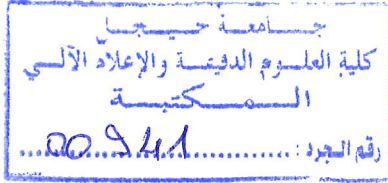


République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Inf SIAD 07/18

Université Mohamed Sadik BENYAHIA-JIJEL

Faculté des Sciences Exactes et Informatique

Département d'informatique



Mémoire

01
02

De fin d'étude pour l'obtention

du diplôme : Master de Recherche en Informatique

Option : Système d'information & Aide à la Décision

Thème

**ANALYSE PREDICTIVE GUIDEE
PAR UNE ONTOLOGIE**

Encadré par :

Mr. YAHIAOUI Abdelbaki

Réalisé par :

Melle. BOUREZAK Fatma Zohra

Melle. SAIDI Khadidja

Promotion 2017 / 2018

Remerciements

*Nous tenons tout d'abord à remercier Dieu
le tout puissant et miséricordieux, qui nous a donné la
force et la patience d'accomplir ce Modeste
travail de fin d'étude.*

*A notre encadreur de mémoire
«Mr. YAHIAOUI Abdel Bakj »
qui nous à donné l'opportunité de travailler sur ce
projet, pour son grand soutien scientifique et moral,
pour les conseils, les suggestions et les encouragements
qu'il nous a apportés pendant l'élaboration de ce mémoire.*

*Aux membres de jury qui ont acceptés l'évaluation
de ce travail.*

*Nous remercions également tous les
enseignants qu'ils nous accompagnés
durant nos études supérieures à
l'université de Jijel.*

*Enfin nous remercions tous ceux
qui, de près ou de loin,
nous ont aidés à la réalisation de ce travail.*

Un Grand MERCI à Tous

Dédicace

*A mes très chers parents Messaoud
et Salima
qui m'ont beaucoup encouragé*

*Et qui sont ma source de
patience que dieu me les gardes*

*A ma chère sœur Meriem et
mes chers frères Tariq et Nedjmeddine*

A mon oncle Boutadjine Karime

A mon fiancé Serhane adbelKadir

A toute ma famille Bourezak,

A tous mes chères amies

Et bien sûr à mon binôme Khadidja

*A toutes les personnes qui m'ont aidé
durant toute ma vie scolaire*

A tous mes collègues de promotion Juin 2018

A tous ce que j'aime je dédie ce travail.

Fatma zohra

Dédicace

*A mes très chers parents Smail
et Zoulikha
qui m'ont beaucoup encouragé*

*Et qui sont ma source de
patience que dieu me les gardes*

*A mes chères sœurs Manal et
Radia Et mes chers frères Soufiane
et Messaoud*

A mon fiancé Boudraa Badis

A toute ma famille Saidi

A tous mes chères amies

Et bien sûr à mon binôme Fatma Zohra

*A toutes les personnes qui m'ont aidé
durant toute ma vie scolaire*

A tous mes collègues de promotion Juin 2018

A tous ce que j'aime je dédie ce travail.

Khadija

لا يعار.
Exclus du Prêt.



Table des matières

Table des matières	i
Liste des tableaux	iv
Table des figures	vi
Liste des abréviations	vii
Introduction générale	1
1 la Fouille de Données et Domaine de Santé Publique	4
1.1 Introduction	4
1.2 La fouille de données	4
1.2.1 L'origine de FD	4
1.2.2 Définition	4
1.2.3 historique	5
1.3 le processus de data mining	5
1.3.1 définition	5
1.4 les objectifs de data mining	6
1.5 Les techniques de fouille de données	6
1.5.1 Apprentissage non supervisé	6
1.5.2 Apprentissage supervisé	7
1.6 Les tâches de fouille de données	9
1.7 Catégorisation des systèmes du data mining	10
1.8 Les domaines d'application du data mining	10
1.9 Les problèmes liés à la fouille de données	11
1.10 les outils de fouille de données	11
1.11 Domaine santé publique	11
1.11.1 Les objectifs de la santé public	12
1.12 Le domaine pharmaceutique	12
1.13 Caractéristiques et spécificités des données biomédicales	13
1.14 Les travaux réalisés	13
1.15 Le data mining dans la médecine et La pharmacie	14

