé par un ensemble de caractéristiques qui sont : la date de capteur, le temps de capteur, le capteur ...Etc.

Le problème d'analyse le comportement d'un habitant est un problème de classification.

✓ Rechercher les données

La recherche des données consiste dans un premier temps à obtenir des données en accord avec les objectifs que l'on s'impose. Afin de sélectionner les attributs à retenir pour l'analyse des données des habitants, nous sommes inspirés dans la littérature, des types d'attributs utilisés au sein des domaines d'intelligence ambiante. Ces données sont intégrées dans notre base de données.

3.2.1 Sélectionner les données pertinentes

Nous présentons les attributs que nous avons retenus. Les attributs pertinents sont : le temps de capteur, l'activité et la position de l'activité.

3.2.2 Prétraitement de la base de données

Nous avons fait le prétraitement des données dans 2 étapes :

- Détection et réparations manquantes : C'est le terme utilisé pour désigner le cas où des champs ne contiennent aucune donnée.
- Supprimer les valeurs en double.

3.2.3 Transformation des données

Une étape très dépendante du choix de l'algorithme de fouilles de données utilisés. Une fois les variables pertinentes et sélectionner les données fiables, une transformation éventuelle s'impose pour les préparer au travail d'analyse. Il s'agit d'intervenir sur les valeurs des variables pour qu'elles soient mieux exploitables par les algorithmes de traitement.

Nous avons choisie la méthode de discrétisation des attributs numériques : partitionner le domaine des attributs en des intervalles de même longueur.

3.2.4 Application de techniques de Data Mining

Le Data Mining répond à de nombreuses applications en divers domaines et cela permet de tirer profit de la disponibilité croissante de données localisées et de leurs richesses potentielles; c'est le cas de l'analyse de comportement des habitants.

- **Choix d'algorithme**

Pour faire la prédiction dans les systèmes ambiante, il faut choisir un algorithme adéquat à cette tâche.

Nous choisissons l'arbre de décisions (decision tree) qui nous semble intéressante .car comme leurs noms l'indique ils permettent le système de faire de décision basées sur des prédictions.

- Principe de cet algorithme :

Les arbres de décision ciblent la classification (prédiction de variables discrètes). Comme son nom l'indique, cette méthode consiste à construire un arbre. Un enregistrement (qu'on veut classifier) entre par le nœud racine, et passe d'un nœud père à un nœud fils s'il

satisfait une condition posée. Le nœud feuille auquel il arrivera est sa classe. Un arbre de décision peut donc être aperçu comme étant un ensemble de règles qui mènent à une classe. [11]

- Exemple

Observation	Petal length	Stem length	Species
1	2.7 cm	20 cm	A
2	2.6 cm	12 cm	C
3	2.1 cm	21 cm	B
4	1.9 cm	20 cm	B

Data

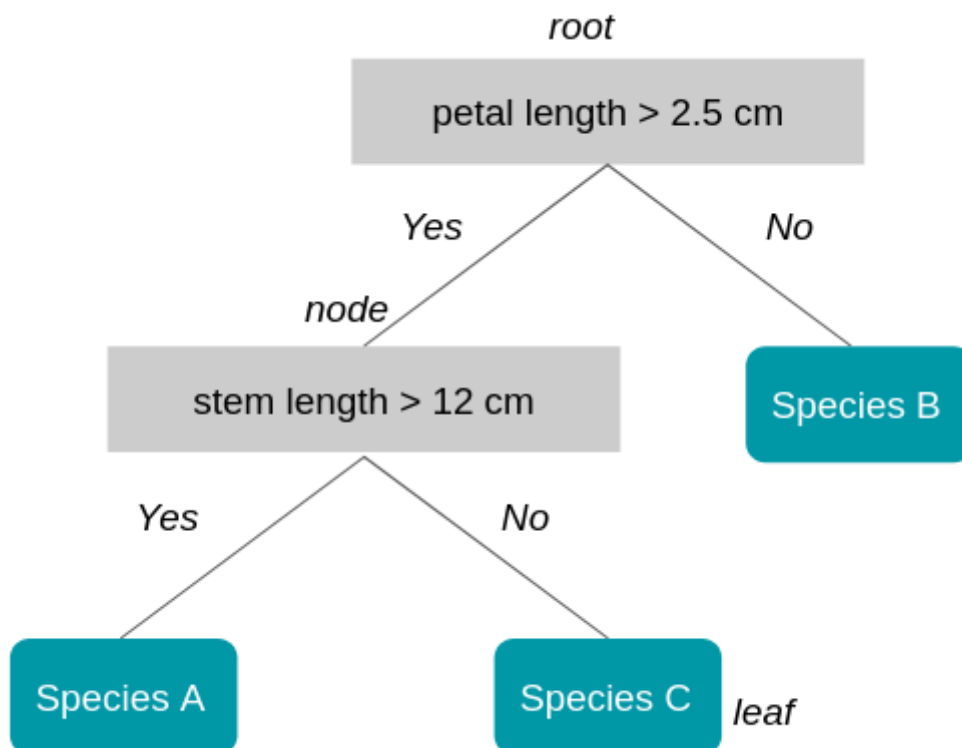


Figure 09 : Exemple de l'arbre de décision [21].

L'arbre a une profondeur de 2 (un nœud plus la racine). La longueur du pétale est la première mesure qui est utilisée car elle sépare le mieux les 4 observations selon l'appartenance aux classes (ici à la classe B) [21].

3.2.5 Évaluer et valider les résultats

Après l'exécution, le résultat est visualisé sous forme arbre. L'évaluation des résultats consiste à mesurer la pertinence du résultat trouvé ; et cela en mesurant l'écart entre les résultats réels et les résultats du modèle.

- **Evaluation et interprétation des connaissances**

Les modèles ou patterns extraits ne sont pas dans la plupart du temps exploitables. En effet, il est difficile d'avoir directement des connaissances valides et utiles, à ce point là. Il existe, cependant, des méthodes d'évaluation des modèles extraits. Ces méthodes peuvent aussi aider à corriger les modèles, et à les ajuster aux données.

Les connaissances obtenues devraient être interprétables, nouvelles, valides et utiles aux utilisateurs. Ces derniers peuvent les utiliser directement, ou les incorporer dans un système de gestion de connaissances. Enfin, cette étape identifie les modèles intéressants qui représentent les connaissances ne se basant pas seulement sur des mesures d'intérêt ou des résultats affichées mais aussi sur l'avis de l'expert.

