

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université Abbes Laghrour-khenchela
Faculté Science et Technologie
Département d'Informatique



Mémoire de Fin d'études Master
Filière : Informatique
Spécialité : Sécurité Et Technologie Web

Impact De L'IoT Sur Le Domaine De L'agriculture

Encadré par:
Dr.Hioual Ouassila

Présenté par:
Chaima Ayeb
Ramzi Ben khelifa

Promotion 2019/2020

Remerciement

La réalisation de ce mémoire a été faite à l'aide des plusieurs personnes qu'on ne pourra jamais leurs exprimer notre gratitude.

Nous voudrions présenter nos remerciements à notre encadreur « Hioual.Ouassila ».Nous voudrions également lui témoigner notre gratitude pour sa patience et son soutien qui nous a été précieux afin de mener notre travail à bon port.

Merci

Dédicace

A mon père Salah Ayeb, pour m'avoir insufflé l'abnégation.

A ma chère mère Abboudi Fatiha, pour sa précieuse attention à mon égard.

A tous ceux qui croient aux vertus du sacrifice, de l'effort et du travail.

Chaima.A

Dédicace

Pour m'avoir permis d'accéder au savoir et d'être ce que je suis devenu aujourd'hui, je voudrais remercier **DIEU** le tout puissant et miséricordieux, qui m'a donné la force, la patience et la persévérance pour accomplir ce Modeste travail.

Je dédie ce modeste travail à celle qui m'a donné la vie, le symbole de tendresse, qui s'est sacrifiée pour mon bonheur et ma réussite, à **ma mère**.

A mon cher **père** pour l'éducation qu'il m'a prodigué; avec tous les moyens et au prix de tous les sacrifices qu'il a consenti à mon égard, pour m'inculper le sens du devoir depuis mon enfance.

A mes frères et sœurs : **Soriya, Ismahane, Saida, Dounia, Walid, Lotfi** et **Hichem**.

A toute la famille **Benkhelifa**.

A mon binôme, **Chaima** pour la patience dont il fait preuve envers moi.

A tous mes amis avec lesquels j'ai partagé des moments de joie et de bonheur, à tous ceux qui ont été à mes côtés jusqu'à aujourd'hui.

A tous les honorables enseignants qui ont contribué à ma formation.

Enfin, toute personne m'ayant aidé de près ou de loin dans ce travail laborieux et de longue haleine, trouve ici l'expression de mes vives reconnaissances et remerciement.

A tous ceux qui m'aiment **RAMZI**.

Ramzi.B

Résumé

L'agriculture est l'un des principaux facteurs qui changent la donne dans le principal secteur générateur de revenus pour les pays du monde entier. Récemment, l'Internet des objets a entraîné des changements majeurs dans l'agriculture, car il s'est accompagné de ce que l'on appelle une agriculture intelligente confrontée aux facteurs qui la contrôlent. Comme nous les savons, les différentes saisons, les modèles de marché et les modèles biologiques affectent la production agricole, mais en raison des changements dans ces modèles, les pertes pour les agriculteurs peuvent être réduites et ces facteurs peuvent être réduits. Ce travail se concentre, principalement, sur les algorithmes utilisés pour prédire le type de cultures rentable set adaptées. Pour cela, nous avons proposé une approche liée à la connaissance du type de culture approprié pour chaque ferme, en utilisant les données capturées par les capteurs sans fils pour extraire des décisions précises à travers l'arbre de décision. Le travail, proposé dans ce manuscrit, peut aider les agriculteurs à prédire le type de cultures les plus rentables.

Mots clés.

Internet des Objets, Agriculture Intelligente, Arbre de décision, Fouille de données, Capteurs sans fils, IoT.

Abstract

Agriculture is mainly one of the principal changing factors in the major income generating sector for countries around the world. Recently, the Internet of Things has brought about fundamental impacts in agriculture, as it has come with what is called smart agriculture confronted with the factors that control it. As we know, different seasons, market and organic patterns affect agricultural production, but due to changes in these patterns, losses for farmers can be reduced as well as these factors. This work substantially focuses on the algorithms used to predict the type of profitable and suitable crops. For this, we previously proposed an approach related to knowing the appropriate crop type for each farm, using the data captured by wireless sensors to extract precise decisions through the decision tree. The work, proposed in this manuscript, can help farmers predict the most profitable type of crops.

Keywords.

Internet of Things, Intelligent Agriculture, Decision tree, Data mining, Wireless sensors, IoT.

ملخص

الزراعة هي أحد العوامل الرئيسية التي تغير قواعد اللعبة في قطاع توليد الدخل الرئيسي للبلدان في جميع أنحاء العالم. أحدثت إنترنت الأشياء مؤخرًا تغييرات كبيرة في الزراعة حيث تزامنت مع ما يسمى بالزراعة الذكية في مواجهة العوامل التي تتحكم فيها. كما نعلم، فإن المواسم المختلفة وأنماط السوق والأنماط العضوية تؤثر على الإنتاج الزراعي، ولكن بسبب التغييرات في هذه الأنماط، يمكن تقليل الخسائر للمزارعين. يركز هذا العمل بشكل أساسي على الخوارزميات المستخدمة للتنبؤ بنوع المحاصيل المربحة والمناسبة. لهذا، اقترحنا سابقًا نهجًا يتعلق بمعرفة نوع المحاصيل المناسب لكل مزرعة، باستخدام البيانات التي تم التقاطها بواسطة أجهزة الاستشعار اللاسلكية لاستخراج قرارات دقيقة من خلال شجرة القرار. يمكن أن يساعد العمل، المقترح في هذه المذكرة، المزارعين على التنبؤ بأنواع المحاصيل الأكثر ربحية.

كلمات مفتاحية.

إنترنت الأشياء، الزراعة الذكية، شجرة القرار، التنقيب عن البيانات، أجهزة الاستشعار اللاسلكية، إنترنت الأشياء

Table des matières

Remerciement.....	II
Dédicace.....	III
Dédicace.....	IV
Résumé.....	V
Liste des figures.....	X
Introduction Générale	XI
Chapitre 01 : Introduction à l'Internet des Objets	13
1.1 Introduction	14
1.2 L'internet des objets (IoT)	14
1.2.1 Définitions de l'internet des objets (IoT)	15
1.2.2 Fonctionnement de l'internet des objets (IoT)	16
1.2.3 Architecture de l'internet des objets (IoT).....	18
1.2.4 Domaine d'application de l'internet des objets (IoT)	19
1.2.5 Les problèmes de l'internet des objets (IoT).....	21
1.3 Agriculture intelligente.....	22
1.3.1 Définition de l'agriculture intelligente	23
1.3.2 Domaines de l'agriculture intelligente	23
1.3.3 Les applications de l'agriculture intelligente	24
1.4 Conclusion.....	26
Chapitre 02 : Les Techniques de fouille de données (Data-Mining)	27
2.1 Introduction	28
2.2 La Fouille de données (Data-mining).....	28
2.2.1 Extraction de connaissance versus Fouille de données.....	28
2.2.2 Historique.....	29
2.2.3 Quelques définitions de Data-Mining.....	29
2.2.4 Processus d'extraction de connaissances à partir des données.....	30
2.2.5 Les tâches du Data-Mining	31
2.2.6 Les méthodes de Data-Mining	33
2.2.7 Les étapes du processus du Data-Mining.....	41
2.2.8 Catégorisation des systèmes du Data-Mining	42
2.2.9 Domaines d'application du Data-Mining.....	43
2.3 Conclusion	44
Chapitre 03 : Conception et Implémentation du Système Proposé.....	45

3.1 Partie I. Conception du système proposé.....	46
3.1.1 Problématique et objectifs.....	46
3.1.2 Travaux connexes.....	46
3.1.3 Modèle proposé.....	47
3.1.4 Aperçu de l'architecture proposée du système.....	47
3.1.5 Fonctionnement du système proposé.....	48
3.1.6 L'Arbre de décision proposé.....	50
3.2 Partie II. Implémentation du système proposé.....	52
3.2.1 Environnement de travail.....	52
3.2.2 Outils de développement.....	53
3.2.3 Description de l'interface.....	54
3.2.4 Difficultés techniques.....	58
3.3 Conclusion.....	59
Conclusion générale et perspectives.....	60
Bibliographie.....	61

Liste des figures

Figure 1: Estimation du nombre d'objets connectés en 2020 [5]	16
Figure 2: Fonctionnement de l'internet des objets (IoT)	17
Figure 3: Architecture de l'internet des objets (IoT)	19
Figure 4: Domaines d'application de l'internet des objets (IoT)	20
Figure 5: Les drones	25
Figure 6: Les agribots	25
Figure 7: Les tracteurs intelligents.....	26
Figure 8: Processus d'extraction des connaissances	30
Figure 9: Arbre de décision	39
Figure 10: Le processus du data-mining	42
Figure 11: Domaines d'application du data-mining	44
Figure 12: L'Architecture du système proposé	48
Figure 13: Diagramme de séquence de notre système	49
Figure 14: L'Arbre de décision proposé	50
Figure 15: Dataset utilisé pour la construction de l'arbre de décision.....	51
Figure 16: Caractéristique se notre PC portable	52
Figure 17: Caractéristiques du système d'exploitation.....	53
Figure 18: Formulaire d'Inscription	54
Figure 19: Interface de connexion	55
Figure 20: Profil d'utilisateur.....	55
Figure 21: Contenu de la barre inferieure	56
Figure 22: La fenêtre Home	56
Figure 23: Résultat d'exécution	57
Figure 24: Contenu de la fenêtre « Apple »	57
Figure 25: Résultat du « Apple »	58

Introduction Générale

L'agriculture est considérée comme la base de la vie humaine car elle est la principale source de nourriture et de revenus pour la plupart des pays du monde. L'économie du pays dépend de la production agricole car elle fournit, à ces citoyens, de la nourriture des matières premières, des emplois, etc. Il a été observé ces derniers temps qu'il n'y a pas de développement significatif dans la production végétale dans le secteur agricole. En outre, il y a une augmentation rapide du prix des aliments parce que la production de la récolte ne répond pas à la demande. L'une des causes de la diminution des récoltes de la production est l'utilisation par les agriculteurs de la méthode de culture traditionnelle, ce qui entraîne une baisse du rendement des cultures. Les agriculteurs, qui sont nouveaux dans le domaine de l'agriculture, ont des connaissances insuffisantes sur les caractéristiques du sol pour la culture végétale. Ils ne sont pas conscients du fait que les terres agricoles doivent être évaluées avant la culture. L'analyse de l'aptitude des terres est une condition préalable et obligatoire pour la culture, ce qui permet d'obtenir une production maximale. Afin d'acquérir les propriétés du sol, les agriculteurs dépendent des analyses de sol laboratoires, qui ne sont pas suffisants pour les aider, et parfois les données fournies par les laboratoires sont inexactes. Afin d'obtenir suffisamment de connaissances sur la culture, les données doivent être collectées manuellement, ce qui est très difficile pour les agriculteurs. La solution est le remplacement des méthodes traditionnelles de données collectées avec des capteurs basés sur l'Internet des objets (IoT). Les capteurs jouent un rôle important dans la collecte de l'information sur divers facteurs tels que le sol, l'eau, le climat, etc. A l'aide des données recueillies à partir de différents capteurs, une analyse de l'aptitude des terres pourrait être effectuée, ce qui aide les agriculteurs à identifier l'état actuel de leurs terres agricoles et à améliorer leur production agricole. Afin d'aider les agriculteurs à prendre une décision sur la culture, plusieurs outils de décision ont été développés jusqu'à présent, parmi ces outils on peut citer les arbres de décision.

L'objectif principal de ce mémoire est de proposer un système intelligent d'aide à la prise de décision qui permet à l'agriculteur de connaître les plantes appropriées pour sa ferme avant de cultiver la terre.

Afin de répondre à notre objectif ce manuscrit est subdivisé en trois chapitres et une conclusion. Le premier est introductif, où nous allons détailler les principes de base concernant les concepts liés à notre travail. Le deuxième chapitre présente les techniques de la

fouille de données. Le troisième chapitre présente la conception du système proposé. Ainsi que son implémentation. Nous terminerons par une conclusion dans laquelle nous discutons quelques perspectives.

Chapitre 01 : Introduction à l'Internet des Objets

1.1 Introduction

En raison de l'énorme croissance des technologies, l'agriculture est devenue plus populaire et plus importante. Ils existent plusieurs outils et techniques pour améliorer le développement de l'agriculture. Selon l'Organisation des Nations Unies pour l'alimentation et l'agriculture, afin de nourrir la population croissante de la Terre, le monde devra produire 70% plus de nourriture en 2050 qu'en 2006. [1] Pour répondre à cette demande, les agriculteurs et les entreprises agricoles se tournent vers l'Internet des objets pour des analyses et de plus grandes capacités de production. L'Internet des objets (IoT) peut jouer un grand rôle dans l'augmentation de la productivité, l'obtention d'un énorme marché mondial, une idée des tendances récentes des cultures. L'IoT est un réseau d'appareils interconnectés qui peuvent transférer efficacement des données sans intervention humaine.

Aujourd'hui, de nombreuses industries agricoles se sont tournées vers l'adoption de la technologie IoT pour une agriculture intelligente afin d'améliorer l'efficacité, la productivité, le marché mondial et d'autres fonctionnalités telles que l'intervention humaine minimale, le temps et les coûts, etc. L'avancement de la technologie garantit que les capteurs deviennent plus petits, sophistiqués et plus économiques. Les réseaux sont également facilement accessibles à l'échelle mondiale afin que l'agriculture intelligente puisse être réalisée avec un engagement total. Puisque l'innovation dans l'agriculture est vivement encouragée, l'agriculture intelligente est la réponse aux problèmes auxquels cette industrie est actuellement confrontée. Tout cela peut être fait à l'aide de téléphones intelligents et d'appareils IoT. Donc, l'agriculteur peut obtenir toutes les données ou informations requises et surveiller son secteur agricole

Dans ce chapitre, dans une première partie, nous présentons l'Internet des Objets (IdO) (Internet of Things (IoT) en Anglais), ses définitions de bases, ses techniques et domaines d'application. Et, la deuxième partie de ce chapitre, sera consacré à la présentation du domaine de l'agriculture intelligente.

1.2 L'internet des objets (IoT)

En 1989, Mark Weiser, professeur à Berkeley, avait une vision d'un monde où la technologie s'intègre dans les objets de la vie quotidienne. Aujourd'hui, sa vision prend forme : avec l'avancé de l'informatique et électronique, tout se miniaturise et se connecte à Internet : du téléphone, montres connectées, aux capteurs. On parle désormais de l'Internet des objets (IoT). Le terme IoT a été utilisé pour la première fois au laboratoire Auto-ID center au MIT

en 1999, où ils travaillaient sur l'identification de la fréquence radio (RFID) en réseau et sur les technologies de détection émergentes.

La quantité d'objets connectés se multiplie à grande vitesse. D'après Cisco, ils prévoient 50 milliards d'objets connectés dans le monde d'ici 2020. En France, de plus en plus de startups émergent dans le domaine de l'IoT.