1.16 Conclusion	14
2 la prévision	15
2.1 Introduction	15
2.2 Définition	15
2.3 Les caractéristiques des prévisions	15
2.4 Les types de prédiction	16
2.5 Les périodes en avant à prévoir	16
2.6 Les étapes du processus de prévision	16
2.7 Les techniques utilisées pour l'AP	17
2.7.1 Méthodes quantitatives	17
2.7.2 Méthodes qualitatives	22
2.8 Conclusion	22
3 les ontologies	23
3.1 Introduction	23
3.2 Origine de notion d'ontologie	23
3.2.1 Définition	24
3.3 Pourquoi développer une ontologie	24
3.4 Domaines d'utilisation	25
3.5 Les composants d'une ontologie	25
3.6 Type d'ontologie	26
3.7 Les étapes de construction des ontologies	28
3.8 Le graphe d'ontologie	30
3.9 Cycle de vie	31
3.10 Les problèmes de construction des ontologies	32
3.11 Les critères d'évaluation des ontologies	32
3.12 Niveau de formalisme de représentation	33
3.13 Outils de développement d'ontologies	33
3.13.1 Les outils dépendants du formalisme de représentation	33
3.14 Les outils indépendants de formalisme de représentation	34
3.15 Langages pour les ontologies	35
3.16 Conclusion	38
4 Conception du système	39
4.1 Introduction	39
4.2 Présentation du domaine	39
4.2.1 Contact avec le champ d'étude	40
4.2.2 Préparation de données	40
4.3 Les produits pharmaceutiques	42
4.3.1 Définition	42

4.3.2	Classification des produits pharmaceutiques	42
4.4	Les données après traitement	43
4.5	Analyse des besoins des utilisateurs	43
4.6	Conception de l'ontologie	44
4.6.1	Présentation d'ontologie	44
4.7	Classification des médicaments	45
4.8	Méthode d'exploitation pour la prédiction	50
4.8.1	Définition du problème	50
4.8.2	La collecte des données	50
4.8.3	Analyse préliminaire	50
4.8.4	Choix de la méthode	52
4.9	Conclusion	53
5	Implémentation et réalisation	54
5.1	Introduction	54
5.2	Les outils d'implémentations	54
5.3	Le schéma général du système	54
5.4	Réalisation de notre ontologie	55
5.4.1	Pour l'édition de l'ontologie	55
5.4.2	Création d'une ontologie	55
5.4.3	Création des classes	56
5.4.4	Création des relations	58
5.4.5	Création des Propriétés	59
5.4.6	Saisie des individus (Instances)	60
5.5	Présentation des outils utilisés pour l'interface et le reporting	61
5.5.1	Développement des interfaces	61
5.5.2	Exécution de l'application	62
5.6	La prévision :	66
5.7	Discussion des résultats de prevision	67
5.8	Conclusion	68
	Conclusion générale	69
	Bibliographie	71

Liste des tableaux

4.1 Classification des Medicaments.	45
--	----

Table des figures

1.1	les principales taches de fouille de données.	9
2.1	Représentation de la tendance.	19
2.2	Représentation des cycles.	19
2.3	Les différents lissages exponentiels.	20
3.1	les domaines d'utilisation des ontologies.	25
3.2	exemple d'une ontologie informatique.	28
3.3	exemple d'un graphe d'ontologie.	31
3.4	cycle de vie d'une ontologie.	32
3.5	exemple d'un graphe RDF.	36
3.6	la relation entre les 3 sous-langages.	38
4.1	format initiale de données	40
4.2	les données au format csv.	41
4.3	Les données au format xlsx	41
4.4	Les données collectées	43
4.5	Une partie d'ontologie	45
4.6	la consommation des produit dentaire entre les 2009 et 2013	50
4.7	résultats de la méthode SMO reg	51
4.8	résultats d'analyse par la méthode Gaussian Processes	51
4.9	résultats d'analyse par la méthode Linéaire Régression	52
4.10	résultats d'analyse par la méthode Multilayer Perceptron	52
5.1	Le schéma général du système	55
5.2	Création d'une ontologie	56
5.3	Création des classes	57
5.4	Mettre les classes "disjoint" les unes des autres	57
5.5	Création des relations	58
5.6	Création de nouveau «objectproperty » est_un	59
5.7	ajouter «DataProperties » à notre ontologie	60
5.8	insertion des individus	60
5.9	association des individus à la classe à laquelle il appartient	61