Ces principales étapes de processus (ECD) visent à partir de données volumineuses l'extraction des connaissances, qui peuvent être exprimés sous forme d'un concept général qui enrichit le champ sémantique de l'utilisateur par rapport à une question qui le préoccupe. Elles peuvent prendre la forme d'un rapport ou d'un graphique. Elles peuvent aussi s'exprimer comme des modèles mathématiques ou logiques. Toutes ces formes des connaissances ont comme but, commun et le plus intéressant, l'aide dans la prise de décision.

Conclusion

Nous avons vu dans ce chapitre quelques notions liées à la reconnaissance d'activité en temps réel dans une maison intelligente à l'aide de capteurs binaires. Puis nous sommes passés à la partie Data Mining.

Nous décrivons dans le prochain chapitre l'architecture technique de notre solution, nous citons les différents outils et technologies que nous avons utilisées.

Chapitre 04

Introduction

Dans ce dernier chapitre nous allons aborder la phase implémentation qui expose les techniques détaillées du système conçu. Ensuite nous allons décrire les ressources utilisées, ainsi que la mise en œuvre de la solution. Nous finirons par définir l'architecture technique de notre système.

4.1 Ressource utilisées

Dans cette section, nous allons présenter les différents outils que nous avons utilisés pour la réalisation de notre solution.

4.1.1 Excel

Microsoft Excel est un logiciel de la suite Microsoft Office. Il est un programme de tableur simple. Il dispose de nombreuses formules susceptibles d'apporter des solutions aussi bien aux particuliers qu'aux entreprises [18].

Bien que cela semble un peu complexe, Excel est assez facile à utiliser et il est important de le maîtriser. Dans certaines offres d'emploi, l'utilisation parfaite des logiciels tels Excel peut être un atout pour avoir le meilleur profil [18].

Excel est un logiciel tableur. Il propose des fonctions de calcul numérique, d'analyse de données, de représentation graphique et de programmation. Le logiciel permet donc d'effectuer des opérations comme des calculs simples tels l'addition ou la soustraction à des calculs plus complexes comme la trigonométrie. Ces différentes fonctions sont nécessaires à la mise en place de solutions diverses tant chez les particuliers que chez les entreprises [18].

4.1.2 Weka

WEKA (Waikato Environment for Knowledge Analysis) est un environnement pour l'analyse de connaissances développé à l'université de Waikato, Nouvelle-Zélande. C'est un logiciel libre disponible sous la licence publique générale GNU (GPL) qui permet de réaliser des analyses en Data Mining. WEKA est écrite en Java et fonctionne sur quasiment tous les systèmes d'exploitation actuels. Il offre une panoplie d'algorithmes d'apprentissage et permet d'appliquer toute la chaîne du processus d'extraction de connaissance à partir de données (prétraitement, classification supervisée, règles d'associations, visualisation, etc...) [10].

La disponibilité du code source permet d'implémenter et de tester des nouveaux algorithmes tout en s'appuyant sur une plate-forme éprouvée et un code objet. Le logiciel comprend plusieurs outils dont un API (Application Programming Interface) qui permet d'utiliser les outils WEKA dans d'autres programmes [10].

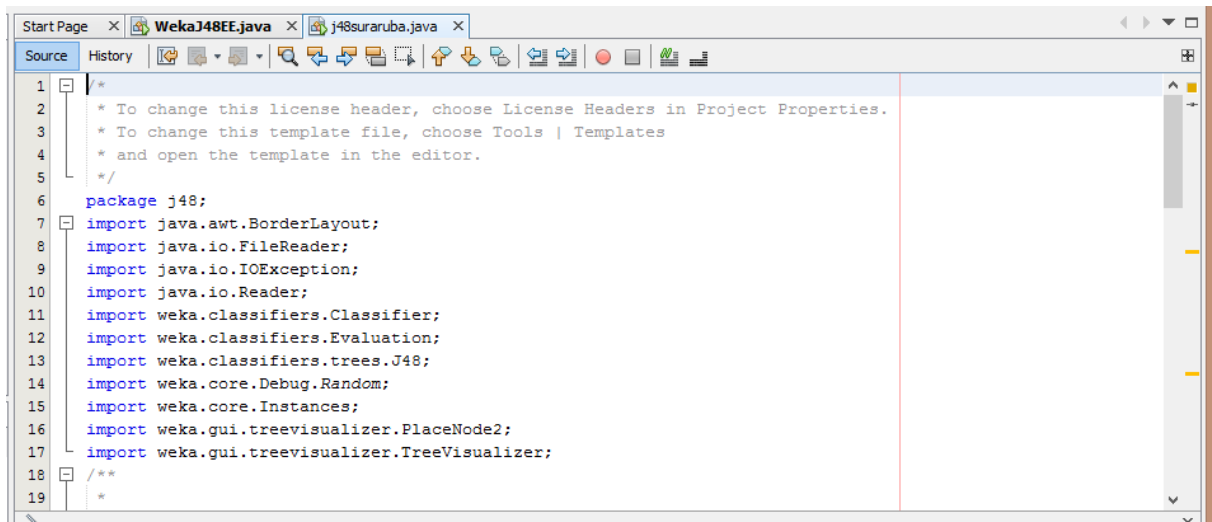
WEKA supporte plusieurs outils d'exploration de données standards, et en particulier, des préprocesseurs de données, des classificateurs, des analyseurs de régression, des outils de visualisation, et des outils d'analyse discriminante. Le format des données d'entrée par défaut de WEKA est ARFF (Attribute Relation File Format). D'autres formats peuvent être importés comme CSV, Binaire, BDD SQL (avec JDBC) à partir d'une URL, etc. WEKA contient plus de 70 algorithmes de classification / régression supervisés, plus de 15 évaluateurs d'attributs et plus de 10 algorithmes de recherche pour la sélection d'attribut, des algorithmes de recherche de règles d'association et plusieurs interfaces graphiques GUI [10].

4.1.3 NetBeans

NetBeans est à l'origine un IDE Java développé par une équipe d'étudiants à Prague, racheté ensuite par Sun Microsystems quelque part en 2002. C'est une plate-forme, qui permet d'écrire des applications Swing ce qui fait de NetBeans une boîte à outils facilement améliorable ou modifiable. La licence de NetBeans permet de l'utiliser gratuitement à des fins commerciales ou non. Elle permet de développer tous types d'applications basées sur la plate-forme NetBeans. Les modules qu'on peut écrire peuvent être open-source comme ils peuvent être closed-source, Ils peuvent être gratuits, comme ils peuvent être payants. Il existe d'autres systèmes de développement rapide mais Netbeans est particulièrement très bien placé [17].