Les usages de l'IoT se diversifient et touchent quasiment tous les domaines : la maison connectée, la santé et le bien-être, les voitures connectées, etc. Mais l'IoT ne se limite pas aux consommateurs finaux ; il s'étend aussi aux collectivités à l'image des villes intelligentes ou encore aux business tels que l'industrie et l'agriculture. [2]

1.2.1 Définitions de l'internet des objets (IoT)

L'Union Internationale des Télécommunications (UIT) définit l'Internet des Objets comme une extension de l'Internet tel que nous le connaissons aujourd'hui, par la création d'un réseau omniprésent et auto-organisé d'objets physiques connectés, identifiables et adressables permettant le développement d'applications au sein de secteurs verticaux clés et entre ces secteurs par le biais de puces intégrées. [3]

L'Internet des Objets est un réseau de réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant. [4]

Le potentiel d'objets qui pourraient être connectés d'ici 2020 est estimé entre 30 et 212 milliards selon de nombreuses études (Gartner, Berg Insight, Cisco ISBG, Idate, IDC) comme le présente la **Figure 1**. [4] Ces objets vont contribuer à la transformation digitale de très nombreux métiers, non seulement par leur usage mais aussi et surtout, parce que les données qu'ils produisent vont être à la source de changements profonds et de la naissance de services inédits. [5]

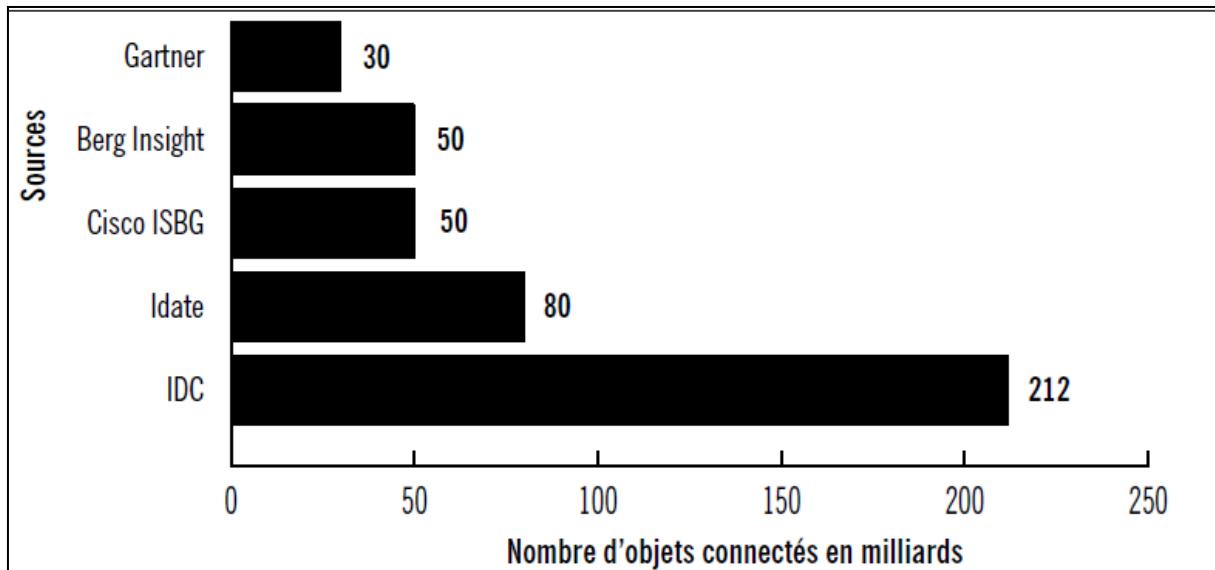


Figure 1: Estimation du nombre d'objets connectés en 2020 [5]

Il existe aussi plusieurs autres définitions possibles de l'IoT. En ce qui suit, nous citons, au moins, trois.

Définition 01 : L'Internet of Things (IoT) est « un réseau qui relie et combine les objets avec l'Internet, en suivant les protocoles qui assurent leurs communication et échange d'informations à travers une variété de dispositifs. » [6]

Définition 02 : L'IoT peut se définir aussi comme étant « un réseau de réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi, de pouvoir récupérer, stocker, transférer et traiter les données sans discontinuité entre les mondes physiques et virtuels. » [7]

Définition 03 : L'IoT est une extension de l'Internet actuel envers tout objet pouvant communiquer de manière directe ou indirecte avec des équipements électroniques eux-mêmes connectés à l'Internet. Cette nouvelle dimension de l'Internet s'accompagne avec de forts enjeux technologiques, économiques et sociaux, notamment avec les économies majeures qui pourraient être réalisées par l'ajout de technologies qui favorisent la standardisation de ce nouveau domaine, surtout en matière de communication, tout en assurant la protection des droits et des libertés individuelles. [8]

1.2.2 Fonctionnement de l'internet des objets (IoT)

Les objets connectés se multiplient et se diversifient tant sur le marché grand public que professionnel. Ceci a engendré un nouveau besoin : celui de créer des interactions entre ces objets, au-delà de leurs constructeurs ou secteurs d'activités. Automatiser certaines tâches de

la vie quotidienne ou professionnelle deviendra ainsi possible. Les plateformes IoT ont vocation à connecter ces objets hétérogènes et les faire communiquer entre eux.

Comme introduit au début, l'écosystème IoT est assez complexe, car il intègre plusieurs technologies et domaines de compétences. Un système IoT englobe, généralement, à la fois du hardware, des protocoles de communication, du software, du cloud et du mobile. Ainsi, un projet IoT nécessite d'avoir une équipe pluridisciplinaire.

On peut décomposer un système IoT en 4 fonctionnalités distinctes comme la montre la **Figure 2**. [9]

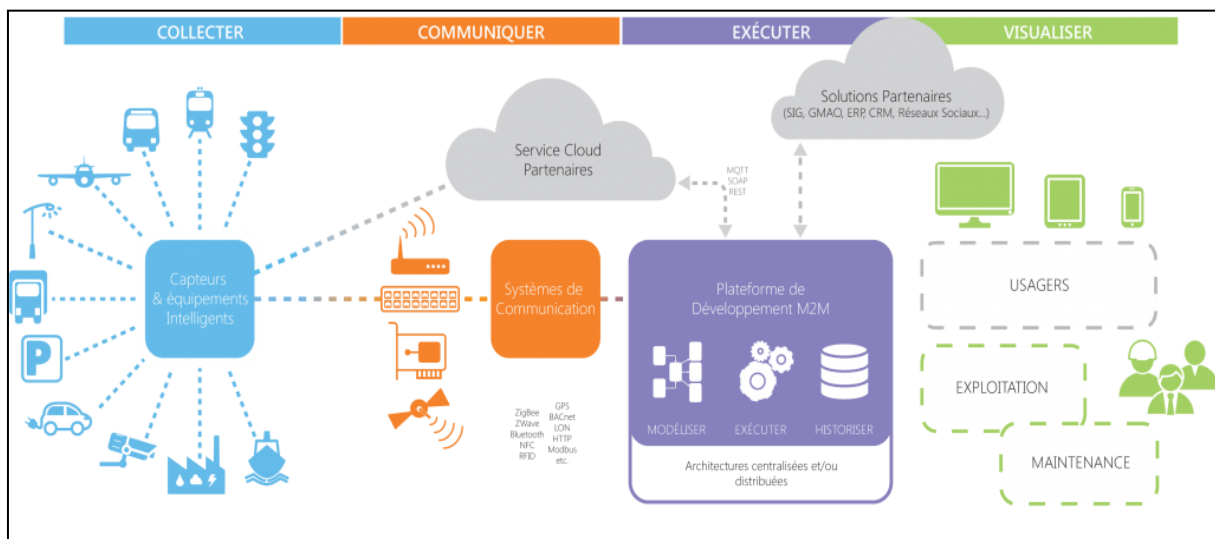


Figure 2: Fonctionnement de l'internet des objets (IoT)

- Collecter /Actionner:** A cette étape, on est au niveau de l'objet connecté. On parle de capteurs qui permettent de faire des mesures de l'environnement physique (ex : température, humidité, bruit) et des actionneurs qui peuvent agir sur l'environnement (ex : des moteurs pour fermer ou ouvrir une porte). Certains objets peuvent être dotés de capacités électroniques, informatiques et réseaux qui leur permettent de se connecter directement au réseau Internet. Mais généralement, ayant des contraintes matérielles et logicielles (autonomie limitée, capacité de traitement limitée, pas de stock réseau, etc), les objets implémentent des protocoles de communication a basse énergie / bas débit et communiquent avec le réseau internet à travers une passerelle (Gateway) Un exemple de cette gateway est notre téléphone portable qui permet de collecter les données de notre montre intelligente. [10]
- Communiquer:** C'est l'étape qui permet l'envoi des données depuis le réseau local vers le cloud. On parle essentiellement des protocoles pour transporter la donnée et on peut en distinguer deux modèles : Le modèle Publish / Subscribe avec des protocoles

de type MQTT et le modèle REST avec des protocoles comme HTTP ou encore CoAP. [10]

- c. **Exécuter** : C'est l'étape de stockage et de traitement de la donnée. A cette étape on parle souvent de " Plate-forme IoT " qui est souvent une solution cloud capable de connecter plusieurs objets connectés, stocker leurs données, les traiter, les analyser et les exposer à travers différentes applications. Les plateformes IoT permettent aussi de faire communiquer d'objets hétérogènes. Ces plateformes se multiplient de nos jours (Amazon, Google, Microsoft, etc.) et on parle même de " guerre des plateformes IoT". [10]
- d. **Visualiser** : C'est l'étape qui permet d'exposer les services des objets connectés à travers différentes applications dédiées. Un utilisateur, à travers une application mobile, peut par exemple communiquer avec ses objets en consultant leurs données ou en envoyant des actions vers ses objets. [10]

1.2.3 Architecture de l'internet des objets (IoT)

Les objets de l'environnement de l'internet des objets permettent de collecter, stocker et transmettre des données issues du monde physique. Ce sont des sources de données, qui possède au minimum un identifiant unique attaché à une identité ayant un lien direct ou indirect avec Internet.

On distingue deux types d'objet :

- a. **Les objets passifs** : ils utilisent généralement un tag (puce RFID, code barre 2D) Ils ont une capacité de stockage faible (de l'ordre du kilooctet) et permettent de jouer le rôle d'identification. Ils peuvent parfois, dans le cas d'une puce RFID, embarquer un capteur (température, humidité) et être réinscriptibles.
- b. **Les objets actifs** : ils peuvent être équipés de plusieurs capteurs, avec une grande capacité de stockage, capables d'accomplir des calculs et être en mesure de communiquer sur un réseau.

On peut préciser le rôle des différents processus présents sur la **Figure 3** :

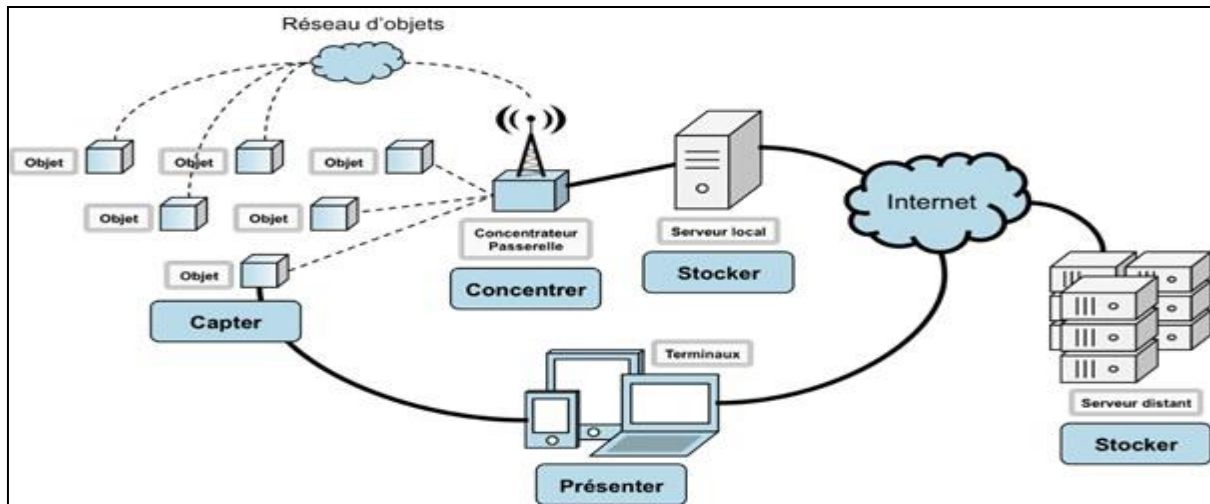


Figure 3: Architecture de l'internet des objets (IoT)

- a. **Capter** : permet de transformer une grandeur physique analogique en un signal numérique.
- b. **Concentrer** permet d'interfacer un réseau spécialisé d'objet à un réseau IP standard.
- c. **Stocker** permet de rassembler des données brutes, produites en temps réel, arrivant de façon non prévue.
- d. **Préserver** permet de collecter les informations de façon compréhensible par l'Homme, en lui offrant un moyen d'agir et/ou d'interagir.

Deux autres processus qui n'apparaissent pas sur le schéma :

- a. **Le traitement des données** : est un processus qui peut intervenir à tous les niveaux de la chaîne, depuis la capture de l'information jusqu'à sa restitution. Une stratégie pertinente, et commune quand on parle d'Internet des objets, consiste à stocker l'information. Cette stratégie est possible aujourd'hui grâce à des architectures distribuées, capables d'emmagasiner de grandes quantités d'information tout en offrant la possibilité de réaliser des traitements complexes en leur sein.
- b. **La transmission des données** : est un processus qui intervient à tous les niveaux de la chaîne. Deux réseaux supportent des transmissions le réseau local de concentration (utilise ANT, ZigBee et Zwave..) et le réseau WAN (Wifi, réseaux cellulaires . . .). [9]

1.2.4 Domaine d'application de l'internet des objets (IoT)

Le marché des objets connectés a promis à une grande croissance dans les années à venir car il a une valeur immense dans les différents domaines d'objets connectés pour les professionnels. Cependant, seules quelques applications sont actuellement déployées.

L'utilisation de l'IoT permettra le développement de plusieurs applications intelligentes qui toucheront essentiellement ceux qu'on citera dans ce qui suit. [10] (Voir la **Figure 4** ci-dessous)

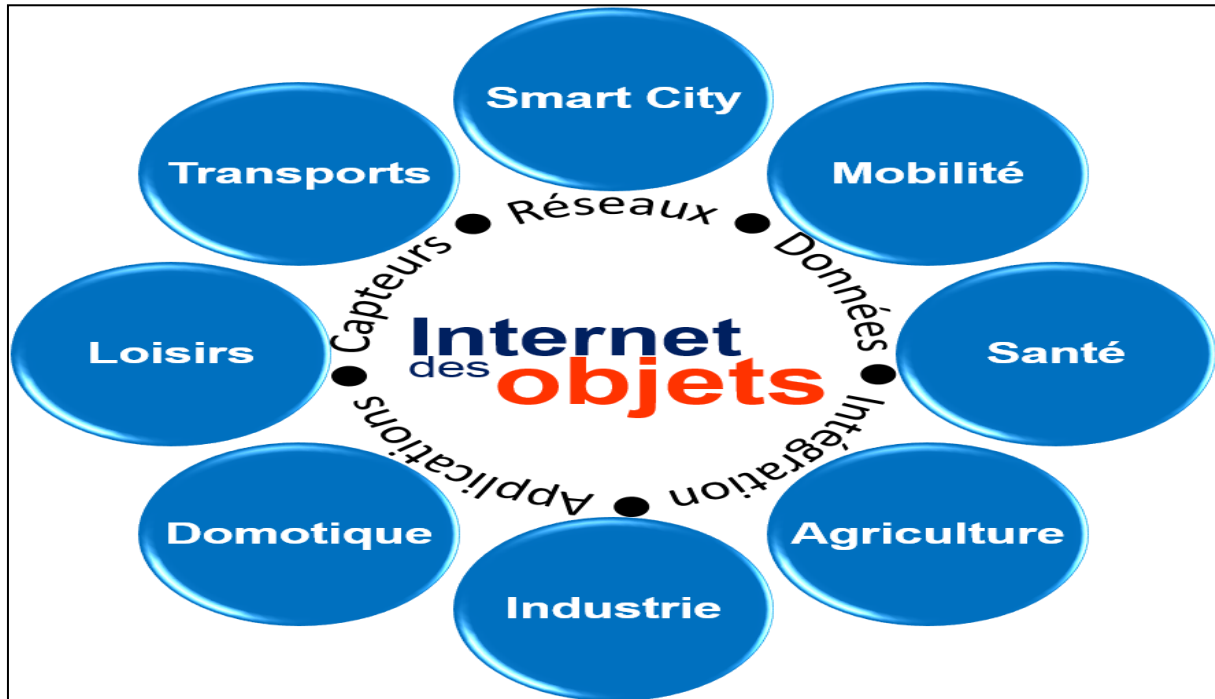


Figure 4: Domaines d'application de l'internet des objets (IoT)