5.10	graphe ontologie	61
5.11	Fenêtre d'Authentification	62
5.12	Fenêtre du Menu Principal	63
5.13	le graphe ontologie	63
5.14	consommation en 2017 par mois/classe	64
5.15	consommation en 2017 par mois/produit	64
5.16	Les produits composite auto	65
5.17	consommation du produits de sous class (Consommable_D) entre trimestre 1(2010) et trimestre 2 (2014).	66
5.18	la méthode Linéaire Régression	66
5.19	Smo Reg.	67
5.20	Représentation des résultats sous forme histogramme et diagramme de secteur	67

Liste des abréviations

- FD** fouille de données
- EPSP** établissement public de santé de proximité
- ODE** Ontology Design Environment
- OntoEdit** Ontology Editor
- RDF** Resource Description Framework
- OWL** Web Ontology Language
- XML** Extensible Markup Language
- URI** Uniform Resource Identifier
- SVM** Support vector machines
- SMO** Séquentiel minimal optimisation
- AP** Analyse prédictive

Introduction générale

Un organisme crée de la valeur en traitant de l'information, en particulier dans le cas des sociétés, des entreprises et des organismes de service. Ainsi, l'information possède une valeur d'autant plus grande qu'elle contribue à atteindre des objectifs de l'organisation.

La mise en place d'un système décisionnel permet d'apporter des réponses efficaces à tous les niveaux de l'entreprise, cet aspect décisionnel est présent dans les organisations depuis de nombreuses années, il revêt l'apparence de rapports. Mais beaucoup d'entre elles s'aperçoivent que ces simples outils de reporting ne satisfont pas entièrement leurs attentes. Elles se rendent compte que la mise en place et l'introduction et la mise en place des techniques de fouille de données est nécessaire pour une meilleure exploitation des données existantes.

Le domaine de la santé est un domaine très riche en données. Si nous prenons le cas de la pharmacie d'un établissement de santé, ces données font référence aux produits pharmaceutiques, les médicaments par exemple, qui sont acquis et consommés au fil du temps, reflétant ainsi leurs disponibilité, leurs coût, leurs péremption, etc.

Les analystes et les décideurs à vouloir analyser ces données pour faire des meilleures décisions à partir d'utilisation des outils d'analyse comme l'extraction des connaissances ou des fouille de données précisément qui permettent d'explorer de grande quantité de données, l'extraction de données se base sur les techniques des fouille de données tel que la classification qui permet de construire des model pour le but de classer un élément dans un classe prédéfinie a fin de prédire le future.

Aujourd'hui il est tendance dans les établissements pharmaceutiques de chercher à prédire les besoins futurs en médicaments et produits pharmaceutique en se basant l'historique de consommation de ces derniers.

Problématique.

La pharmacie d'un établissement de santé comme le cas de l'EPSP (Etablissement Public de Santé de Proximité) de Jijel qui représente notre domaine d'étude, est un domaine très riche en données qu'il faut exploiter pleinement pour une meilleure prise de décision.

La prévision des besoins en produits pharmaceutiques est une préoccupation majeure du gestionnaire d'un établissement de santé afin d'assurer d'une part la disponibilité permanente de ces produits et d'éviter d'autre part leurs péremptions.

En effet, avant tout achat, la première étape consiste à évaluer le plus précisément possible, aussi bien qualitativement que quantitativement, les besoins en produits pharmaceutiques. L'évaluation des besoins sert à assurer à chaque structure de santé ou médicale un approvi-

sionnement approprié en médicaments pour leur permettre de traiter le nombre de cas prévu et, de façon générale, de promouvoir et de maintenir une utilisation rationnelle et économique des médicaments.