4.1.4 Fichier WEKA. JAR

Le fichier JAR est un fichier pour exposer l'API WEKA pour Netbeans de sorte que nous vous pouvons programmer en toute souplesse qu'avec WEKA. On ajoute le fichier « weka.jar » à notre projet et on va avoir l'accès à toutes les classes de l'API WEKA. Dans la figure suivante, nous présentons les principaux imports de l'API [17].



```
1  | /*
2  |  * To change this license header, choose License Headers in Project Properties.
3  |  * To change this template file, choose Tools | Templates
4  |  * and open the template in the editor.
5  |  */
6  | package j48;
7  | import java.awt.BorderLayout;
8  | import java.io.FileReader;
9  | import java.io.IOException;
10 | import java.io.Reader;
11 | import weka.classifiers.Classifier;
12 | import weka.classifiers.Evaluation;
13 | import weka.classifiers.trees.J48;
14 | import weka.core.Debug.Random;
15 | import weka.core.Instances;
16 | import weka.gui.treevisualizer.PlaceNode2;
17 | import weka.gui.treevisualizer.TreeVisualizer;
18 | /**
19 |  *
```

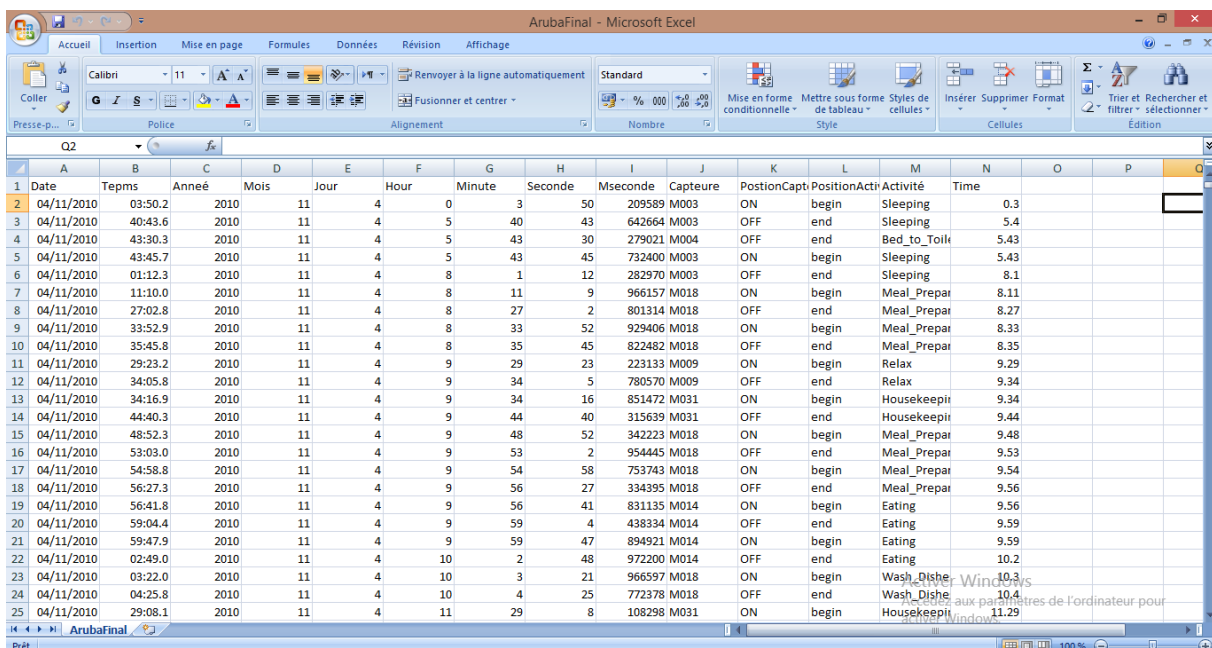
Figure 10 : Les principales classes du fichier weka.jar.

4.2 Mise en œuvre de la solution

4.2.1 Présentation de prétraitement de notre base de données

Le prétraitement de notre base de données s'effectue en 2 étapes essentielles comme suit :

- Détection et réparations manquantes
- Supprimer les valeurs en double



1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2	Date	Tepms	Année	Mois	Jour	Hour	Minute	Seconde	Mseconde	Capteure	PositionCapt	PositionActi	Activité	Time			
3	04/11/2010	03:50.2	2010	11	4	0	3	50	209589	M003	ON	begin	Sleeping	0.3			
4	04/11/2010	40:43.6	2010	11	4	5	40	43	642664	M003	OFF	end	Sleeping	5.4			
5	04/11/2010	43:30.3	2010	11	4	5	43	30	279021	M004	OFF	end	Bed_to_Toilk	5.43			
6	04/11/2010	43:45.7	2010	11	4	5	43	45	732400	M003	ON	begin	Sleeping	5.43			
7	04/11/2010	01:12.3	2010	11	4	8	1	12	282970	M003	OFF	end	Sleeping	8.1			
8	04/11/2010	11:10.0	2010	11	4	8	11	9	966157	M018	ON	begin	Meal_Prepar	8.11			
9	04/11/2010	27:02.8	2010	11	4	8	27	2	801314	M018	OFF	end	Meal_Prepar	8.27			
10	04/11/2010	33:52.9	2010	11	4	8	33	52	929406	M018	ON	begin	Meal_Prepar	8.33			
11	04/11/2010	35:45.8	2010	11	4	8	35	45	822482	M018	OFF	end	Meal_Prepar	8.35			
12	04/11/2010	29:23.2	2010	11	4	9	29	23	223133	M009	ON	begin	Relax	9.29			
13	04/11/2010	34:05.8	2010	11	4	9	34	5	780570	M009	OFF	end	Relax	9.34			
14	04/11/2010	34:16.9	2010	11	4	9	34	16	851472	M031	ON	begin	Housekeepi	9.34			
15	04/11/2010	44:40.3	2010	11	4	9	44	40	315639	M031	OFF	end	Housekeepi	9.44			
16	04/11/2010	48:52.3	2010	11	4	9	48	52	342223	M018	ON	begin	Meal_Prepar	9.48			
17	04/11/2010	53:03.0	2010	11	4	9	53	2	954445	M018	OFF	end	Meal_Prepar	9.53			
18	04/11/2010	54:58.8	2010	11	4	9	54	58	753743	M018	ON	begin	Meal_Prepar	9.54			
19	04/11/2010	56:27.3	2010	11	4	9	56	27	334395	M018	OFF	end	Meal_Prepar	9.56			
20	04/11/2010	56:41.8	2010	11	4	9	56	41	831135	M014	ON	begin	Eating	9.56			
21	04/11/2010	59:04.4	2010	11	4	9	59	4	438334	M014	OFF	end	Eating	9.59			
22	04/11/2010	59:47.9	2010	11	4	9	59	47	894921	M014	ON	begin	Eating	9.59			
23	04/11/2010	02:49.0	2010	11	4	10	2	48	972200	M014	OFF	end	Eating	10.2			
24	04/11/2010	03:22.0	2010	11	4	10	3	21	966597	M018	ON	begin	Wash_Dishe	10.4			
25	04/11/2010	04:25.8	2010	11	4	10	4	25	772378	M018	OFF	end	Wash_Dishe	10.4			
26	04/11/2010	29:08.1	2010	11	4	11	29	8	108298	M031	ON	begin	Housekeepi	11.29			