Donc, selon la Figure 4, ils existent plusieurs domaines:

- a. **La santé :** Les objets connectés peuvent servir à réduire quelques éléments de dépenses pour les remplacer par d'autres. Ils permettent aussi de favoriser l'hospitalisation à domicile, qui assurera le contrôle et le suivi des signes cliniques des patients par la mise en place des réseaux personnels de surveillance, ces réseaux seront constitués de biocapteurs posés sur le corps des patients ou dans leurs lieux d'hospitalisation. Cela facilitera la télésurveillance des patients qui permettras de réduire les erreurs médicales, optimiser la consommation de médicaments ou encore leur prise régulière, et même encourager la prévention de certaines maladies, l'internet des objets permettre aussi de suivre la tension d'un patient, son rythme cardiaque, la qualité de sa respiration ou encore sa masse, etc. [11]
- b. **Le secteur sportif :** Dans le domaine sportif, de nombreux objets connectés comme des montres ou des bracelets connectés permettent de calculer, pendant la journée, le nombre de pas effectuée, la distance par courue, le temps d'activités d'un sportif, les calories brulées, ainsi pendant la nuit en calculant les heures de sommeil [12]

- c. Le secteur de la domotique :** Les objets connectés permettent de rendre une maison connectée, d'où le nom très utilisé de smart home. Dans une maison intelligente, on peut avoir de nombreux dispositifs de sécurité, comme par exemple : un détecteur de fumée, ou encore une serrure connectée. Dans le domaine de l'électroménager, on parle de l'électroménager connecté comme par exemple : un réfrigérateur connecté. Pour la décoration de la maison, on peut citer comme exemples : une lampe connectée, un cadre lumineux, une plante connectée, etc. Il existe aussi des villes intelligentes (Smart Cities). En effet, grâce à des services avancés, il est en effet possible d'optimiser l'utilisation des infrastructures physiques de la ville (par exemple, les réseaux routiers, le réseau électrique, etc.) et la qualité de vie des citoyens. Donc, on peut dire que l'internet des objets change les modalités d'accès au réseau et produit de nouvelles interactions homme-machine. [13]
- d. L'automobile :** Le marché des transports a déjà anticipé l'arrivée des objets connectés. Parmi les enjeux les plus fréquents que ce domaine fait naître, on peut énumérer la réduction des accidents et des embouteillages, le partage de voitures, le développement des offres de VTC et de TAX ou encore la gestion des flots automobiles. [13]
- e. L'industrie :** Le déploiement de L'IoT dans l'industrie sera certainement un grand support pour le développement de l'économie et le secteur des services, puisque L'IoT permettra d'assurer un suivi total des produits, de la chaîne de production, jusqu'à la chaîne logistique et de distribution en supervisant les conditions d'approvisionnements. Cette traçabilité de bout en bout facilitera la lutte contre la contrefaçon, la fraude et les crimes économiques transfrontaliers.
- f. L'agriculture :** L'IoT présentera des outils de choix pour la supervision de l'environnement des cultures, ce qui permettra une meilleure aide à la décision en agriculture. L'IoT servira non seulement à optimiser l'eau d'irrigation, l'usage des intrants et la planification des travaux agricoles, mais aussi, cette technologie peut être utilisée pour lutter contre la pollution (l'air, le sol et les eaux) et améliorer la qualité de l'environnement en général.

1.2.5 Les problèmes de l'internet des objets (IoT)

Comme toutes les autres technologies, l'Internet des objets a des difficultés d'application et de réalisation. Les problèmes majeurs sont :

- a. Personne ne portera 50 appareils. Les appareils sont de plus en plus spécifiques ils sont conçus pour faire une chose et la faire vraiment bien, que ce soit mesurer le sommeil, sa vitesse d'alimentation. Au nom de la simplicité, les entreprises cherchent trop à résoudre un seul problème, un seul cas d'utilisation.
- b. Plus d'inférence moins de détection. Avons-nous vraiment besoin d'un bracelet qui détecte le niveau d'UV auquel nous sommes exposés ? L'inférence, qui prend en charge le contexte est plus nécessaire que la détection permanente. Dans l'Internet des objets, les objets qui sont capables de comprendre l'information par inférence plutôt que par une détection permanente ont plus de chance de survivre que les autres.
- c. Le problème du découpage des données. Dans le monde de l'Internet des objets, trop souvent, les interfaces et la façon dont sont produites les données découragent l'échange. Le manque d'interopérabilité et la guerre des formats est la règle. Nombreux sont ceux qui veulent créer le système d'exploitation unique de l'IoT.
- d. Le contexte est la clef. Si les données ne changent pas notre comportement, alors il ne sert à rien de les recueillir. V1bes ou les solutions sociométriques mesurent le niveau de stress, mais s'ils ne captent pas le contexte qui génère ce stress, alors à quoi le rattacher, comment l'interpréter, le comprendre ?

L'Internet des Objets, ou M2M à l'échelle industrielle, est en pleine expansion depuis quelques années : nous passerons de 15 milliards d'objets connectés dans le monde en 2012 à 80 milliards d'ici 2020 selon les prévisions.

Entrez dans la révolution IOT du 21ème siècle basé sur des objets connectés, source de données tout au long de leur cycle de vie, qui se focalise sur la connectivité permanente et l'exploitation systématique de ces données grâce au Big Data. Ces données, relatives à l'état et l'environnement d'exploitation des objets en question, élargissent les possibilités de services du fabricant dans de nombreux domaines tels que l'observance, l'optimisation des flux, la maintenance prédictive et l'automatisation. [16]

1.3 Agriculture intelligente

À l'ère de la technologie, l'agriculture intelligente est devenue un vaste domaine à plusieurs dimensions, dont nous parlerons plus en détail ci-dessous.

1.3.1 Définition de l'agriculture intelligente

L'agriculture intelligente est une révolution de l'agriculture classique qui implique la réorientation des systèmes agricoles afin de soutenir efficacement le développement alimentaire. Le principal objectif de l'agriculture intelligente est d'accroître la productivité et les revenus agricoles. L'agriculture intelligente implique l'utilisation des technologies de la communication de l'information (TIC) et en particulier de l'Internet des objets (IoT) et de l'analyse de données volumineuses (Big data) associées pour faire face à ces défis via la surveillance électronique des cultures, ainsi que pour l'environnement, le sol, la fertilisation et conditions d'irrigation. Ces données de surveillance peuvent ensuite être analysées pour identifier les cultures qui répondent le mieux aux objectifs de productivité de toute exploitation agricole dans le monde. [17]

1.3.2 Domaines de l'agriculture intelligente

En raison de l'expansion et du développement de l'agriculture intelligente, cette dernière a fait l'objet de plusieurs domaines, dont les suivants :

- a. **Agriculture de précision** : L'agriculture de précision est un principe de gestion des parcelles agricoles qui vise l'optimisation des rendements et des investissements, en cherchant à mieux tenir compte des variabilités des milieux et des conditions entre parcelles différentes ainsi qu'à des échelles intra-parcellaires. [18] Elle s'intéresse à l'intégration et l'utilisation des technologies de l'information et de divers éléments tels que capteurs, systèmes de contrôle, robots et véhicules autonomes pour faciliter la gestion de l'élevage et de la croissance des cultures. L'agriculture de précision est l'une des applications Internet les plus populaires dans le secteur agricole. De nombreuses organisations bénéficient de cette technologie dans le monde entier. Elle peut impliquer l'utilisation de divers types de capteurs qui remontent des informations riches de manière régulière, notamment des capteurs de sol qui recueillent des données sur la teneur des sols en azote, capteurs d'irrigation mesurent le niveau d'eau, capteurs d'inondation surveillent également les niveaux d'eau et capteurs de gel. Les informations fournies par ces capteurs permettront de contrôler automatiquement le système d'irrigation et d'arrosage et d'alerter automatiquement les utilisateurs lorsque les conditions météorologiques ou autres événements susceptible d'endommager les plants sensibles se produisent.

- b. Contrôle du bétail:** L'internet des objets peut aider également à contrôler les bétails. Les grandes exploitations peuvent tirer parti des applications Internet sans fil pour collecter des données sur l'emplacement, le bien-être et la santé de leurs animaux, ce qui les aide à identifier les animaux malades et à les séparer du troupeau, évitant ainsi la propagation de la maladie. [19]
- c. Serres intelligentes :** La serre intelligente est une serre qui intègre la technologie de l'Internet des objets en vue d'améliorer la productivité des légumes, des fruits et des cultures. La technologie de l'Internet des objets est ainsi utilisée pour collecter et analyser en temps réel les indicateurs bioclimatiques de la serre et par conséquent appliquer (d'une manière automatique, semi automatique ou manuelle) les mesures et actions requises pour maintenir les conditions environnementales nécessaires aux plantes. Pour contrôler la serre, différents capteurs (connectés à internet ou non) sont utilisés. Pour mesurer les normes environnementales en fonction des exigences de chaque plante. Ceci élimine la nécessité d'une surveillance statique dans les serres. Ces capteurs fournissent des informations sur les niveaux de lumière, de pression, d'humidité et de température, qui contrôlent automatiquement les déclencheurs pour ouvrir la fenêtre, allumer les lumières, contrôler le chauffage et allumer le ventilateur.

1.3.3 Les applications de l'agriculture intelligente

La transformation numérique du secteur agricole rassemble l'utilisation de plusieurs technologies qui s'articulent autour de trois couches.

- a. Les Drones :** Les drones sont utilisés pour collecter des données. Ils observent l'état des cultures afin de cartographier les variations en matière de, notamment, mauvaises herbes, rendements et état des sols, (voir la **Figure 5** suivante).



Figure 5: Les drones

- b. Les Agribots :** Les agribots sont des petits robots conçus pour accomplir certaines tâches dans le domaine de l'agriculture. Leurs utilisations sont nombreuses : injection d'engrais, arrosage des cultures, désherbage au laser, pulvérisation, traite des vaches, nettoyage d'une étable, récolte de légumes, (voir la **Figure 6**).



Figure 6: Les agribots

- c. Les tracteurs intelligents :** Les tracteurs intelligents sont utilisés pour maximiser les rendements sur des tâches de grande échelle. Des tracteurs connectés peuvent, par exemple, intégrer directement des informations cartographiques pour un apport

d'engrais sur mesure automatique. Leur équipement GPS permet d'optimiser les chemins au travers des cultures, (voir la **Figure 7**).



Figure 7: Les tracteurs intelligents

1.4 Conclusion

L'agriculture jouera un rôle vital au cours des prochaines années dans le pays. Il faut donc une agriculture intelligente. L'Internet des objets contribuera à améliorer l'agriculture intelligente. L'IoT joue un rôle important dans différents domaines de l'agriculture, il permet d'améliorer l'efficacité du temps, la gestion de l'eau, la surveillance des cultures, la gestion des sols, le contrôle des insecticides et des pesticides, etc. Il minimise également les efforts humains, simplifie les techniques agricoles et aide à gagner une agriculture intelligente. Parallèlement à ces fonctionnalités, l'agriculture intelligente peut aider à développer le marché des agriculteurs avec une seule touche et un minimum d'efforts.

Chapitre 02 : Les Techniques de fouille de données (Data- Mining)

2.1 Introduction

Le développement des technologies de l'information a généré une grande quantité de bases de données et d'énormes données dans divers domaines. La recherche dans les bases de données et les technologies de l'information a donné lieu à une approche pour stocker et manipuler ces précieuses données pour une prise de décision ultérieure. L'exploration de données est un processus d'extraction d'informations et de modèles utiles à partir d'énormes données. Il est également appelé processus de découverte de connaissances, exploration de connaissances à partir de données, extraction de connaissances ou analyse de données /modèles, traite généralement de données qui ont déjà été collectées dans un but plutôt que l'analyse de l'exploration de données. Cela signifie que les objectifs de l'exercice d'exploration de données ne jouent aucun rôle dans la stratégie de collecte de données. Les ensembles de données examinés dans l'exploration de données sont souvent volumineux. Dans le reste de ce chapitre, nous allons commencer par introduire les principes de base du data-mining. Ensuite, nous expliquons les différentes tâches et les méthodes de la fouille de données. Puis, nous présentons quelques domaines d'application et, nous terminerons par une petite conclusion.

2.2 La Fouille de données (Data-mining)

La fouille de données (data mining en anglais), est le cœur du processus d'ECD (Extraction de Connaissances dans les bases de Données). Il s'agit, à ce niveau, de trouver des pépites de connaissances à partir des données. Tout le travail consiste à appliquer des méthodes intelligentes dans le but d'extraire cette connaissance. Il est possible de définir la qualité d'un modèle en fonction de critères comme les performances obtenus, la fiabilité, la compréhensibilité, la rapidité de construction et d'utilisation et enfin l'évolutivité. Tout le problème de la fouille de données réside dans le choix de la méthode adéquate à un problème donné. Il est possible de combiner plusieurs méthodes pour essayer d'obtenir une solution optimale globale. [20]

2.2.1 Extraction de connaissance versus Fouille de données

Les concepts de fouille de données et d'extraction de connaissances à partir de données sont parfois confondus et considérés comme synonymes. Mais, formellement on considère la fouille de données comme une étape centrale du processus d'Extraction de Connaissances des Bases de Données (ECBD ou KDD pour Knowledge Discovery in Databases en anglais). [21]

2.2.2 Historique

‘‘Data Mining’’ que l’on peut traduire par “la fouille de données” apparaît au milieu des années 1990 aux États-Unis comme une nouvelle discipline à l’interface de la statistique et des technologies de l’information: bases de données, intelligence artificielle, apprentissage automatique « *machine learning* ». [21]

Les premières applications se sont faites dans le domaine de la gestion de la relation client qui consiste à analyser le comportement de la clientèle pour mieux la fidéliser et lui proposer des produits adaptés. La recherche d’information dans les grandes bases de données médicales ou de santé (enquêtes, données hospitalières etc.) par des techniques de *Data Mining* est encore relativement peu développée, mais devrait se développer très vite à partir du moment où les outils existent.

La communauté de “*data mining*” a initié sa première conférence en 1995 à la suite de nombreux ateliers (workshops) sur le *KDD* entre 1989 et 1994. La première revue du domaine “*Data mining and knowledge discovery journal*” publiée par “*Kluwers*” a été lancée en 1997. [22]

2.2.3 Quelques définitions de Data-Mining

Selon le Groupe Gartner, le data-mining appelé aussi « fouille de données » est le processus de découverte de nouvelles corrélations, modèles et tendances en analysant une grande quantité de données. Ceci, est en utilisant les technologies de reconnaissance des formes ainsi que d’autres techniques statistiques et mathématiques. [23]

Ils existent d’autres définitions. Parmi lesquelles on peut citer :

Définition 1 : « Le data-mining, ou fouille de données, est l’ensemble des méthodes et techniques destinées à l’exploration et l’analyse de bases de données informatiques (souvent grandes), de façon automatique ou semi-automatique, en vue de détecter dans ces données des règles, des associations, des tendances inconnues ou cachées, des structures particulières restituant l’essentiel de l’information utile tout en réduisant la quantité de données ». [24]

Définition 2 : « Le data-mining est un domaine interdisciplinaire utilisant dans le même temps des techniques d’apprentissage automatiques, de reconnaissance des formes, des statistiques, des bases de données et de visualisation pour déterminer les manières d’extraction des informations de très grandes bases de données ». [25]

Définition 3 : «Le data-mining est un processus non trivial qui consiste à identifier, dans des données, des schémas nouveaux, valides, potentiellement utiles et surtout compréhensibles et utilisables ». [26]

2.2.4 Processus d'extraction de connaissances à partir des données

Le processus d'extraction de l'information consiste à parcourir les données volumineuses contenues dans une base, à la recherche de connaissance. Ce processus est décrit dans le schéma suivant la (figure). Ce processus comprend des étapes de définition du problème (définition du domaine, but de l'utilisateur final), de préparation des données (sélection, préparation, transformation), de fouille de données (sélection, des outils de data mining appropriés, recherche des patrons) et d'évaluation des résultats pour aboutir aux nouvelles connaissances. Le processus présenté est itératif et plusieurs retours en arrière dans les différentes étapes peuvent être nécessaires pour affiner les résultats,[27] (voir la **Figure 8**).