Cette étape d'estimation des besoins en produits pharmaceutiques est d'autant plus importante que les budgets sont généralement limités.

Objectif

Le but principal de notre projet qui a comme domaine d'étude l'établissement de santé de proximité (EPSE) de Jijel est de concevoir et d'implémenter un système de prédiction des besoins en produits pharmaceutiques guidé par une ontologie. Ce système vise à prêter aide aux décideurs et analystes de l'établissement afin de remédier aux problèmes d'approvisionnement insuffisant ou excessif de ces produits.

Pour atteindre nos objectifs, nous contribuons par les éléments suivants :

- ❖ **L'étude du terrain** : qui consiste premièrement à l'étude du domaine, deuxièmement à la collecte des données d'acquisition et des consommations de toutes les années précédentes, partir de l'année 2009 et troisièmement à l'analyse des besoins des utilisateurs.
- ❖ **Structuration des données** : utiliser les ontologies pour faire la préparation et classification des produits pharmaceutiques pour avoir une meilleure organisation de donnés.
- ❖ **Analyser les données** :choisir la technique de prévision le plus adéquate.

Mais comment ?

Pour réaliser notre travail nous allons exploiter et utiliser les outils et les environnements suivants :

- ❖ **Java** :comme langage de programmation.
- ❖ **NetBeans** :comme environnement de développement.
- ❖ **Protéger** :comme un éditeur d'ontologie permet de manipuler des formats très divers.
- ❖ **Weka** :L'espace de travail Weka contient une collection d'outils de visualisation et d'algorithmes pour l'analyse prédictive des données .

Notre mémoire est organisé comme suit :

Dans chapitre 1, nous présentons une introduction générale sur le domaine de la santé publique, la définition de feuille de données aussi les processus et les déferents techniques de fouille de données.

Dans chapitre 2, nous présentons la définition de la prévision, ces caractéristiques ainsi que les étapes et les différentes méthodes de prévision.

Dans chapitre 3, présente les notions fondamentales d'ontologie, et comment elle est utilisée pour la classification des médicaments.

Dans chapitre 4, concerne la conception de notre système, il est consacré l'ensemble des traitements sur les données après l'extraction ainsi qu'au choix des techniques de prévision retenues.

Dans chapitre 5, présente l'implémentation de notre système et les différents outils utilisés pour la réalisation de notre application ensuite nous y présentons le résultat de prédiction obtenu pour montrer l'utilité de notre projet.

la Fouille de Données et Domaine de Santé Publique

1.1 Introduction

Actuellement Les établissements de santé traitent une grande quantité de données de son environnement chaque année, pour rester toujours concurrente. Ces données proviennent de multiple source. Par conséquent ces institutions ont besoin d'un outil approprié d'extraction des données et des connaissances, tel que le data mining, ce qui a facilité la gestion de toutes données avec un gain de temps appréciable.

Dans ce chapitre nous présentons les éléments de fouille de données aussi le domaine de la santé public.

1.2 La fouille de données

1.2.1 L'origine de FD

La Fouille de Données est à l'origine de trois types de méthodes qui sont toutes intégrées dans de nombreux systèmes commerciaux :

- ❖ La recherche des associations, en particulier la découverte des théorèmes incertains confirmés par les données et les multiples mesures d'intérêt qui sont associées au choix des associations.
 - ❖ La découverte de séries temporelles qui correspond à une extension de la première méthode, et qui représente l'un des plus grands succès de la Fouille de Données.
- En fin, et ce sous l'influence du monde industriel, la Fouille de Données a développé de multiples méthodes de nettoyage et de segmentation des données.[8]

1.2.2 Définition

Il existe plusieurs définitions de la fouille de données, nous citons les suivantes :

- ✓ Selon Wook (2009), cité dans [1] , « data mining est le processus d'extraction automatique d'information et des relations à partir d'immenses quantités des données»