Figure 11 : La base de données après le prétraitement des données.

Transformation des données

Partitionner le domaine des attributs en des intervalles de même longueur.

Relation: ArubaFinal-weka.filters.unsupervised.attribute.Discretize-B46-M-1.0-R.14-precision1-unset-class-temporality

No.	1: Date Nominal	2: Temps Nominal	3: Année Numeric	4: Mois Numeric	5: Jour Numeric	6: Hour Numeric	7: Minute Numeric	8: Seconde Numeric	9: Mseconde Numeric	10: Capture Nominal	11: Position/Capture Nominal	12: Position/Activité Nominal	13: Activité Nominal	14: Time Nominal
1	04/11/2010	03:50.2	2010.0	11.0	4.0	0.0	3.0	50.0	209589.0	M003	ON	begin	Sleeping	{-inf-0.5]}
2	04/11/2010	40:43.6	2010.0	11.0	4.0	5.0	40.0	43.0	642664.0	M003	OFF	end	Sleeping	{5-5.5]}
3	04/11/2010	43:30.3	2010.0	11.0	4.0	5.0	43.0	30.0	279021.0	M004	OFF	end	Bed_to_Toilet	{5-5.5]}
4	04/11/2010	43:45.7	2010.0	11.0	4.0	5.0	43.0	45.0	732400.0	M003	ON	begin	Sleeping	{5-5.5]}
5	04/11/2010	01:12.3	2010.0	11.0	4.0	8.0	1.0	12.0	282970.0	M003	OFF	end	Sleeping	{8-8.5]}
6	04/11/2010	11:10.0	2010.0	11.0	4.0	8.0	11.0	9.0	966157.0	M018	ON	begin	Meal_Preparation	{8-8.5]}
7	04/11/2010	27:02.8	2010.0	11.0	4.0	8.0	27.0	2.0	801314.0	M018	OFF	end	Meal_Preparation	{8-8.5]}
8	04/11/2010	33:52.9	2010.0	11.0	4.0	8.0	33.0	52.0	929406.0	M018	ON	begin	Meal_Preparation	{8-8.5]}
9	04/11/2010	35:45.8	2010.0	11.0	4.0	8.0	35.0	45.0	822482.0	M018	OFF	end	Meal_Preparation	{8-8.5]}
10	04/11/2010	29:23.2	2010.0	11.0	4.0	9.0	29.0	23.0	223133.0	M009	ON	begin	Relax	{9-9.5]}
11	04/11/2010	34:05.8	2010.0	11.0	4.0	9.0	34.0	5.0	780570.0	M009	OFF	end	Relax	{9-9.5]}
12	04/11/2010	34:16.9	2010.0	11.0	4.0	9.0	34.0	16.0	851472.0	M031	ON	begin	Housekeeping	{9-9.5]}
13	04/11/2010	44:40.3	2010.0	11.0	4.0	9.0	44.0	40.0	315639.0	M031	OFF	end	Housekeeping	{9-9.5]}
14	04/11/2010	48:52.3	2010.0	11.0	4.0	9.0	48.0	52.0	342223.0	M018	ON	begin	Meal_Preparation	{9-10]}
15	04/11/2010	53:03.0	2010.0	11.0	4.0	9.0	53.0	2.0	954445.0	M018	OFF	end	Meal_Preparation	{9-10]}
16	04/11/2010	54:58.8	2010.0	11.0	4.0	9.0	54.0	58.0	753743.0	M018	ON	begin	Meal_Preparation	{9-10]}
17	04/11/2010	56:27.3	2010.0	11.0	4.0	9.0	56.0	27.0	334395.0	M018	OFF	end	Meal_Preparation	{9-10]}
18	04/11/2010	56:41.8	2010.0	11.0	4.0	9.0	56.0	41.0	831135.0	M014	ON	begin	Eating	{9-10]}
19	04/11/2010	59:04.4	2010.0	11.0	4.0	9.0	59.0	4.0	438334.0	M014	OFF	end	Eating	{9-10]}
20	04/11/2010	59:47.9	2010.0	11.0	4.0	9.0	59.0	47.0	894921.0	M014	ON	begin	Eating	{9-10]}
21	04/11/2010	02:49.0	2010.0	11.0	4.0	10.0	2.0	48.0	972200.0	M014	OFF	end	Eating	{10-10.5]}
22	04/11/2010	03:22.0	2010.0	11.0	4.0	10.0	3.0	21.0	966597.0	M018	ON	begin	Wash_Dishes	{10-10.5]}
23	04/11/2010	04:25.8	2010.0	11.0	4.0	10.0	4.0	25.0	772378.0	M018	OFF	end	Wash_Dishes	{10-10.5]}
24	04/11/2010	29:08.1	2010.0	11.0	4.0	11.0	29.0	8.0	108298.0	M031	ON	begin	Housekeeping	{11-11.5]}
25	04/11/2010	33:48.3	2010.0	11.0	4.0	11.0	33.0	48.0	313567.0	M031	OFF	end	Housekeeping	{11-11.5]}
26	04/11/2010	41:34.0	2010.0	11.0	4.0	11.0	41.0	34.0	29848.0	D004	OPEN	begin	Leave_Home	{11-11.5]}
27	04/11/2010	41:43.3	2010.0	11.0	4.0	11.0	41.0	43.0	345957.0	D004	CLOSE	end	Leave_Home	{11-11.5]}
28	04/11/2010	43:30.1	2010.0	11.0	4.0	11.0	43.0	30.0	94537.0	D004	OPEN	begin	Enter_Home	{11-11.5]}
29	04/11/2010	43:34.7	2010.0	11.0	4.0	11.0	43.0	34.0	683398.0	D004	CLOSE	end	Enter_Home	{11-11.5]}
30	04/11/2010	43:53.8	2010.0	11.0	4.0	11.0	43.0	53.0	784103.0	D004	OPEN	begin	Leave_Home	{11-11.5]}
31	04/11/2010	43:58.4	2010.0	11.0	4.0	11.0	43.0	58.0	371827.0	D004	CLOSE	end	Leave_Home	{11-11.5]}
32	04/11/2010	47:48.0	2010.0	11.0	4.0	11.0	47.0	48.0	42367.0	D004	OPEN	begin	Enter_Home	{11.5-11.9]}
33	04/11/2010	47:52.6	2010.0	11.0	4.0	11.0	47.0	52.0	575602.0	D004	CLOSE	end	Enter_Home	{11.5-11.9]}
34	04/11/2010	46:25.6	2010.0	11.0	4.0	14.0	46.0	25.0	648668.0	M009	ON	begin	Relax	{14.4-14.9]}
35	04/11/2010	13:24.3	2010.0	11.0	4.0	15.0	13.0	24.0	311043.0	M009	OFF	end	Relax	{14.9-15.4]}
36	04/11/2010	23:00.0	2010.0	11.0	4.0	15.0	23.0	0.0	12287.0	M018	ON	begin	Meal_Preparation	{14.9-15.4]}
37	04/11/2010	25:33.6	2010.0	11.0	4.0	15.0	25.0	33.0	571046.0	M017	OFF	end	Meal_Preparation	{14.9-15.4]}
38	04/11/2010	25:35.2	2010.0	11.0	4.0	15.0	25.0	35.0	233346.0	M014	ON	begin	Eating	{14.9-15.4]}