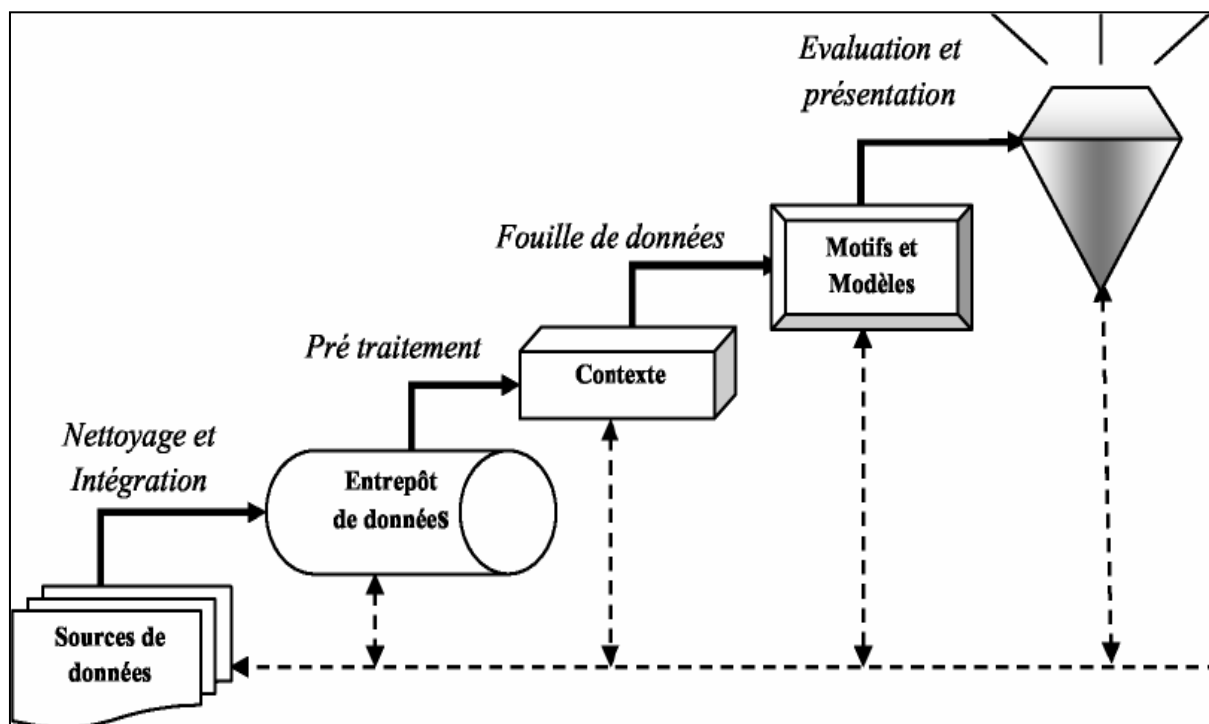


Figure 8: Processus d'extraction des connaissances

- a. **Nettoyage et intégration des données :** Le nettoyage des données consiste à retravailler ces données bruitées, soit en les supprimant, soit en les modifiant de manière à tirer le meilleur profit. L'intégration est la combinaison des données provenant de plusieurs sources (base de données, sources externes, etc.). Le but de ces deux opérations est de générer des entrepôts de données et/ou des magasins de données

spécialisés contenant les données retravaillées pour faciliter leurs exploitations futures. [28]

- b. Prétraitement des données :** Il peut arriver parfois que les bases de données contiennent des données incomplètes et/ou bruitées. Ces données erronées, manquantes ou inconsistantes doivent être retravaillées si cela n'a pas été fait précédemment. Dans le cas contraire, durant l'étape précédente, les données sont stockées dans un entrepôt. Cette étape permet de sélectionner et transformer des données de manière à les rendre exploitables par un outil de fouille de données. [28]

2.2.5 Les tâches du Data-Mining

Beaucoup de problèmes intellectuels, économiques ou même commerciaux peuvent être exprimés en termes des six tâches suivantes :

- La classification
- L'estimation
- La prédiction
- Le groupement par similitude
- L'analyse des clusters
- La description.

Les trois premières tâches sont des exemples du data-mining supervisé dont le but est d'utiliser les données disponibles pour créer un modèle décrivant une variable particulière prise comme but en termes de ces données. Le groupement par similitude et l'analyse des clusters sont des tâches non-supervisées où le but est d'établir un certain rapport entre toutes les variables. [30] La description appartient à ces deux catégories de tâche, elle est vue comme une tâche supervisée et non-supervisée en même temps. [29]

- a. La classification :** La classification est la tâche la plus commune du Data Mining et qui semble être une obligation humaine. Afin de comprendre notre vie quotidienne, nous sommes constamment classifiés, catégorisés et évalués. [29] Quelques exemples de l'utilisation des tâches de classification dans les domaines de recherche et commerce sont les suivants :

- Déterminer si l'utilisation d'une carte de crédit est frauduleuse.
- Diagnostiquer si une certaine maladie est présente.
- Déterminer quels numéros de téléphone correspondent aux fax.
- Déterminer quelles lignes téléphoniques sont utilisées pour l'accès à Internet.

b. L'estimation: L'estimation est similaire à la classification à part que la variable de sortie est numérique plutôt que catégorique. En fonction des autres champs de l'enregistrement, l'estimation consiste à compléter une valeur manquante dans un champ particulier. Par exemple, on cherche à estimer la lecture de tension systolique d'un patient dans un hôpital, en se basant sur l'âge du patient, son genre, son indice de masse corporelle et le niveau de sodium dans son sang. La relation entre la tension systolique et les autres données vont fournir un modèle d'estimation. Et, par la suite nous pouvons appliquer ce modèle dans d'autres cas. [29] Quelques exemples de l'utilisation des tâches d'estimation dans les domaines de recherche et commerce sont les suivants :

- Estimer le nombre d'enfants dans une famille.
- Estimant le montant d'argent qu'une famille de quatre membres choisis aléatoirement dépensera pour la rentrée scolaire.
- Estimer la valeur d'une pièce immobilière.

Souvent, la classification et l'estimation sont utilisés ensemble, comme quand le data-mining est utilisée pour prévoir qui va probablement répondre à une offre de transfert d'équilibre (de solde) de carte de crédit et aussi évaluer la taille de l'équilibre (du solde) à être transféré.

c. La prédiction : La prédiction est la même que la classification et l'estimation, à part que dans la prédiction les enregistrements sont classés suivant des critères (ou des valeurs) prédites (estimées). La principale raison qui différencie la prédiction de la classification et l'estimation est que dans la création du modèle prédictif on prend en charge la relation temporelle entre les variables d'entrée et les variables de sortie. [29] Quelques exemples de l'utilisation des tâches de prédiction dans les domaines de recherche et commerce sont les suivants :

- Prévoir le prix des actions dans les trois prochains mois.
- Prévoir le champion de la coupe du monde en football en se basant sur la comparaison des statistiques des équipes.

d. Le groupement par similitude : Le groupement par similitude consiste à déterminer quels attributs "vont ensemble". La tâche la plus répandue dans le monde du business, où elle est appelée l'analyse d'affinité ou l'analyse du panier du marché, est l'association des recherches pour mesurer la relation entre deux et plusieurs attributs. Les règles d'associations sont de la forme "Si antécédent, alors conséquent". Quelques

exemples de l'utilisation des tâches du groupement par similitude dans les domaines de recherche et commerce sont les suivants :

- Trouver dans un supermarché quels produits sont achetés ensemble et quels sont ceux qui ne s'achètent jamais ensemble.
- Déterminer la proportion des cas dans lesquels un nouveau médicament peut générer des effets dangereux. [31]

e. L'analyse des clusters : Le clustering (ou la segmentation) est le regroupement d'enregistrements ou des observations en classes d'objets similaires; un cluster est une collection d'enregistrements similaires l'un à l'autre, et différents à ceux existants sur les autres clusters. La différence entre le clustering et la classification est que dans le clustering il n'y a pas de variables sortantes. La tâche de clustering ne classe pas, n'estime pas, ne prévoit pas la valeur d'une variable sortantes. Au lieu de cela, les algorithmes de clustering visent à segmenter la totalité de données en de sous-groupes relativement homogènes. Ils maximisent l'homogénéité à l'intérieur de chaque groupe et la minimisent entre ces derniers. [31] Les algorithmes du clustering peuvent être appliqués dans des différents domaines, tel que :

- Découvrir des groupes de clients ayant des comportements semblables.
- Classification des plantes et des animaux étant donné leurs caractéristiques.
- Segmentation des observations des épices pour identifier les zones dangereuses.

f. La description : Parfois le but du Data Mining est de décrire, tout simplement, ce qui se passe sur une Base de Données compliquée en expliquant les relations existantes dans les données pour comprendre le mieux possible les individus, les produits et les processus présents sur cette base. Une bonne description d'un comportement implique souvent une bonne explication de celui-ci. [29]

2.2.6 Les méthodes de Data-Mining

Dans cette partie, nous présentons des techniques d'extraction de données qui sont les plus utilisées et nous mettons en avant la méthode de l'arbre de décision qui sera utilisée dans notre étude.

a. Segmentation (Clustering) : La segmentation est l'opération qui consiste à regrouper les individus d'une population en un nombre limité de groupes. Les segments (ou clusters, ou partitions), qui ont deux propriétés, d'une part, ils ne sont pas prédéfinis,

mais découverts automatiquement au cours de l'opération, contrairement aux classes de la classification. D'autre part, les segments regroupent les individus ayant des caractéristiques similaires et séparent les individus ayant des caractéristiques différentes (homogénéité interne et hétérogénéité externe). La segmentation est une tâche d'apprentissage "*non supervisée*" car on ne dispose d'aucune autre information préalable que la description des exemples. Après application de l'algorithme et donc lorsque les groupes ont été construits, d'autres techniques ou une expertise doivent dégager leur signification et leur éventuel intérêt. Nous présentons ici la méthode des *k*-moyennes car elle est très simple à mettre en œuvre et très utilisée. Elle comporte de nombreuses variantes et est souvent utilisée en combinaison avec d'autres algorithmes.

b. Méthode des k-moyennes : La méthode est basée sur une notion de similarité entre enregistrements. Pour introduire l'algorithme, il faut considérer un espace géométrique muni d'une distance. Deux points sont similaires s'ils sont proches pour la distance considérée. Pour pouvoir visualiser le fonctionnement de l'algorithme, le nombre de champs des enregistrements est limité. L'algorithme suppose choisi *a priori* un nombre *k* de groupes à constituer. On choisit alors *k* enregistrements, soit *k* points de l'espace appelés « centres ». On constitue alors les *k* groupes initiaux en affectant chacun des enregistrements dans le groupe correspondant au centre le plus proche. Pour chaque groupe ainsi constitué, on calcule son nouveau centre en effectuant la moyenne des points du groupe et on réitère le procédé. Le critère d'arrêt est la stabilité, par lequel d'une itération à la suivante, aucun point n'a changé de groupe.

- **Description de l'algorithme :** On travaille avec des enregistrements qui sont des *n*-uplets de valeurs. On suppose définie une notion de similarité qui permet de comparer les distances aux centres. L'algorithme est paramétré par le nombre *k* de groupes que l'on souhaite constituer, le pseudo-code de cet algorithme est présenté dans la table ci-dessous :

Algorithme des k-moyennes

Paramètre : le nombre *k* de groupes

Entrée : un échantillon de *m* objets x_1, \dots, x_m

1- Choisir *k* centres initiaux c_1, \dots, c_k

2- Pour chacun des *m* objets, l'affecter au groupe *i* dont le centre c_i est le plus proche

3- Si aucun élément ne change de groupe alors arrêter et sortir les groupes

4- Calculer les nouveaux centres : pour tout *i*, c_i est la moyenne des éléments du groupe *i*

5- Aller en 2

- **Avantages**

- _ Apprentissage non supervisé : la méthode des k -moyennes et ses variantes ne nécessitent aucune information sur les données. La segmentation peut être utile, pour découvrir une structure cachée qui permettra d'améliorer les résultats de méthodes d'apprentissage supervisé (classification, estimation, prédiction).
- _ Applicable à tout type de données : en choisissant une bonne notion de distance, la méthode peut s'appliquer à tout type de données (mêmes textuelles).

- **Inconvénients**

- _ Problème du choix de la distance : les performances de la méthode (la qualité des groupes constitués) sont dépendantes du choix d'une bonne mesure de similarité ce qui est une tâche délicate surtout lorsque les données sont de types différents.
- _ Le choix des bons paramètres : la méthode est sensible au choix des bons paramètres, en particulier, le choix du nombre k de groupes à constituer. Un mauvais choix de k produit de mauvais résultats. Ce choix peut être fait en combinant différentes méthodes, mais la complexité de l'algorithme augmente.
- _ L'interprétation des résultats : il est difficile d'interpréter les résultats produits, en d'autres termes, d'attribuer une signification aux groupes constitués.

c. Règles d'association : Les règles d'association sont, traditionnellement, liées au secteur de la distribution car leur principale application est "*l'analyse du panier de la ménagère (market basket analysis)*" qui consiste en la recherche d'associations entre produits sur les tickets de caisse. Le but de la méthode est l'étude de ce que les clients achètent pour obtenir des informations sur "*qui*" sont les clients et "*pourquoi*" ils font certains achats. La méthode peut être appliquée à tout secteur d'activité pour lequel il est intéressant de rechercher des groupements potentiels de produits ou de services : services bancaires ou services de télécommunications, par exemple. Elle peut être également utilisée dans le secteur médical pour la recherche de complications dues à des associations de médicaments ou à la recherche de fraudes en recherchant des associations inhabituelles. [32]

- **Avantages**

- non supervisée: à l'exception de la classification de différents articles en produits.
- Clarté des résultats: les règles sont faciles à interpréter.
- Traite des données de taille variables: le nombre de produits dans un achat n'est pas défini.
- Simplicité de programmation: même avec un tableur.
- **Inconvénients**
 - Pertinence des résultats: ils peuvent être triviaux ou inutiles.
 - Efficacité faible dans certains cas : pour les produits rares.
 - Traitement préalable des données : classement des articles en produits.

d. La méthode de plus proches voisins : (*PPV* en bref, *nearest neighbor* en anglais) est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimation. La méthode *PPV* est une méthode de raisonnement à partir de cas. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire. [30]

Contrairement aux autres méthodes de classification qui seront étudiées dans les sections suivantes (arbres de décision, réseaux de neurones, ...), il n'y a pas d'étape d'apprentissage consistant en la construction d'un modèle à partir d'un échantillon d'apprentissage. C'est l'échantillon d'apprentissage, associé à une fonction de distance et d'une fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constitue le modèle. L'algorithme générique de classification d'un nouvel exemple par la méthode *PPV* est dressé suivant :

Algorithme de classification par k-PPV

Paramètre : le nombre k de voisins

Donnée : un échantillon de m exemples et leurs classes //La classe d'un exemple X est $c(X)$

Entrée : un enregistrement Y

- 1- Déterminer les k plus proches exemples de Y en calculant les distances
- 2- Combiner les classes de ces k exemples en une classe c

Sortie : la classe de Y est $c(Y) = c$

- **Avantages**
 - **Absence d'apprentissage:** c'est l'échantillon qui constitue le modèle. L'introduction de nouvelles données permet d'améliorer la qualité de la méthode sans nécessiter la reconstruction d'un modèle. C'est une différence

majeure avec des méthodes telles que les arbres de décision et les réseaux de neurones.