- ✓ Selon fayyad en 1996, cité dans [1] , «l'extraction de connaissances à partir des données et un processus non trivial d'identification de structures inconnues ,valide et potentiellement exploitable dans les base de données »
- ✓ La fouille de données Le data mining est l'ensemble des méthodes scientifiques destinées à l'exploration et l'analyse de(souvent) grandes bases de données informatiques en vue de détecter dans ces données des profils-type, des comportements récurrents, des règles, des liens, des tendances inconnues (non fixées a priori), des structures particulières restituant de façon concise l'essentiel de l'information utile pour l'aide à la décision.[3]

1.2.3 historique

L'expression data mining est apparue vers le début des années 1960, employée par les statisticiens qui ont commencé à traiter les tableaux de données relatifs à des enquêtes ou des expériences. Les statisticiens considéraient toutefois cette démarche comme peu scientifique et utilisèrent alors les termes data mining ou data fishing.[2]

Cette expression réapparaît dans les années 80. Des chercheurs, tels que Rakesh Agrawal, ont commencé à travailler sur l'exploitation du contenu des bases de Données volumineuses pour la recherche de règles d'association. Ils utilisèrent l'expression data base mining.

En 1989, Shapiro Piatetski propose le terme knowledge discovery à l'occasion d'une conférence sur la découverte des connaissances dans les bases de données.

En 1991, le concept du data mining apparait pour la première fois aux Etats Unis comme une nouvelle discipline à l'interface de la statistique et des technologies de l'information.

En 1995, la communauté de data mining a initié sa première conférence.

En 1997 a été lancé le premier journal spécialisé « data mining and knowledge discovery journal » publiée par Kluwers.[2]

1.3 le processus de data mining

1.3.1 définition

Le processus de data mining a été développé en 1996 pour répondre aux besoin des projets industriels, il constitué six étapes :

- ❖ **Compréhension du business** :cette phase porte sur la compréhension des objectifs, la définition des problématiques et le plan préliminaire pour accomplir ces objectifs .
- ❖ **Compréhension des données** :cette phase commence par une collection de données, identification les problèmes de qualité des données, détection des sous-ensembles pour

construire des hypothèses pour l'information cachée.

- ❖ **Préparation des données** : cette phase contient toutes les activités nécessaires pour construire une base de données finale.
- ❖ **Modélisation** : dans cette étape sont appliquées plusieurs techniques sur les données et leur paramètre.
- ❖ **Evaluation** : dans cette étape les modèles sont évalués aussi les étapes suivies pour la construction des modèles sont réévaluées pour assurer que le projet respecte les objectifs.
- ❖ **Déploiement** : La création du modèle ne représente pas la fin du projet, même si le but initial du projet est d'augmenter les connaissances de données. Ces connaissances acquises ont besoin d'être organisées et présentées d'une manière utilisable par le client. [10]

1.4 les objectifs de data mining

- ❖ es fouille de données permet de développer des systèmes et des techniques efficaces et extensibles.
- ❖ Représentation simple des connaissances.
- ❖ Visualisation sous forme exploitables.
- ❖ Limiter l'intervention des utilisateurs.

1.5 Les techniques de fouille de données

Pour effectuer les tâches du Data Mining il existe plusieurs techniques issues de disciplines scientifiques diverses (statistiques, intelligence artificielle, base de données) afin de construire des modèles à partir de ces données. Dans ce chapitre, nous présentons les techniques du data mining les plus connues.

1.5.1 Apprentissage non supervisé

L'objectif de la fouille de données non supervisée est donc de trouver des relations entre caractéristiques (variables exogènes) suffisamment significatives et permettant d'augmenter les connaissances du domaine étudié. Plusieurs techniques peuvent être, on en cite :

- ❖ **les règles d'association** :

Ces méthodes sont intuitivement faciles à interpréter car elles montrent comment des produits ou des services se situent les uns par rapport aux autres. Ces règles sont particulièrement utiles en marketing.

- Exemple sur les règles d'association** :

- ✓ Si un client achète des plantes alors il achète du terreau.
- ✓ Si un client achète une télévision, il achètera un magnétoscope dans un an.

-Avantages de la méthode :

- ✓ Méthode non supervisée : à l'exception de la classification de différents articles en produits.
- ✓ Clarté des résultats : les règles sont faciles à interpréter.
- ✓ Traitement 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.