Figure 12 : La base de données après transformation des données.

4.3 Capture d'écran

Pour mieux expliquer l'usage de notre système, nous décrivons dans ce qui suit les fonctionnalités à travers les différentes interfaces.

- **Interface accueil**

La figure suivante illustre l'interface principale de notre outil assistant nommé **Comportement des habitants**.

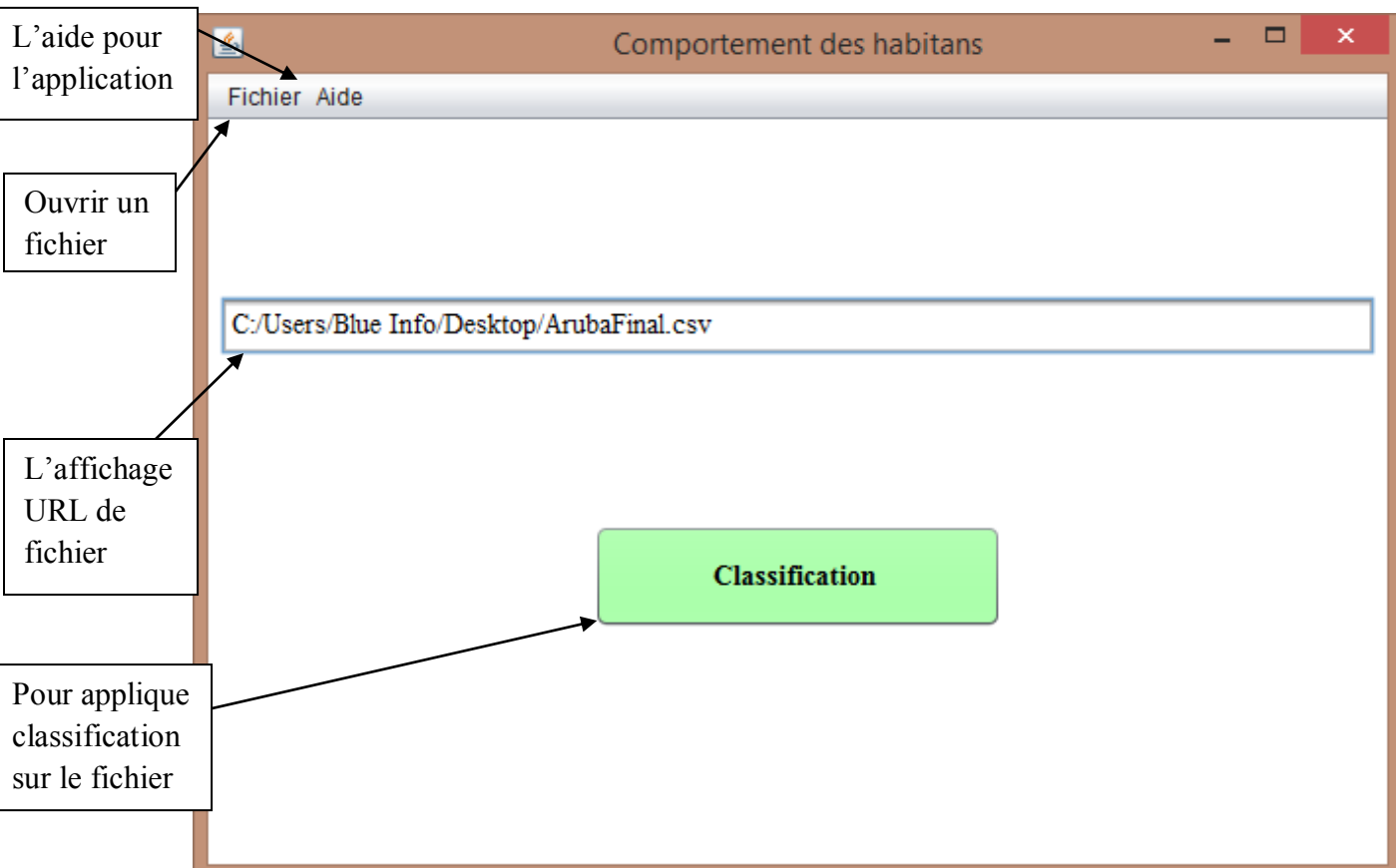


Figure 13 : Interface Accueil

- **Interface des résultats de Classification**

Dans cette interface, nous avons voir classification les activités par rapport aux temps.

Nous avons deux classifications :

Le premier c'est la classification pour un jour de la base de données et la deuxième c'est la classification pour la base de données Aruba.