- **Clarté des résultats:** bien que la méthode ne produise pas de règle explicite, la classe attribuée à un exemple peut être expliquée en exhibant les plus proches voisins qui ont amené à ce choix.
- **Inconvénients**
 - **Sélection des attributs pertinents:** pour que la notion de proximité soit pertinente, il faut que les exemples couvrent bien l'espace et soient suffisamment proches les uns des autres. Si le nombre d'attributs pertinents est faible relativement au nombre total d'attributs, la méthode donnera de mauvais résultats car la proximité sur les attributs pertinents sera noyée par les distances sur les attributs non pertinents. Il est donc parfois utile de sélectionner tout d'abord les attributs pertinents.
 - **Le temps de classification:** si la méthode ne nécessite pas d'apprentissage, tous les calculs doivent être effectués lors de la classification. Ceci est la contrepartie à payer par rapport aux méthodes qui nécessitent un apprentissage (éventuellement long) mais qui sont rapides en classification (le modèle est créé, il suffit de l'appliquer à l'exemple à classer). Certaines méthodes permettent de diminuer la taille de l'échantillon en ne conservant que les exemples pertinents pour la méthode *PPV*, mais il faut, de toute façon, un nombre d'exemple suffisamment grand relativement au nombre d'attributs.
 - **Définir les distances et nombre de voisins:** les performances de la méthode dépendent du choix de la distance, du nombre de voisins et du mode de combinaison des réponses des voisins. En règle générale, les distances simples fonctionnent bien. Si les distances simples ne fonctionnent pour aucune valeur de k , il faut envisager le changement de distance, ou le changement de méthode.
- a. **Les réseaux de neurones :** les réseaux de neurones sont apparus dans les années cinquante avec les premiers perceptrons, et sont utilisés industriellement depuis les années quatre-vingt. Un réseau de neurone "*ou réseau neuronal*" a une architecture calquée sur celle du cerveau, organisée en neurones et synapses, et se présente comme un ensemble de nœuds "*ou neurones formels, ou unités*" connectés entre eux, chaque variable prédictive continue correspondant à un nœud d'un premier niveau, appelé

couche d'entrée, et chaque variable prédictive catégorique (ou chaque modalité d'une variable catégorique) correspondant également à un nœud de la couche d'entrée. [34]

▪ **Les avantages**

- **Lisibilité du résultat:** le résultat de l'apprentissage est un réseau constitué de cellules organisées selon une architecture, définies par une fonction d'activation et un très grand nombre de poids à valeurs réelles.
- **Les données réelles:** les réseaux traitent facilement les données réelles "préalablement normalisées" et les algorithmes sont robustes au bruit. Ce sont, par conséquent, des outils bien adaptés pour le traitement de données complexes éventuellement bruitées comme la reconnaissance de formes (son, images sur une rétine, etc.).
- **Classification efficace:** le réseau étant construit, le calcul d'une sortie à partir d'un vecteur d'entrée est un calcul très rapide.
- **En combinaison avec d'autres méthodes:** pour des problèmes contenant un grand nombre d'attributs pour les entrées, il peut être très difficile de construire un réseau de neurones. On peut, dans ce cas, utiliser les arbres de décision pour sélectionner les variables pertinentes, puis générer un réseau de neurones en se restreignant à ces entrées.

▪ **Les inconvénients**

- **Temps d'apprentissage:** l'échantillon nécessaire à l'apprentissage doit être suffisamment grand et représentatif des sorties attendues. Il faut passer un grand nombre de fois tous les exemples de l'échantillon d'apprentissage avant de converger et donc le temps d'apprentissage peut être long.
- **Evolutivité dans le temps:** comme pour les arbres de décision, l'apprentissage n'est pas incrémental et, par conséquent, si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage pour s'adapter à cette évolution.

b. Arbre de décision : la méthode des arbres de décision est l'une des plus intuitives et des plus populaires du data-mining, d'autant plus qu'elle fournit des règles explicites de classement et supporte bien les données hétérogènes, manquantes et les effets non linéaires. Pour les applications relevant du marketing de bases de données, actuellement la seule grande concurrente de l'arbre de décision est la régression logistique, cette méthode étant préférée dans la prédiction du risque en raison de sa plus grande robustesse. Remarquons que les arbres de décision sont à la frontière entre

les méthodes prédictives et descriptives, puisque leur classement s'opère en segmentant la population à laquelle ils s'appliquent : ils ressortissent donc à la catégorie des classifications hiérarchiques descendantes supervisées, (voir la **Figure 9**).

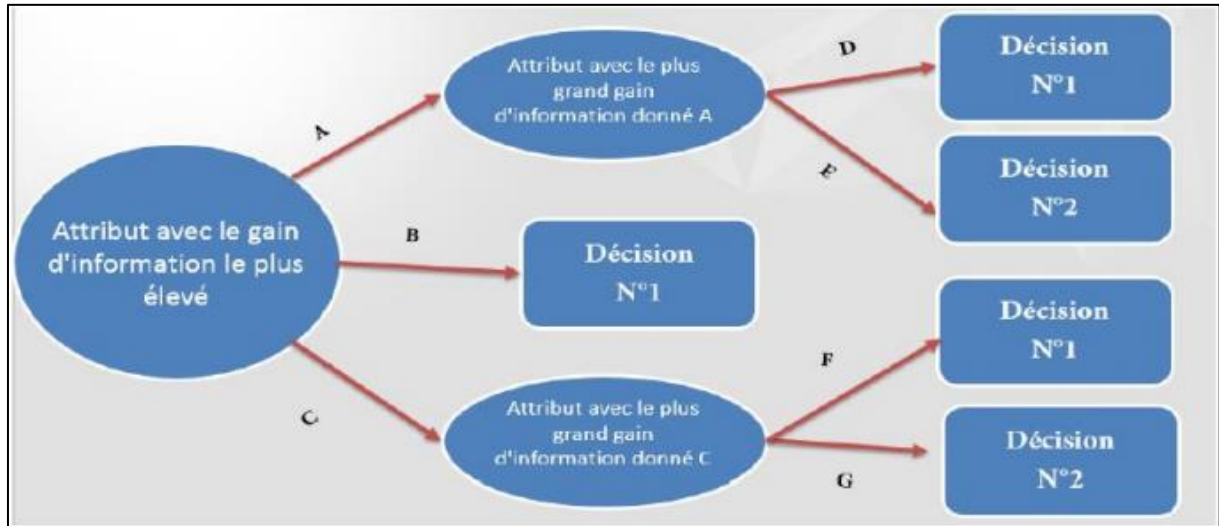


Figure 9: Arbre de décision

- Description de l'algorithme :** la technique de l'arbre de décision est employée en classement pour détecter des critères permettant de répartir les individus d'une population en n classes (souvent $n=2$) prédéfinies. On commence par choisir la variable qui, par ses modalités, sépare le mieux les individus de chaque classe, de façon à avoir des sous-populations, que l'on appelle nœuds, contenant chacune le plus possible d'individus d'une seule classe, puis on réitère la même opération sur chaque nouveau nœud obtenu jusqu'à ce que la séparation des individus ne soit plus possible ou plus souhaitable. Par construction, les nœuds terminaux (les feuilles) sont tous majoritairement constitués d'individus d'une seule classe avec une assez forte probabilité, quand il satisfait l'ensemble des règles permettant d'arriver à cette feuille. L'ensemble des règles de toutes les feuilles constitue. L'algorithme d'apprentissage par l'arbre de décision est décrit comme suit :

Algorithme d'apprentissage par arbre de décision

Donnée : un échantillon S de m enregistrements classés $(x, c(x))$

Initialisation : $A \leftarrow$ arbre vide
 noeud_courant \leftarrow racine
 échantillon_courant $\leftarrow S$

Répéter
 Décider si le nœud courant est terminal

Si (noeud_courant est terminal) alors

Étiqueter le nœud courant par une feuille

Sinon

Sélectionner un test :

Créer les fils

Définir les échantillons sortants du nœud

Finsi

noeud_courant ← un nœud non encore étudié de A

échantillon_courant : échantillon atteignant noeud_courant

Jusque (noeud_courant = \emptyset)

élaguer l'arbre de décision A obtenu

Sortie : l'arbre A élagué

▪ **Avantages**

- **Adaptabilité aux attributs de valeurs manquantes:** les algorithmes peuvent traiter les valeurs manquantes (descriptions contenant des champs non renseignés) pour l'apprentissage, mais aussi pour la classification.
- **Bonne lisibilité du résultat:** un arbre de décision est facile à interpréter et à la représentation graphique d'un ensemble de règles. Si la taille de l'arbre est importante, il est difficile d'appréhender l'arbre dans sa globalité. Cependant, les outils actuels permettent une navigation aisée dans l'arbre (parcourir une branche, développer un nœud, élaguer une branche) et, le plus important, est certainement de pouvoir expliquer comment est classé un exemple par l'arbre, ce qui peut être fait en montrant le chemin de la racine à la feuille pour l'exemple courant.
- **Traitement de tout type de données:** l'algorithme peut prendre en compte tous les types d'attributs et les valeurs manquantes. Il est robuste au bruit.
- **Sélectionne des variables pertinentes:** l'arbre contient les attributs utiles pour la classification. L'algorithme peut donc être utilisé comme prétraitement qui permet de sélectionner l'ensemble des variables pertinentes pour ensuite appliquer une autre méthode.
- **Donne une classification efficace:** l'attribution d'une classe à un exemple à l'aide d'un arbre de décision est un processus très efficace (parcours d'un chemin dans un arbre).
- **Disponibilité des outils:** les algorithmes de génération d'arbres de décision sont disponibles dans tous les environnements de fouille de données.

- **Méthode extensible et modifiable:** la méthode peut être adaptée pour résoudre des tâches d'estimation et de prédiction. Des améliorations des performances des algorithmes de base sont possibles grâce aux techniques qui génèrent un ensemble d'arbres votant pour attribuer la classe.
- **Inconvénients**
 - **Méthode sensible au nombre de classes:** les performances tendent à se dégrader lorsque le nombre de classes devient trop important.
 - **Manque d'évolutivité dans le temps:** l'algorithme n'est pas incrémental, c'est-à-dire, que si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage sur l'échantillon complet (anciens exemples et nouveaux exemples).

2.2.7 Les étapes du processus du Data-Mining

Le processus du data-mining se fait par plusieurs étapes, on cite comme suit :

- a. **Collecte des données :** la combinaison de plusieurs sources de données, souvent hétérogènes, dans une base de données. [35]
- b. **Nettoyage des données :** la normalisation des données : l'élimination du bruit (les attributs ayant des valeurs invalides et les attributs sans valeurs). [36]
- c. **Sélection des données :** Sélectionner de la base de données les attributs utiles pour une tâche particulière du data-mining. [38]
- d. **Transformation des données :** le processus de transformation des structures des attributs pour être adéquates à la procédure d'extraction des informations. [39]
- e. **Extraction des informations (data-mining):** l'application de quelques algorithmes du data-mining sur les données produites par l'étape précédente (*Knowledge Discovery in Databases*, ou KDD). [36]
- f. **Visualisation des données :** l'utilisation des techniques de visualisation (histogramme, camembert, arbre, visualisation 3D) pour exploration interactive de données (la découverte des modèles de données). [36]
- g. **Evaluation des modèles :** l'identification des modèles strictement intéressants en se basant sur des mesures données. [39]

Ces étapes sont montrées dans la **Figure 10** :

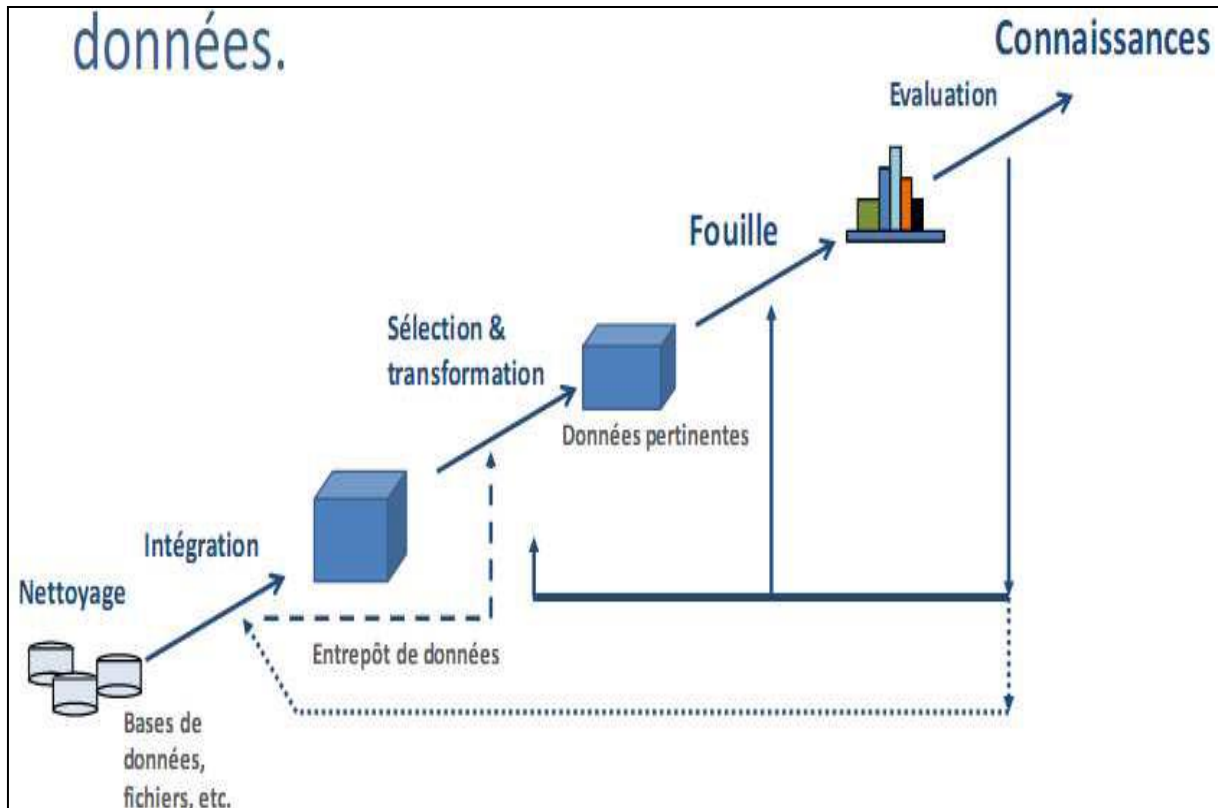


Figure 10: Le processus du data-mining

2.2.8 Catégorisation des systèmes du Data-Mining

Les systèmes de data-mining peuvent être catégorisés selon plusieurs critères. Parmi les catégorisations existantes nous pouvons citer:

- a. **Classification selon le type de données à explorer** : dans cette classification les systèmes de data-mining sont regroupés selon le type des données qu'ils manipulent tel que les données spatiales, les données de séries temporelles, les données textuelles et le Word Wide Web, etc. [39]
- b. **Classification selon les modèles de données avancés** : cette classification catégorise les systèmes de data-mining en se basant sur les modèles de données avancés tel que les bases de données relationnelles, les bases de données orientées objets, les data warehouses, les bases de données transactionnelle, etc. [39]
- c. **Classification selon le type de connaissance à découvrir** : cette classification catégorise les systèmes de data-mining en s'appuyant sur le type de connaissance à découvrir ou les tâches de data-mining tel que la classification, l'estimation, la prédiction, etc. [39]
- d. **Classification selon les techniques d'exploration utilisées** : cette classification catégorise les systèmes de data-mining suivant l'approche d'analyse de données

utilisés la reconnaissance des formes, les réseaux neurones, les algorithmes génétiques, les statistiques, la visualisation, orienté-base de données ou orienté-data warehouse, etc. [39]

2.2.9 Domaines d'application du Data-Mining

Dans cette section, nous présentons des exemples des applications possibles du datamining par secteur d'activités :

- a. **Grande distribution et VPC (Nuage Privé Virtuel):** Analyse des comportements des consommateurs, recherche des similarités des consommateurs en fonction de critères géographiques ou sociodémographiques, prédiction des taux de réponse en marketing direct, vente croisée et activation sélective dans le domaine des cartes de fidélité, optimisation des réapprovisionnements.
- b. **Laboratoires pharmaceutiques:** Modélisation comportementale et prédiction de médications ou de visites, optimisation des plans d'action des visiteurs médicaux pour le lancement de nouvelles molécules, identification des meilleures thérapies pour différentes maladies.
- c. **Banques :** Modélisation prédictive des clients partants, détermination de pré-autorisations de crédit.
- d. **Assurance :** Modèles de sélection et de tarification, analyse des sinistres, recherche des critères explicatifs du risque ou de la fraude, prévision d'appel sur les plateformes d'assurance directe.
- e. **Aéronautique, automobile et industries :** Contrôle qualité et anticipation des défauts, prévision des ventes, dépouillement d'enquêtes de satisfaction.
- f. **Transport et voyagistes :** Optimisation des tournées, prédiction de carnets de commande, marketing relationnel dans le cadre de programmes de fidélité.
- g. **Télécommunications, eau, énergie :** Simulation de tarifs, détection de formes de consommation frauduleuses, classification des clients selon la forme de l'utilisation des services, prévision de ventes.