✧ **Le clustering (segmentation)**

Le clustering est l'opération qui consiste à regrouper les individus d'une population en un nombre limité de groupes.

-Propriété de segmentation :

- ✓ Ils ne sont pas prédéfinis, mais découverts automatiquement au cours de l'opération, contrairement aux classes de la classification.
- ✓ Les segments regroupent les individus ayant des caractéristiques similaires et séparent les individus ayant des caractéristiques différentes.

1.5.2 Apprentissage supervisé

L'objectif de la fouille de données supervisée est de trouver des relations entre les variables exogènes permettant d'expliquer et/ou de prévoir le comportement de la variable endogène.

Plusieurs techniques peuvent être inscrites : on en cite :

✧ **Les arbres de décision**

Les arbres de décision sont des outils puissants et connus, tant pour la classification que pour la prédiction les arbres de décision permettent de distinguer différentes classes et de leur associer une ou plusieurs règles.

-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 noeud, 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 variable pertinentes** : l'arbre contient les attributs utiles pour la classification.

-Inconvénients :

- ✓ Manque d'évolutivité dans le temps.
- ✓ Méthode sensible au nombre de classes .

✧ Les réseaux de neurones :

Sont des outils utiliser pour les classifications et les prédictions, il représentent une transposition simplifiée des neurones des cerveau humain.

- 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.

- 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.[4]

1.6 Les tâches de fouille de données

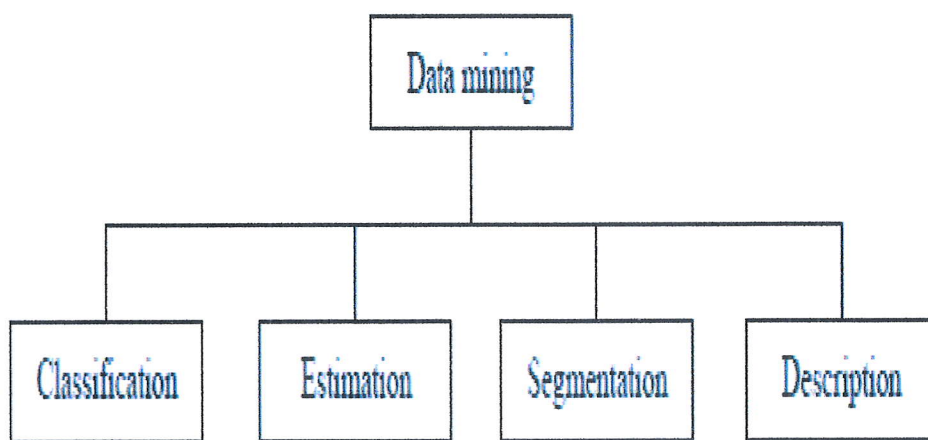


FIGURE 1.1 – les principales tâches de fouille de données.

- ✓ **La classification** : Elle consiste à examiner les caractéristiques d'un objet et lui attribuer une classe, la classe est un champ particulier à valeurs discrètes.
Des exemples de tâche de classification sont :
 - attribuer ou non un prêt à un client.
 - établir un diagnostic.
 - accepter ou refuser un retrait dans un distributeur.
 - attribuer un sujet principal à un article de presse.
- ✓ **L'estimation** : Contrairement à la classification, le résultats d'une estimation permet de procéder aux classification grâce à barème.[5]
- ✓ **La description** : C'est souvent l'une des premières tâche demandées à un outil de data mining .On lui demande de décrire les données d'une base complexe, cela engendre souvent une exploitation supplémentaire en vue de fournir des explications.[5]
- ✓ **La segmentation (clustering)** : La segmentation consiste à segmenter une population hétérogène en sous populations homogène, contrairement à la classification, les sous

population ne sont pas préétablies.[5]

1.7 Catégorisation des systèmes du data mining

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

- ✓ **Classification selon le type de données à explorés** : 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.
- ✓ **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 datawarehouses, les bases de données transactionnelles, etc.
- ✓ **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.
- ✓ **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.