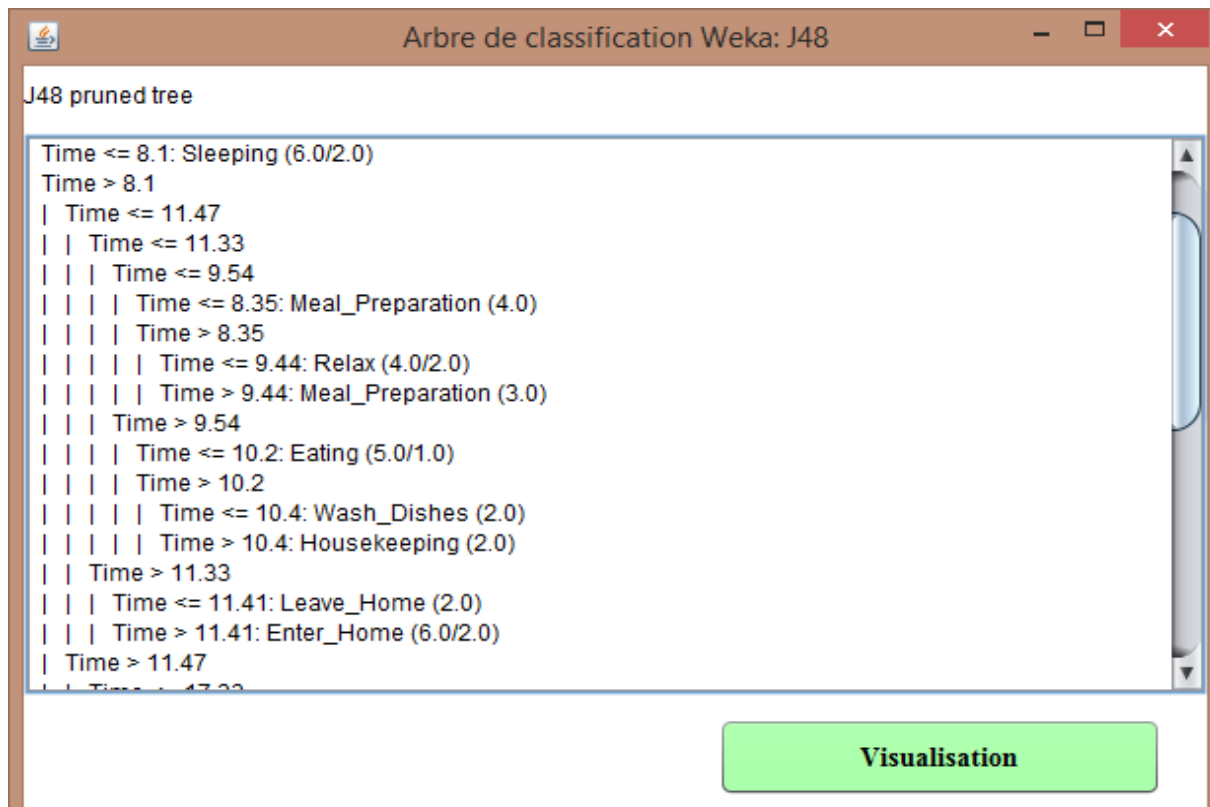


Figure 14 : Interface de classification des activités par rapport aux temps d'un 1 jour.

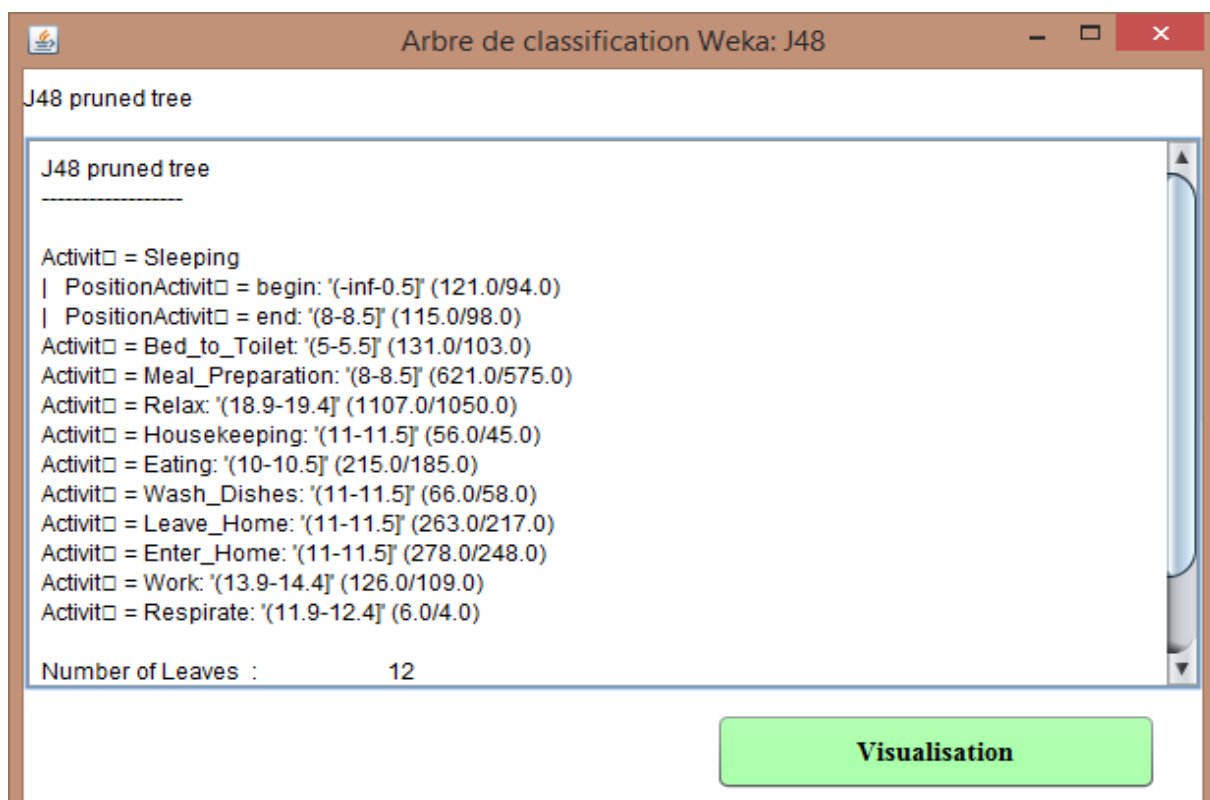


Figure 15 : Interface de classification des activités par rapport aux temps de la BDD

- **Interface de visualisation de l'arbre de décision**

Après faire la visualisation, nous avons voir un arbre de décision.