Comme on peut le voir, le datamining peut s'appliquer à tous les domaines, (voir la **Figure 11** ci-dessous) :

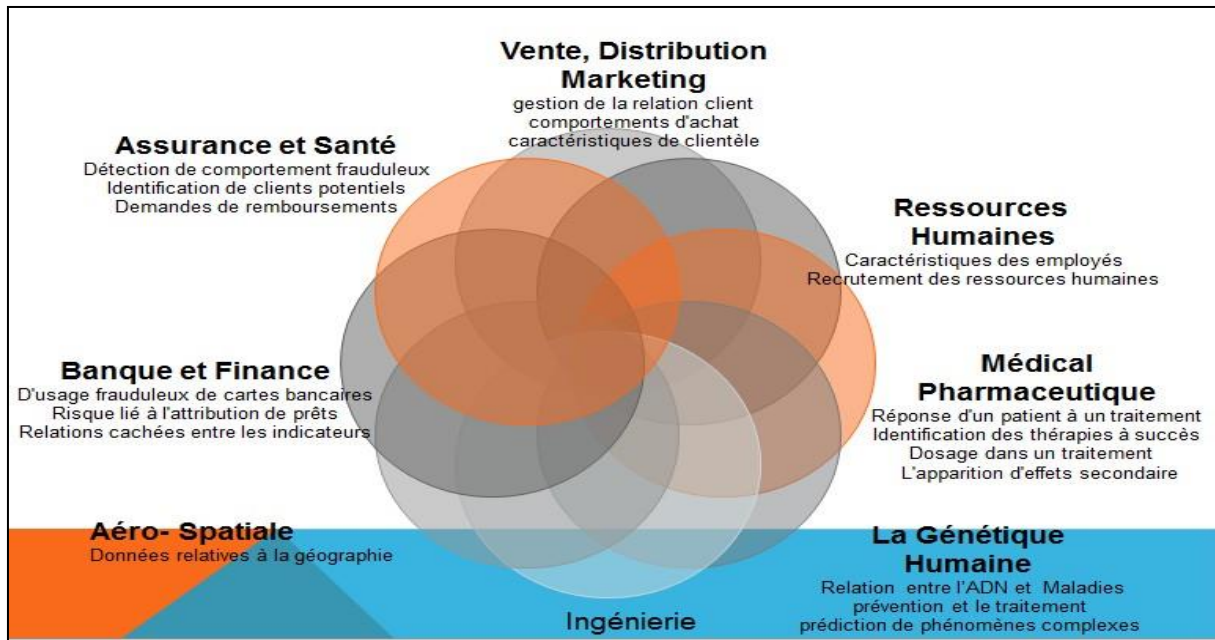


Figure 11: Domaines d'application du data-mining

2.3 Conclusion

Dans ce deuxième chapitre, nous avons présenté les principaux concepts de fouille de données, les processus, les tâches et les méthodes les plus utilisés en datamining ainsi que les avantages et les inconvénients de chaque méthode. Le data-mining est une méthodologie qui automatise la synthèse de connaissances à partir de gros volumes de données. L'essor de cette technologie est le résultat d'un accroissement dramatique de l'information numérique qui, de par son abondance, est sous-exploitée sans outil et expertise adéquats. Cette technologie repose sur une diversité de techniques (intelligence artificielle, statistiques, théorie de l'information, génie logiciel, bases de données,...) qui requièrent des compétences variées et de haut niveau.

Chapitre 03 : Conception et Implémentation du Système Proposé

3.1 Partie I. Conception du système proposé

Après avoir fait un bon état de l'art sur l'internet des objets et ses déferents domaines d'application (santé, sécurité, industrie et l'agriculture), nous avons choisi le domaine agriculture comme un domaine d'étude parce que c'est un domaine très riche d'idées et très influent dans l'économie, ainsi nous essayons de trouver une nouvelle idée et de la présenter à travers ce travail.

Dans ce chapitre nous commençons par une petite introduction et ensuite nous présentons la problématique et l'objectif de ce travail. Nous discutons quelques travaux connexes, puis nous présentons le modèle proposé. Nous terminerons ce chapitre par une conclusion.

3.1.1 Problématique et objectifs

Ils existent plusieurs travaux qui sont réalisés à l'aide de l'internet des objets pour faciliter certaines tâches et les terminer en un temps record. En agriculture, l'agriculteur souffre en apportant une nouvelle culture dans son ferme et en essayant de l'essayer, il ne sait pas si cette culture est compatible ou non à son ferme, alors il essaie et attend que sa maturité grandisse et a peur de réussir ou non. Si son expérience ne réussira pas, il aurait perdu du temps et de l'argent. Alors, comment pouvons-nous aider l'agriculteur à accomplir ces tâches rapidement et sans perdre du temps et d'argent?

Notre objectif principal est de créer un système intelligent qui travaille sur la prise de décisions concernant la compatibilité de l'usine avec notre ferme, afin de réaliser plus de profits lorsque nous connaissons les plantes appropriées et d'éviter l'inapproprié.

3.1.2 Travaux connexes

Dans l'IoT, il ya plusieurs systèmes proposés dans le domaine de l'agriculture intelligente pour trouver des solutions aux problèmes de l'agriculture traditionnels. Cette dernière gaspillaient l'argent, le temps et l'énergie humaine. Dans cette section, nous présentons quelques travaux connexes qui sont récemment, été proposés :

- a. En Inde, en 2019, Abhishek et Rishi Barath ont présenté un article qui vise à améliorer l'efficacité du secteur agricole en Inde en utilisant les facteurs météorologiques retardés pour prédire le meilleur produit pouvant être cultivé à travers l'arbre de décision. Ce travail a été appelé TDMA.
- b. En Italie, en 2018, Fabrizio Balducci, Donato Impedovo et Giuseppe Pirlo ont présenté un travail visant à gérer et adapter des données hétérogènes collectées à partir de

différentes sources et à les exploiter dans une agriculture intelligente en extrayant des connaissances qui contribuent à réduire le gaspillage d'argent et d'énergie.

- c. En juin 2019, les deux chercheurs ont proposé un travail dans le domaine de l'agriculture intelligente qui dépend de l'Internet des Objets et de l'apprentissage automatique. Ce travail vise à traiter les données, à les analyser et à prévoir des décisions précises qui contribuent à la maîtrise de l'agriculture et à l'augmentation de la production.

Les travaux cités, précédemment, tournent, tous, autour de la même idée, à savoir, réduire les pertes pendant la culture, soit en exploitant les données pour extraire de nouvelles connaissances sur la culture et ce qui lui est utile, soit en anticipant des décisions précises qui peuvent être utiles pour augmenter ou améliorer la production. De ce point de vue, nous avons voulu ajouter une nouvelle idée dans l'Internet des objets, qui est d'étudier la compatibilité de la plantation avec notre ferme en exploitant plusieurs facteurs et cette étude se fait avant le début du processus de plantation et cela à travers l'utilisation d'un arbre de décision. Si le champ est approprié, la plantation est faite et si elle n'est pas adaptée au type de culture, le processus de plantation est annulé sans pertes.

3.1.3 Modèle proposé

Le modèle proposé est un système composé de plusieurs éléments et chacun d'eux a un rôle important. Nous l'expliquons en détail dans les sous-sections suivantes.

3.1.4 Aperçu de l'architecture proposée du système

L'Internet des objets introduit le concept d'objet connecté, c'est-à-dire l'interconnexion entre l'objet et Internet, afin de surveiller et d'améliorer les performances de l'objet. L'architecture de notre système est illustrée dans la **Figure 12** ci-dessous. Les principaux éléments du système sont :

- a. **Le capteur** : son rôle est de capturer les données de l'objet connecté et de les envoyer au traiteur de données.
- b. **Le traiteur** : est celui qui fait le prétraitement de données collectées par le capteur afin de les préparer pour le passage au décideur.
- c. **Le décideur**: il fait passer les données traitées à travers l'arbre de décision pour prendre la décision.
- d. **L'interface** : joue le rôle d'intermédiaire entre l'utilisateur et le système.
- e. **L'utilisateur**: c'est lui qui a la capacité d'utiliser l'application.

f. La base de données: permet de stocker les données.

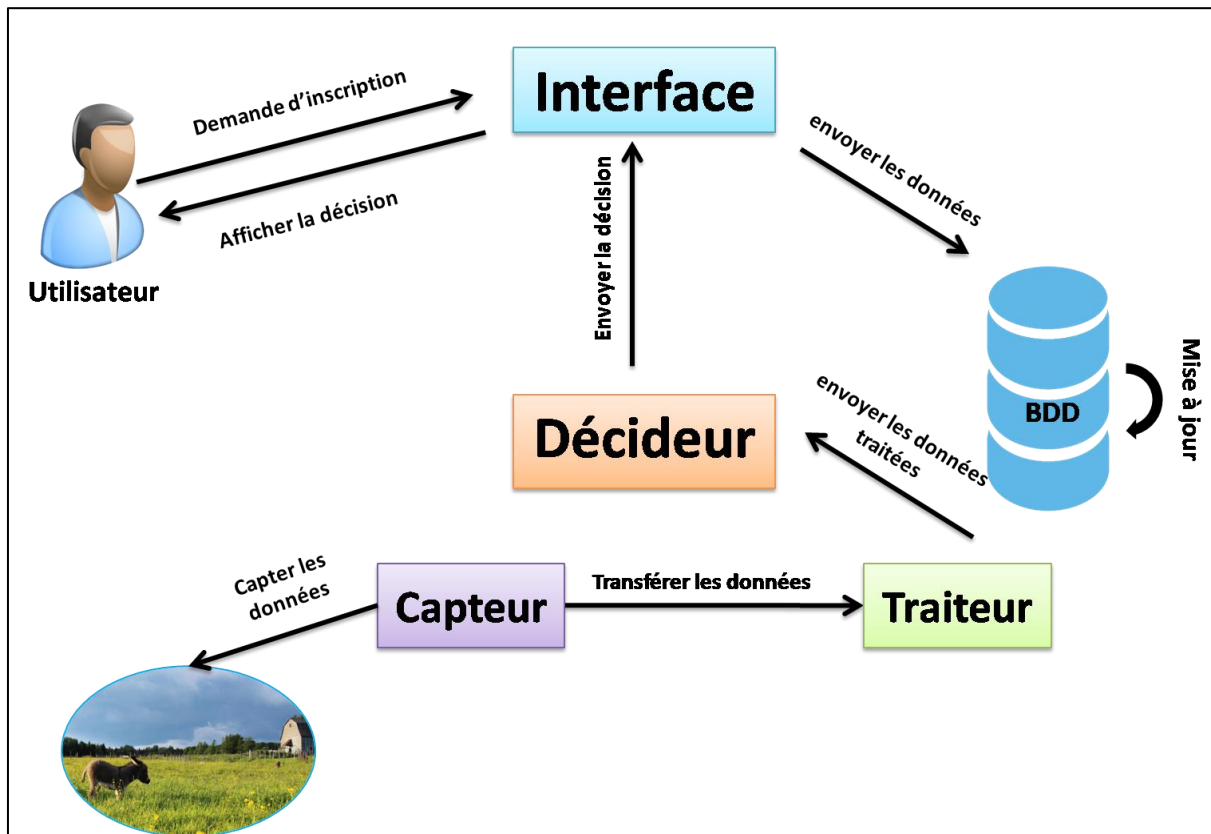


Figure 12: L'Architecture du système proposé

3.1.5 Fonctionnement du système proposé

Dans cette section, nous présentons comment notre système fonctionne et répond aux objectifs fixés au début de ce travail. Ceci est fait à travers un diagramme de séquence comme le montre la **Figure 13**.

Chapitre 03 : Conception et Implémentation du Système Proposé

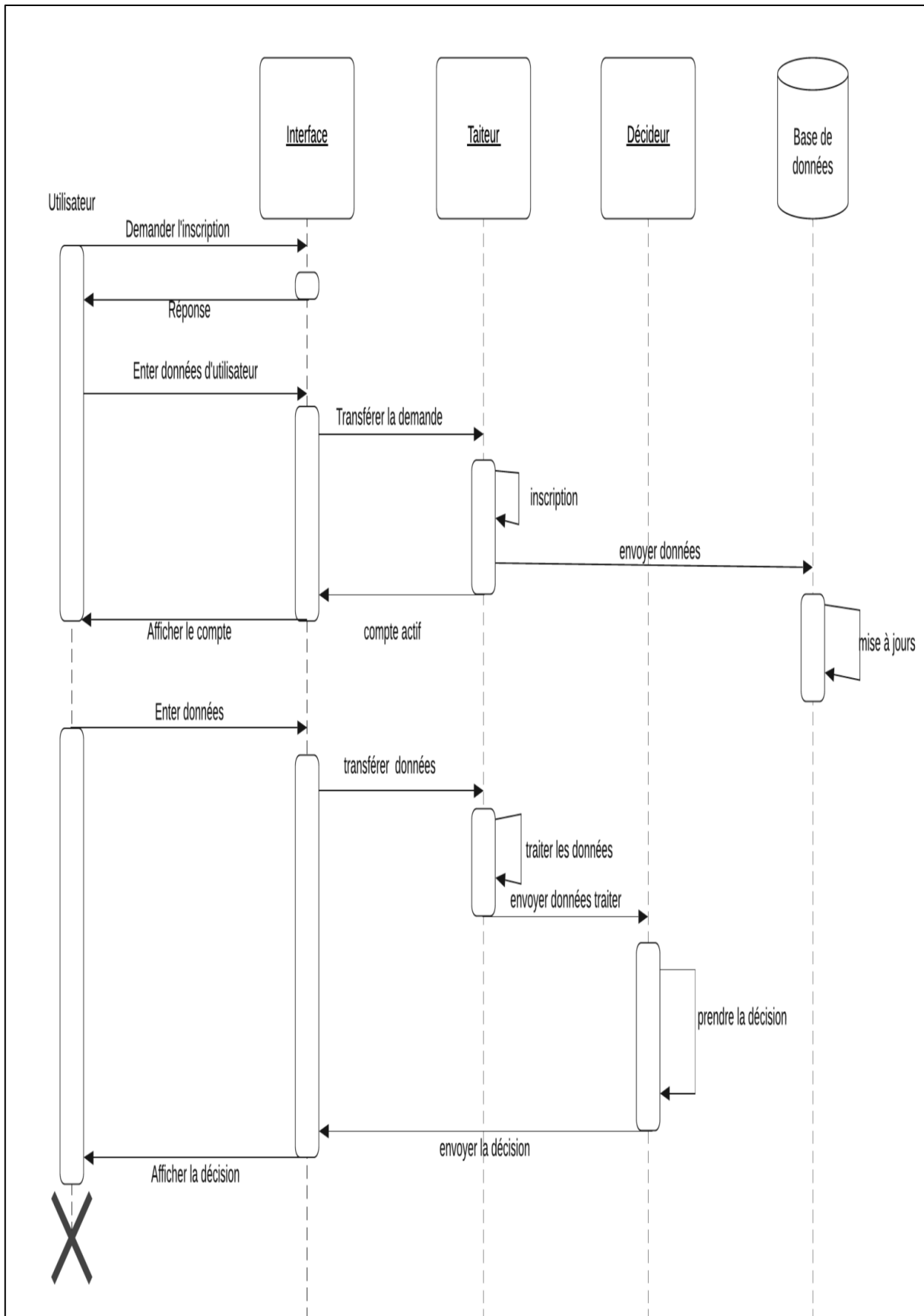


Figure 13: Diagramme de séquence de notre système

L'utilisateur de l'application doit, pour la première utilisation du système, créer son propre

compte de données. Une fois qu'un compte est créé, il devient en mesure d'utiliser l'application. L'utilisateur saisit les données de sa ferme avec ses propres caractéristiques (cette tâche se fait uniquement en l'absence du capteur). Dans le cas de la présence du capteur sans fils, c'est celui qui capture les données nécessaires et les envoie au traiteur. Le traiteur déclenche un prétraitement des données pour la transmission au décideur. Le décideur transmet les données à travers l'arbre de décision pour prendre une décision à la fin qui l'envoie à l'interface et qui soit affiché à l'utilisateur.