1.8 Les domaines d'application du data mining

Les fouille de données utilisent dans un grand nombre de domaines et d'activités. Nous présentons le plus connus :

- ◆ Analyse financière.
- ◆ Médecine et pharmacie.
- ◆ Gestion et analyse de risque.
- ◆ Marketing direct.
- ◆ Gestion et analyse des marchés.
- ◆ Détection des fraudes.
- ◆ Gestion de stocks.

1.9 Les problèmes liés à la fouille de données

Les données sont collectées dans le but de répondre à une question posée par le métier. Un risque de l'exploration de données est l'utilisation de ces données dans un autre but que celui assigné au départ. Le détournement des données est l'équivalent d'une citation hors de son contexte. En outre, elle peut conduire à des problèmes éthiques.

La qualité des données, c'est-à-dire la pertinence et la complétude des données, est une nécessité pour l'exploration des données, mais ça ne suffit pas. Les erreurs de saisies, les enregistrements doublonnés affectent aussi la qualité des données. [6]

1.10 les outils de fouille de données

Il y a beaucoup d'outils généraux d'exploitation de données qui fournissent des algorithmes d'exploitation, quelques exemples d'outil commercial et scolaire sont tanagra, weka, dbminer, clementine.

- ✧ **Tanagra** : Est un logiciel de data mining gratuit pour l'enseignement et la recherche il implémente une série de méthodes de fouilles de données issues du domaine de la statistique exploratoire, de l'analyse de données, de l'apprentissage automatique et des bases de données.
Tanagra est un projet ouvert au sens qu'il est possible à tout chercheur d'accéder au code et d'ajouter ses propres algorithmes pour peu qu'il respecte la licence de distribution du logiciel. [9]
- ✧ **Weka** : Est une collection d'algorithmes d'apprentissage automatique pour des tâches d'exploitation de données. Les algorithmes peuvent être appliqués directement à un ensemble de données ou s'appeler de vos propres codes java.
- ✧ **Clementine** : avec clémentine il est facile de découvrir des perspicacités dans vos données. Son interface graphique simple met la puissance de l'exploitation de données dans les mains des clients de l'entreprise et augmente la productivité d'analyste.
- ✧ **DBMiner** : DBMiner est spécialisé dans le domaine économique. Il utilise des processus intelligents et automatisés pour analyser de grands volumes de données détaillées des bases de données relationnelles, des entrepôts de données et des données de Web avec la facilité d'utilisation exceptionnelle et la souplesse élevée. En encapsulant des techniques de data mining, vous pouvez présenter vos produits ou services aux clients.

1.11 Domaine santé publique

Les établissements de santé (EPSP) sont très riches en données qu'il faut exploiter pour voir une meilleure décision.

1.11.1 Les objectifs de la santé public

La santé publique peut être définie par ses objectifs qui sont :

- Réduire l'apparition de la maladie.
- Maintenir la population en santé.
- Améliorer et augmenter la qualité de la vie en assurant des conditions des lesquelles les gens peuvent être bonne santé.

Ceci implique la réalisation d'un certain nombre de fonction de base de santé publique tel que :

- **Prévention, surveillance et contrôle de maladie** : prévention de risque, surveillance des maladies, etc.
- **Surveillance de l'état de santé** : évaluation des besoins et des risques de population afin de déterminer quels sous-groupes ont besoin de service, évaluation de l'efficacité de fonctions de la santé public, etc.
- **Promotion de la santé** : développement de compétence de vie, établissement et maintien des liens avec les décideurs.
- **Protection de l'environnement** : prise en compte des enjeux environnementaux dans les politiques de développement tel que la lutte contre les pollutions.
- **Législation et réglementation en santé publique** : établissement de législation de santé aussi des règlements et des procédures administratives .
- **Planification et gestion de santé publique** : amélioration de qualité des services de santé.[7]

1.12 Le domaine pharmaceutique

le domaine pharmaceutique représente un enjeu très important il est au centre de plusieurs disciplines qui on peut classer sous quatre aspect principaux :

- ✓ **Aspect de santé publique** : On y trouve des organismes gouvernementaux et non gouvernementaux s'intéressant à l'amélioration de la santé publique tel que les ministères de la santé.
- ✓ **Aspect commercial** : Il concerne les producteurs ou fabricant de produit pharmaceutique revendeurs de ces produit affectant les stratégies de marketing, gestion de stocks, réalisation de profits financiers, etc.
- ✓ **Aspect scientifique** : Il concerne la communauté des chercheurs en pharmacologie tel que les recherches moléculaires.
- ✓ **Aspect d'intérêt et d'économie national** : Comme les ministères de l'économie et la sécurité sociale pour assure la sécurité des personnes, l'équilibre budgétaire.