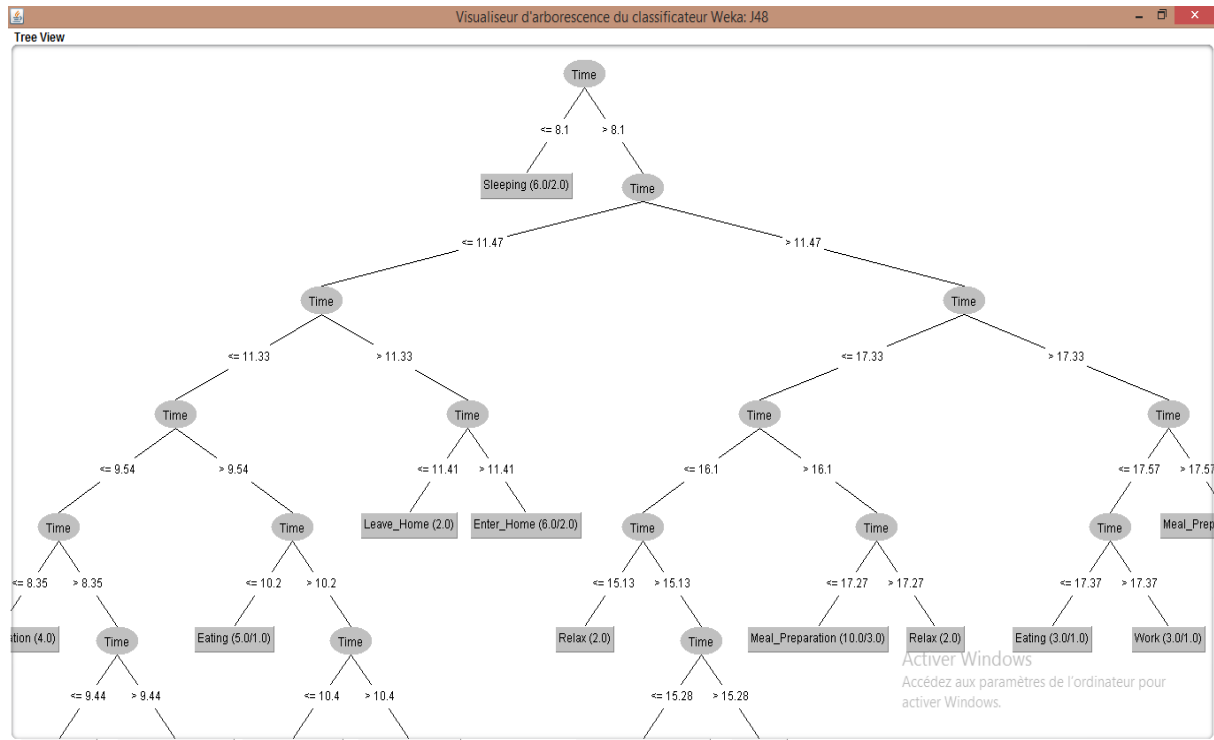


Figure 16 : Interface de visualisation l'arbre de décision pour 1 jour.

- Quelques explication :

Meal preparation : [8.35 ,9.45]

Relax :] 8.35 ,9.44]

Eating :] 9.54 ,10.2]

Wach dish :] 9.2 , 9.4]

... Etc.

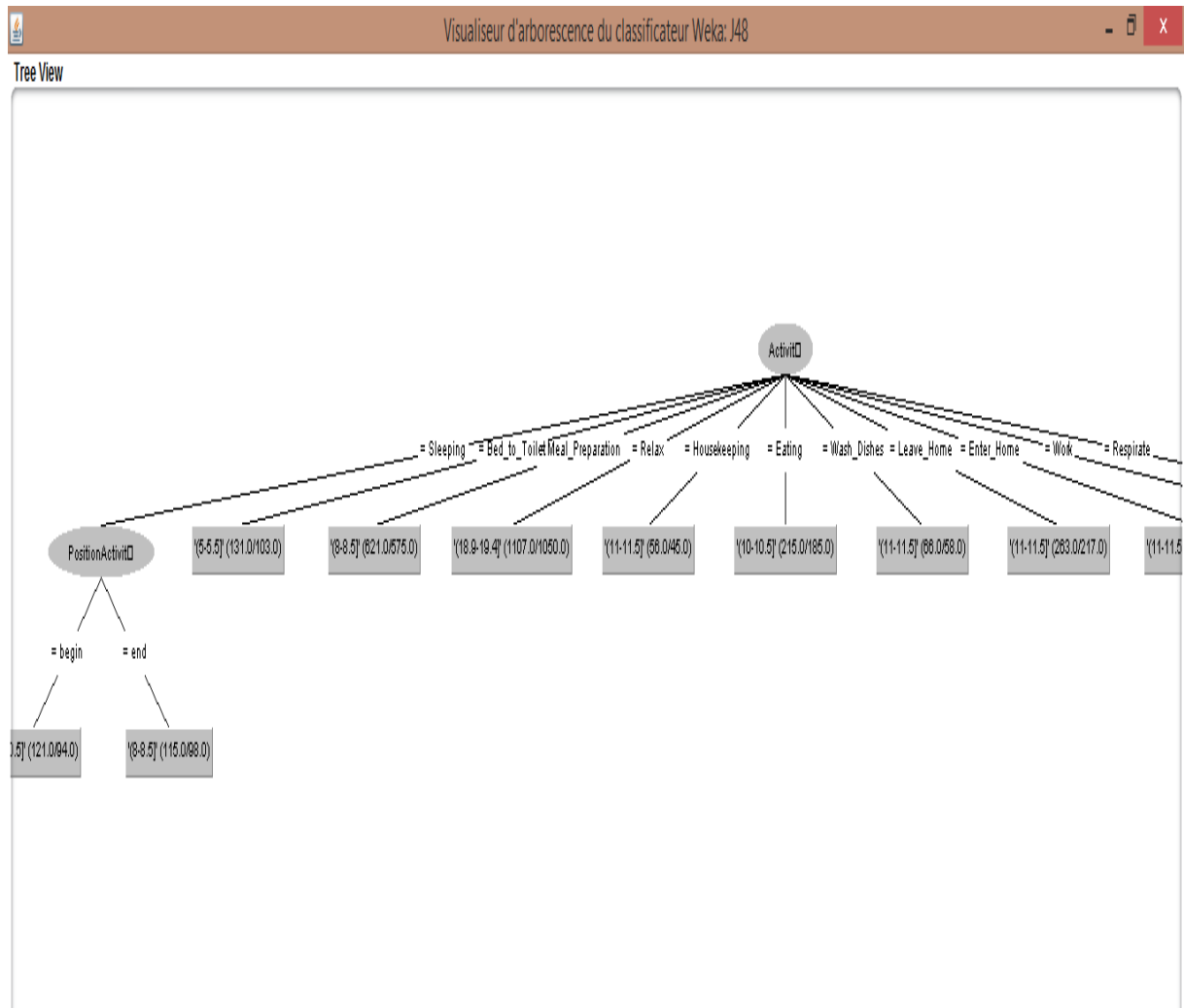


Figure 17 : Interface de visualisation l'arbre de décision par la BDD Aruba.