3.1.6 L'Arbre de décision proposé

L'arbre de décision correspondant est décrit ci-dessous (voir la **Figure 14**)

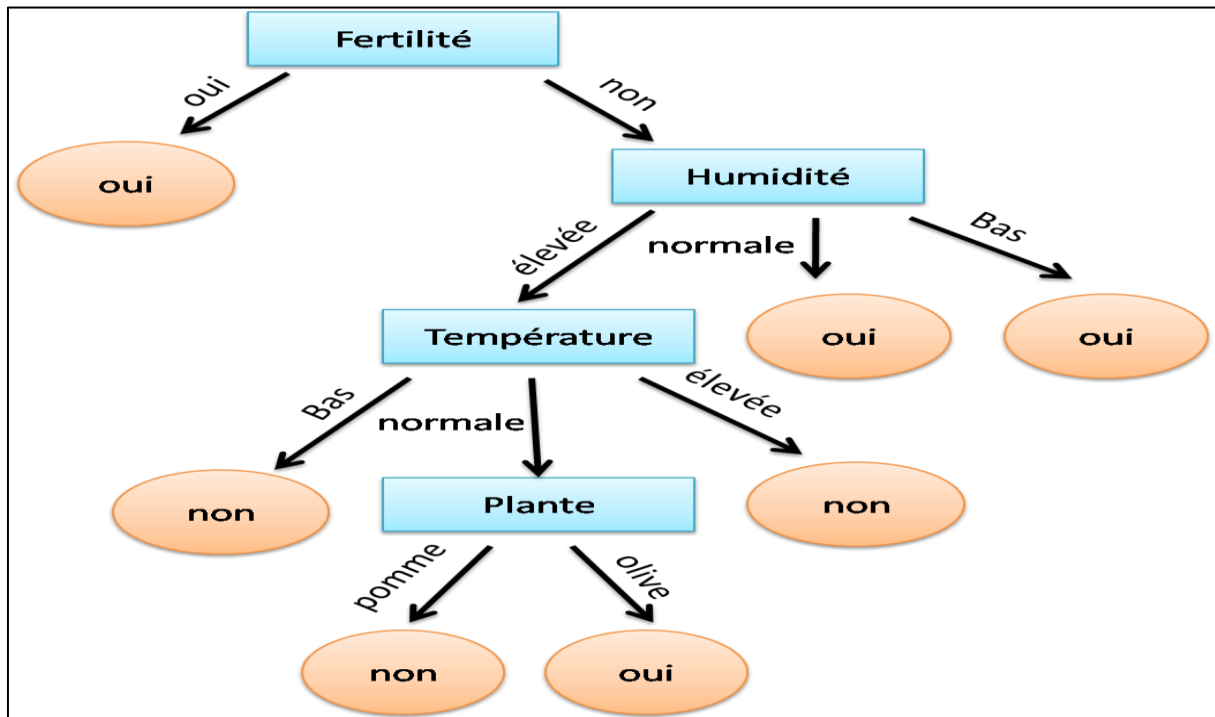


Figure 14: l'Arbre de décision proposé

- Le premier sommet est appelé la « racine » de l'arbre. Il est situé sur le premier niveau. Nous y observons la distribution de fréquence de la variable à prédire «Fertilité». Nous constatons qu'il y a bien 13 observations, dont 6 « oui » et 7 « non ».
- La variable «Humidité» est la deuxième variable utilisée ; on parle de variable de segmentation. Comme elle est composée de 3 modalités {élevée, normale, Bas}, elle produit donc 3 sommets enfants.

Chapitre 03 : Conception et Implémentation du Système Proposé

- La première arête (la première branche), à gauche, sur le deuxième niveau, est produite à partir de la modalité « élevée » du variable « Humidité ».
- La seconde arête, au centre, correspond à la modalité « normale » de la variable de segmentation « Humidité » ;
- La dernière arête, correspond à la modalité « Bas » de la variable de segmentation « Humidité » ;
- La variable «Température» est la troisième variable utilisée ; Comme elle est composée de 3 modalités {élevée, normale, Bas}, elle produit donc 3 sommets enfants.
- La quatrième variable utilisée est «Plante», Comme elle est composée de 3 modalités {Pomme, olive, pomme-terre}, elle produit donc 3 sommets enfants.

Cet arbre de décision a été créé en fonction des données présentées dans la **Figure 15** ci-dessous :

	Plante	Température	Humidité	Fertilité	Résultat
1	Pomme	Elevée	Elevée	Oui	Oui
2	Pomme	Elevée	Normale	Oui	Non
3	Olive	Elevée	Faible	Oui	Oui
4	Pomme-terre	Normal	Elevée	Oui	Oui
5	Pomme-terre	Normal	Normale	Oui	Non
6	Pomme-terre	Normal	Faible	Oui	Oui
7	Olive	Faible	Elevée	Oui	Non
8	Pomme	Faible	Normale	Non	Non
9	Pomme	Faible	Faible	Non	Non
10	Pomme-terre	Elevée	Elevée	Non	Oui
11	Pomme	Elevée	Normale	Non	Non
12	Olive	Elevée	Faible	Non	Non
13	Olive	Normal	Elevée	Non	oui

Figure 15: Dataset utilisé pour la construction de l'arbre de décision

Cette première partie a donné une vision sur notre travail. Elle a donné l'aspect conceptuel de notre système à travers la présentation de son architecture, son fonctionnement

en utilisant le diagramme de séquence et l'arbre de décision proposé. Dans la prochaine partie, on se concentrera sur l'implémentation de notre application.

3.2 Partie II. Implémentation du système proposé

L'implémentation est la phase la plus importante après celle de la conception. Le choix des outils de développement influence énormément sur le coût en temps de programmation, ainsi que sur la flexibilité du produit à réaliser. Cette phase consiste à transformer le modèle conceptuel établi précédemment en des composants logiciels formant notre système. Dans cette partie, nous allons commencer par la description de l'environnement de travail puis, nous allons présenter les composants élaborés, de notre système, via la présentation et la description de quelques fenêtres de notre application.

3.2.1 Environnement de travail

L'environnement de travail est constitué de deux parties nommées environnement matériel et environnement logiciel.

- a. **Environnement matériel** : le développement de l'environnement matériel est caractérisé par : (voir la **Figure 16**)

Type(s)	Portable bureautique
Processeur	Intel Core i3-370M
Fréquence processeur	2,4 GHz
Mémoire vive (RAM)	4 Go
Type de mémoire vive	DDR3-SDRAM
Capacité de stockage principal	750 Go
Type de stockage principal	5400tr/mn
Interface stockage principal	Serial ATA

Figure 16: Caractéristique se notre PC portable

- b. **Environnement logiciel** : le système d'exploitation installé sur le PC utilisé dans notre projet et caractérisé par : (voir la **Figure 17**)

Édition	Windows 10 Entreprise 2015 LTSB
ID de produit	00329-50000-00001-AA753
Processeur	Intel(R) Core(TM) i3-5010U CPU @ 2.10GHz 2.10 GHz
Mémoire RAM installée	4,00 Go
Type du système	Système d'exploitation 64 bits, processeur x64
Stylet et fonction tactile	La fonctionnalité d'entrée tactile ou avec un stylet n'est pas disponible sur cet écran

Figure 17: Caractéristiques du système d'exploitation

3.2.2 Outils de développement

Nous avons utilisé plusieurs outils de programmation. Dans ce qui suit, nous allons les mentionner.

- a. **Plateforme Node.js** : **Node.js** est une plateforme logicielle libre en JavaScript orientée vers les applications réseau, événementielles hautement concurrentes, qui doivent pouvoir monter en charge. Elle utilise la machine virtuelle V8, la librairie libuv pour sa boucle d'évènements, et implémente sous licence MIT les spécifications CommonJS. Parmi les modules natifs de Node.js, on trouve http qui permet le développement de serveur HTTP. Il est donc possible de se passer de serveurs web tels que Nginx ou Apache lors du déploiement de sites et d'applications web développés avec Node.js. Concrètement, Node.js est un environnement bas niveau permettant l'exécution de JavaScript côté serveur.
- b. **JavaScript Object Notation (JSON)** : un format de données textuelles est dérivé de la notation des objets du langage JavaScript. Il permet de représenter de l'information structurée comme le permet XML par exemple. Créé par Douglas Crockford entre 2002 et 2005, la première norme du JSON est ECMA-404 qui a été publiée en octobre 2003, il est actuellement décrit par les deux normes en concurrence : RFC 82593 de l'IETF et ECMA-4044 de l'ECMA. La dernière version des spécifications du format date de décembre 2017. Des bibliothèques pour le format JSON existent dans la plupart des langages de programmation.
- c. **React Native** : React Native est un framework d'applications mobiles open source créé par Facebook. Il est utilisé pour développer des applications pour Android 3, iOS et UWP 4 en permettant aux développeurs d'utiliser React avec les fonctionnalités natives de ces plateformes.
- d. **Langage de Manipulation de Bases de Données MongoDB** : MongoDB est un système de gestion de base de données orienté documents, répartitionnable sur un nombre

quelconque d'ordinateurs et ne nécessitant pas de schéma prédéfini des données. Il est écrit en C++. Le serveur et les outils sont distribués sous licence SSPL, les pilotes sous licence Apache et la documentation sous licence Creative Commons. Il fait partie de la mouvance NoSQL.

Nous avons utilisé aussi une suite de logiciels de développement pour Windows et mac OS conçue par Microsoft. La dernière version s'appelle Visual Studio 2019.

- a. **Logiciel de développement Microsoft Visual Studio** : Visual Studio est un ensemble complet d'outils de développement permettant de générer des applications web ASP.NET, des services web XML, des applications bureautiques et des applications mobiles. Visual Basic, Visual C++, Visual C# utilisent tous le même environnement de développement intégré (IDE), qui leur permet de partager des outils et facilite la création de solutions faisant appel à plusieurs langages. Par ailleurs, ces langages permettent de mieux tirer parti des fonctionnalités du Framework .NET, qui fournit un accès à des technologies clés simplifiant le développement d'applications web ASP et de services web XML grâce à Visual Web Developer.

3.2.3 Description de l'interface

Notre application mobile s'appelle « Smart Farm », en français la ferme intelligente. Lors de sa première utilisation, on trouve au début un formulaire d'inscription à remplir par l'utilisateur .(Voir la **Figure 18**)



Figure 18: Formulaire d'Inscription

Quant aux personnes qui ont déjà un compte, il leur apparaîtra tel qu'il est montré dans la **Figure 19**:

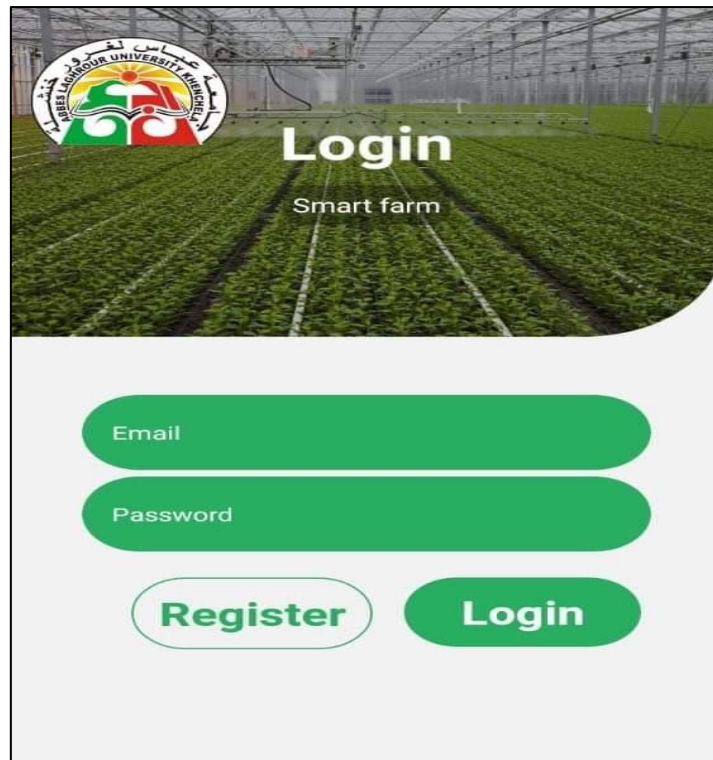


Figure 19: Interface de connexion

Après avoir connecté au compte, la fenêtre illustrée par la **Figure 20** apparaîtra :

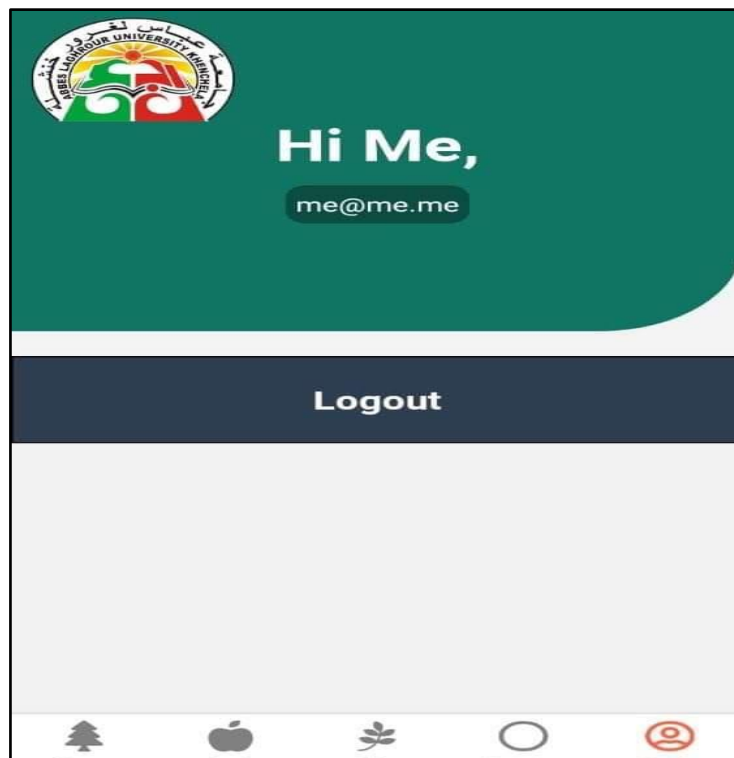


Figure 20: Profil d'utilisateur

La barre inférieure contient cinq composants. (Voir la **Figure 21**)