Conclusion

A travers ce chapitre nous avons abordé la mise en œuvre de notre solution. Nous avons commencé par voir les ressources utilisées, puis présenter le prétraitement de notre base de données, nous avons fini par une présentation des captures d'écran de l'application de Data Mining sur le dataset choisi.

Conclusion générale

Conclusion générale

Conclusion générale

Notre projet consiste à concevoir et réaliser une application Data Mining pour l'extraction des connaissances à partir un système ambiante (Prédiction le comportement des habitants).

Comme une appréciation personnelle, nous tenons à préciser qu'au cours de la réalisation de ce projet nous avons énormément appris à réaliser un outil concret. En résumé, nous pouvons dire que ça été une agréable expérience qui nous a aidé à découvrir notre potentiel et à nous préparer pour continuer nos chemins vers ce qui est de mieux.

Perspectives

Comme projection future de ce travail, il est utile de :

- Envisager une validation empirique de notre approche, à grande échelle, pour l'améliorer et apporter les correctifs nécessaires.
- nous proposons d'appliquer les autre techniques de data Mining.

Bibliographie

- [1] KAOUBI ADEL, La maison intelligente, MEMOIRE DE STAGE DE FIN D'ETUDES Pour l'obtention du «Mastère professionnel en Nouvelles Technologies des Télécommunications et Réseaux (N2TR)».Université Virtuelle de Tunis, 2018, page1
- [2] Plan « intelligence ambiante » : défis et opportunités Document de réflexion conjoint du comité d'experts « Informatique Ambiante » du département ST2I du CNRS et du Groupe de Travail « Intelligence Ambiante » du Groupe de Concertation Sectoriel (GCS3) du Ministère de l'Enseignement Supérieur et de la Recherche, DGRI A3 Version 1.2 finale - 14 octobre 2008
- [3] AHMED MALEK Nada. L'intelligence ambiante et les systèmes de transport intelligents. MEMOIRE En vue de l'obtention du diplôme MAGISTER en informatique, UNIVERSITE BADJI MOKHTAR ANNABA, 2014, page3
- [10] CHAMI Djazia , Une plate forme orientée agent pour le data mining, Mémoire En vue de l'obtention du diplôme de Magister en informatique. Université HADJ LAKHDAR – BATNA, 2009-2010. Page 4-32
- [11] Abdelouahab ATTIA, Extraction des connaissances par les règles d'association : Extension au cas flous, MEMOIRE Pour l'Obtention du Diplôme de MAGISTER. UNIVERSITE FARHAT AMMOUR Zahia, BENGUEDDA Nour el yakine kheira, L'extraction des connaissances Pour Le suivi épidémiologique, Mémoire Deuxième année Master Ingénierie des Systèmes d'Information. UNIVERSITE ABDELHAMID IBN BADIS MOSTAGANEM ,2012-2013.page 7-16
- [13] CHELIK Nassima, OULD ALI Nadia, Application de méthodes de data Mining dans le but d'analyser l'échec/succès des étudiants, Mémoire Pour l'obtention du diplôme de : MASTER Spécialité : Génie Logiciel. UNIVERSITE IBN KHALDOUN - TIARET ,2018-2019, page12
- [14] Hadjer BENYAMINA, Extraction de nouvelles connaissances d'une image vidéo indexée par le contenu: Implémentation d'un nouvel algorithme du Data Mining, MEMOIRE DE FIN D'ETUDES Pour l'Obtention du Diplôme de Master en Informatique. UNIVERSITE ABDELHAMID IBN BADIS - MOSTAGANEM, 2016-2017, page17-27
- [16] Guillaume CALAS. Études des principaux algorithmes de data mining. Kremlin-Bicêtre, France. Page2-4

Bibliographie

- [17] AMMOUR Zahia, BENGUEDDA Nour el yakine kheira, L'extraction des connaissances Pour Le suivi épidémiologique, Mémoire Deuxième année Master Ingénierie des Systèmes d'Information. UNIVERSITE ABDELHAMID IBN BADIS MOSTAGANEM ,2012-2013
- [18] BRAHIMI Belgacem, Extraction de connaissances à partir de données incomplètes et imprécise, Mémoire Présenté pour l'obtention du diplôme de MAGISTER. UNIVERSITE DE M'SILA, 2011.page 10
- [19] Kateb Nabila, Une approche multi agents pour Le Data Mining. MEMOIRE Pour l'obtention du diplôme de Master en informatique, Université Larbi Ben M'hidi Oum El Bouaghi.2010-2011

Webographie

Visité le 28/07/2020

- [4] <https://fr.slideshare.net/nikhilpatteri/ambient-intelligence-28697238>
- [5] <http://dictionnaire.sensagent.leparisien.fr/Informatique%20ubiquitaire/fr-fr/>
- [6] <https://fr.slideshare.net/ShifaliJindal/ambient-intelligence-made-by-shifali-jindal>
- [7] <https://fr.slideshare.net/ShifaliJindal/ambient-intelligence-seminar-report-made-by-shifali-jindal>
- [8] <http://flowmate.com/sensor/>
- [9] <https://fr.slideshare.net/chandrika95/ambient-intelligence-58604649>
- [12] Introduction au Data Mining et à l'apprentissage statistique. Gilbert Saporta.page 3.
Disponible sur <http://cedric.cnam.fr/~saporta/DM.pdf>
- [15] Data Mining«Fouille de données».page4. Disponible sur
<https://www.uv.es/nemiche/cursos/DataMining.pdf>
- [20] <https://medium.com/@nickmal/real-time-activity-recognition-in-a-smart-home-using-binary-sensors-efd147ec694>
- [21] <https://datakeen.co/8-machine-learning-algorithms-explained-in-human-language/>