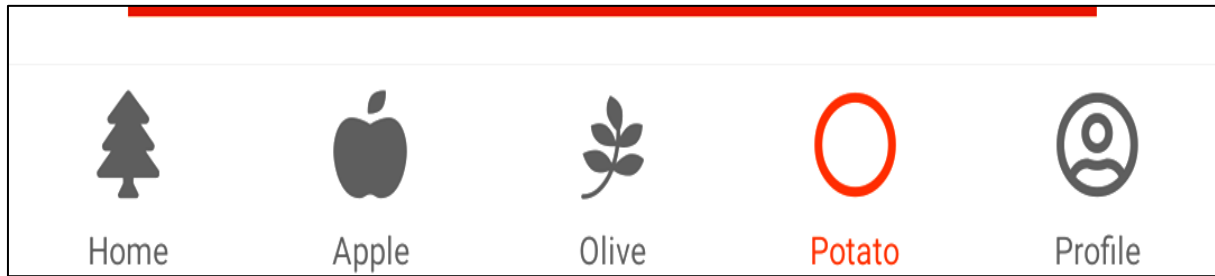


Figure 21: Contenu de la barre inférieure

En entrant dans le « Home », nous trouverons les données spécifiques à l'exploitation de la ferme, voir la **Figure 22** ci-dessous :



Figure 22: La fenêtre Home

En sélectionnant les données pour la ferme d'utilisateur et en cliquant sur "check", le résultat apparaîtra et il s'agit d'une liste des plantes dans DataSat apparaît, et chacune d'elle a sa propre décision, qui peut être cultivée ou non. (Voir la **Figure 23**)

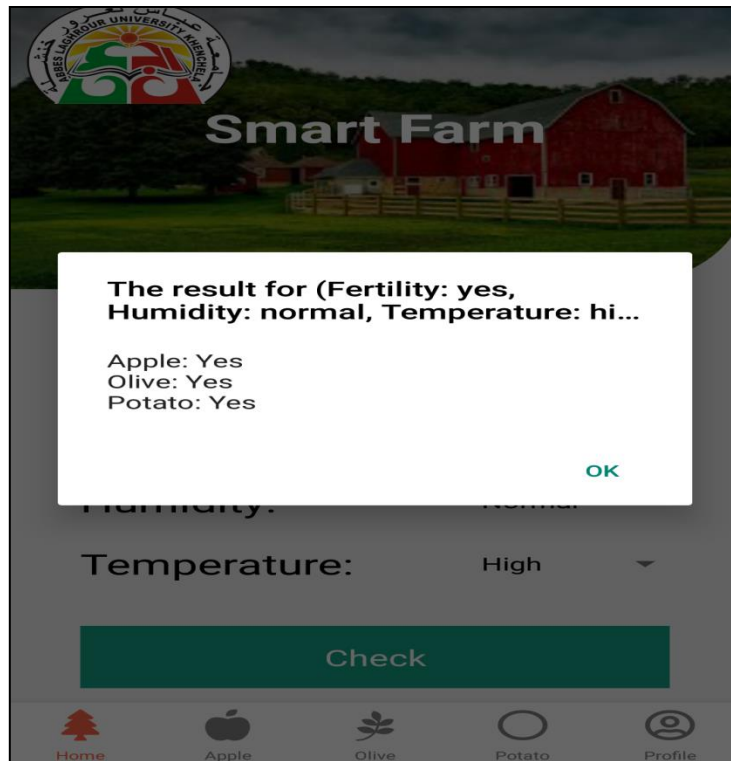


Figure 23: Résultat d'exécution

En cliquant sur « Apple », la fenêtre illustrée par la Figure 24 apparaîtra :

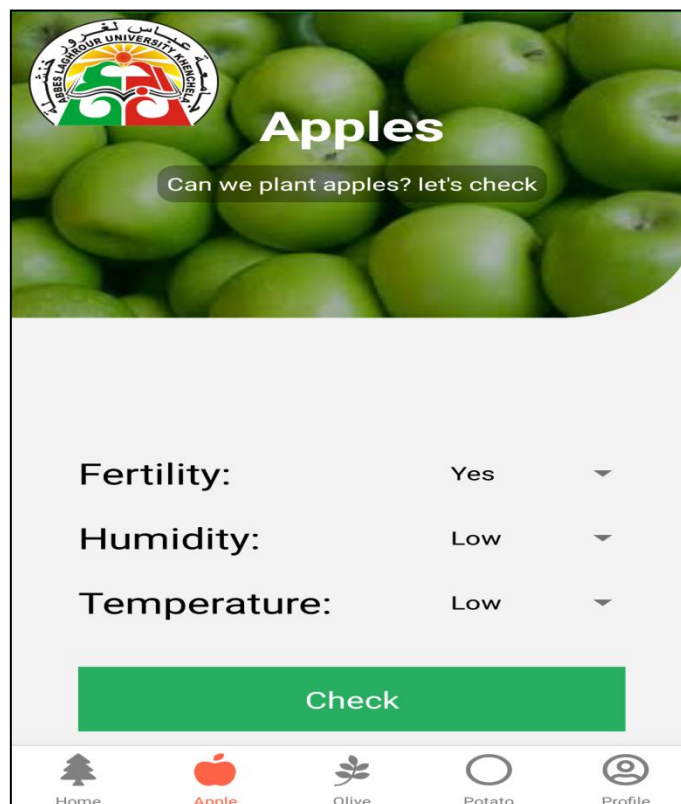


Figure 24: Contenu de la fenêtre « Apple »

En sélectionnant les données pour la ferme d'utilisateur et en cliquant sur "check", le résultat montre une décision, à savoir si les pommes peuvent être cultivées dans cette ferme ou non. (Voir la **Figure 25**)

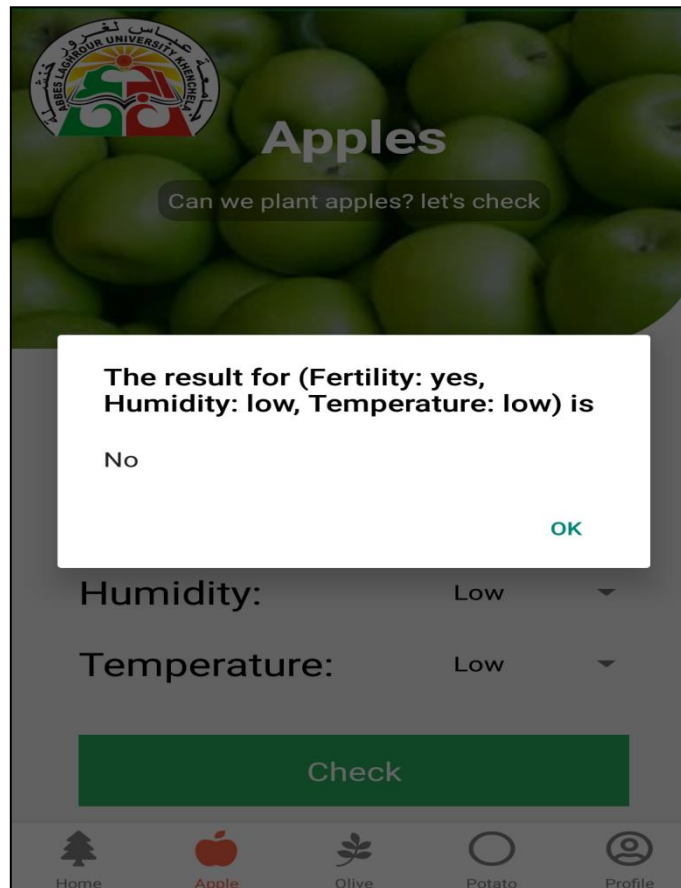


Figure 25: Résultat du « Apple »

Le fonctionnement de l'application est le même pour les fenêtres « Olive » et « Potato ».

3.2.4 Difficultés techniques

On a rencontré plusieurs problèmes lors de la réalisation de notre travail. Parmi ces derniers, on peut citer:

- Le confinement soudain, qui a conduit à la fermeture complète des universités et des instituts, c'est-à-dire la difficulté d'obtenir des travaux antérieurs.
- L'absence de bureau d'études agricoles dans notre pays qui nous a empêchés d'obtenir un DataSet.
- La difficulté d'obtenir un capteur sans fil.

Mais malgré les nombreuses difficultés, nous avons essayé de trouver des solutions et d'achever notre travail. Grâce à notre implémentation, nous pouvons dire qu'il est très

important de faire le bon choix concernant la bonne plateforme à utiliser pour obtenir des résultats fiables.

3.3 Conclusion

Dans ce chapitre, nous avons présenté quelques résultats obtenus de l'implémentation de notre système. Nous avons utilisé, pour l'implémentation de notre système, plusieurs outils tels que : Plateforme Node.js, JavaScript Object Notation (JSON), React Native, Langage de Manipulation de Bases de Données MongoDB et Logiciel de développement Microsoft Visual Studio. L'utilisation de notre application permet d'améliorer les cultures d'une ferme donnée et ceci en aidant un agriculteur à prendre les décisions nécessaires et dans les délais voulus.

Conclusion générale et perspectives

L'agriculture étant l'épine dorsale de tout pays, il est nécessaire d'assurer sa croissance durable au cours des années. Un système décisionnel approprié avec des instructions précises permettrait toujours de fournir de meilleurs résultats.

Pour réaliser ce mémoire, nous avons d'abord rassemblé le plus de documents concernant le sujet. Tant un domaine relativement récent, malheureusement, il existe peu de livres liés à ce sujet, on a aussi le problème de confinement et heureusement, on a trouvé des thèses, articles et autres papiers qui étaient disponibles à ce sujet.

La deuxième étape consiste à introduire notre domaine d'étude à l'aide de documentations en précisant les grandes lignes et les notions basiques pour pouvoir réaliser la troisième étape qui introduit une application décisionnelle basé sur les arbres de décision, elle détermine une décision pour les plantes appropriées pour chaque ferme.

Pour conclure, nous espérons obtenir des capteurs sans fil qui contribueront grandement au développement de notre application et la rendre plus intelligente et plus précise, ce qui pourrait nous permettre de mettre l'application à la disposition des personnes concernées à savoir les agriculteurs.

Bibliographie

1. José, Graziano Da Silva (sans date). Nourrir la planète de manière durable [article]. Disponible sur : <https://www.un.org/fr/chronicle/article/nourrir-la-planete-de-maniere-durable>.(Consulté le 02.09.2020).
2. L'Internet des Objets : 101, Publié par Sameh Ben Fredj <http://blog.xebia.fr/2015/12/02/linternet-des-objets-101/>.
3. Gilles Babinet et Al, Big data et objets connectés faire de la France un champion de la révolution numérique, article, Institut Montaigne, Avril 2015, 228 pages.
4. Open Edition Books, <http://books.openedition.org/editionsmsh/84>, 12/12/2015.
5. Orange, acteur de l'Internet des objets et du Big Data, article, Novembre 2015, 17 pages.
6. .M. Han and H. Zhang, "Business intelligence architecture based on internet of things " Journal of Theoretical & Applied Information Technology, vol. 50, no. 1, pp. 90-95, 2013.
7. P.-J. Benghozi, S. Bureau, F. Massit-Folléa, C. Waroquiers, and S. Davidson, L'internet des objets: quels enjeux pour l'Europe, Éd. de la Maison des sciences de l'homme éd., 2009, 66 p.
8. M. Weill and M. Souissi, "L'Internet des objets: concept ou réalité?," dans Réalités industrielles, ESKA, Éd., Les Annales des Mines. ESKA, 2010, pp. 90-96.
9. KARA Nadjah : " Conception d'un réseau de communication pour une maison intelligente en utilisant la technique d'Internet des Objets.", Université Abderrahmane
10. Mira de Bejaïa. Pour obtenir le diplômé de Master en Télécommunications, 2017.
11. <http://blog.octo.com/modeles-architectures-internet-des-objets/>, (consultée le 12 Janvier 2017).
12. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. Internet of things : A survey on enabling technologies, protocols, and applications. IEEE Communications Surveys & Tutorials, 17(4) :2347–2376, 2015.
13. P.J . Benghozi,S. Bureau,and F. Massit-Folea. L'Internet des objets. Quels Enjeux Pour les Européens ? Technical report, Orange Ecole Polytechnique et TELECOM Paris Tech,2008.
14. Y.ait mouhoub, F.Bouchebbah . Propotion d'un modèle de confiance pour l'internet des Objets, Université A/MIRA de Bejaia . 2015

15. R.ACHOUR, N.Makhloufi, Authentification dans l'internet des objets . Université A/MIRAdé Bejaia, 2017
16. <https://www.itu.int/rec/T-RREC-E.164-201011-1/fr>.
17. P. Benghazi, S. Bureau, and F. Massit-Folea. L'internet des objets. Quels enjeux Pour les européens ? 2008.
18. «Solution IoT,» [En ligne]. Available: <http://www.alpwise.com/>.
19. Jayaraman, Prem, et al. "Internet of things platform for smart farming: Experiences and lessons learnt." Sensors 16.11 (2016): 1884.
20. Philippe Zwaenepoel & Jean-Michel Le Bars (1977), L'agriculture de précision ;Ingénieries EAT no 12, décembre 1997, p. 67 à 79.
21. http://www.omafra.gov.on.ca/french/livestock/vet/disease_pre.html.
22. J. Lieber, Fouille de données : notes de cours, 2007.
23. 20Référence BRAHIMI Belgacem ; Extraction de connaissances à partir de données incomplètes et imprécises ; 2011
24. L. Chaabane,« fusion et fouille de donneesguidees par les connaissances : application a l'analyse d'image »,Doctorat,Universite Mohamedkhider – biskra,2013.
25. The Gartner Group, www.gartner.com.
26. M. Kantardzic, "Data Mining–Concepts, Models, Methods,and Algorithms,"IEEE Press, Piscataway, NJ, USA, 2003.
27. P. CABENA, P. HADJINIAN, R. STADLER, J. VERHEES et A. ZANASI, Discovering Data Mining: From Concept to Implementation, Prentice Hall, Upper Saddle River, NJ, 1998.
28. E-G. TALBI, Fouille de données (Data Mining) : Un tour d'horizon, Laboratoire d'Informatique Fondamentale de Lille.
29. B. Fatiha, « DATA MINING DISTRIBUE», Magister, Université Mohamed Boudiaf (USTO-MB) Oran
30. M. J. BERRY, G. S. LINOFF, Data Mining Techniques For Marketing, Sales, and Customer Relationship, Management, Second Edition, 2004.
31. M. J. BERRY, G. S. LINOFF, Mastering Data Mining: The Art and Science
32. D.T. LAROSE, Discovering Knowledge In Data: An Introduction to Data Mining, Central Connecticut State University, 2005.
33. B. HOUMADI, « Etude exploratoire d'outils pour le data mining », Master ,l'Université du Québec aTrois-Rivières,2007.

34. B. Fatiha, « DATA MINING DISTRIBUE», Magister, Université Mohamed Boudiaf (USTO-MB) Oran
35. R. ZAIANE, Principles of Knowledge Discovery in Databases CMPUT690, University of Alberta, 1999.
36. Ph. PREUX, Fouille de données : Notes de cours, Université de Lille 3, 9 octobre 2008.
37. G. DONG, J. PEI, Sequence Data Mining, Springer Edition, 2007.
38. S. PRABHU, N. VENKATESAN, Data Mining and Warehousing, New Age International (P) Ltd., Publishers, New Delhi, 2007.
39. O. R. ZAIANE, Principles of Knowledge Discovery in Databases CMPUT690, University of Alberta, 1999.