[7]

1.13 Caractéristiques et spécificités des données biomédicales

En médecine, plus que d'autres disciplines, les observations sur lesquelles se basent les décisions médicales sont très imparfaites. Elles peuvent être floues, ambiguës, incomplètes, Incertaines. Le processus décisionnel, qu'il soit diagnostique, thérapeutique ou pronostique est donc un processus sous incertitude. Plusieurs éléments participent à cette situation :

- Les observations sont floues parce qu'il n'y a pas de vocabulaire standardisé communément utilisé par la communauté médicale et répondant à des définitions clairement exprimées.
- Elles sont ambiguës car le malade peut exprimer une plainte et le médecin en attendre une autre. De plus, des observations différentes ont des seuils de perception différents dans l'appréciation de la présence ou l'absence d'un signe.
- Les observations sont incomplètes car, en situation de décision, le médecin doit agir sans connaître l'ensemble de données relatives à un patient et bien entendu toute la connaissance spécifique de la situation. Ceci est particulièrement vrai dans des situations d'urgences, où un patient peut se retrouver inconscient, lorsque les résultats de laboratoire ne sont pas encore disponibles ou lorsqu'une technique particulière fournissant des informations utiles pour le diagnostic n'est pas pratiquée dans la structure de soins.
- Elles sont incertaines car les connaissances cliniques sont l'expression d'observations statistiques sur des échantillons de patients présentant des maladies plus ou moins fréquentes, ayant des formes cliniques différentes et ne s'exprimant pas toujours par la même symptomatologie, partageant certains signes avec d'autres maladies ou présent an des réponses variables à un traitement donné.
- La reproductibilité des observations médicales est fonction des méthodes de mesure (Variabilité analytique), des observations (variabilité intra et inter-observateurs et du sujet observé (variabilité intra et inter-individuelle).[4]

1.14 Les travaux réalisés

L'utilisation des méthodes de fouille de données dans le au domaine biomédical est en forte croissance.

Parmi les travaux de recherche il y'a : Fouille de données biomédicales basées sur une ontologie. L'application des techniques du data mining au domaine biomédical est une aventure Certes enrichissantes, mais très difficile. Cette difficulté est non seulement liée aux grandes bases de données mais aussi principalement aux problèmes cités ci-dessous :

- ✓ Le premier problème concerne la spécificité des données médicales : il s'agit de données particulières, complexes, hétérogènes, hiérarchiques souvent imprécises, subjectives, bruitées, entachées de valeurs manquantes, ou incomplètes.

- ✓ problème concerne la nécessité d'une validation par les experts de l'ensemble des traitements effectués sur des données médicales.[4]

1.15 Le data mining dans la médecine et La pharmacie

Des exemples de l'usage médicaux et pharmaceutiques des techniques de Data Mining pour l'analyse de bases de données médicales :

- Prédiction de présence de maladies et/ou de complications.
- Le choix d'un traitement pour le cancer.
- Le choix d'une technique particulière (de sutures, matériel de suture, etc.) dans une des procédures chirurgicales.
- Approvisionnement des médicaments les plus fréquemment prescrits.[5]

1.16 Conclusion

Dans ce premier chapitre, nous avons présenté les principaux concepts de fouille de données, les processus, les tâches et les méthodes les plus utilisés en data mining ainsi que les avantages et les inconvénients de chaque méthode, aussi les outils de fouille de données tel que Weka que nous allons utiliser dans notre application. Nous avons donné un aperçu global sur le domaine de la santé publique et le domaine pharmaceutique tel que les objectifs de la santé public et les différentes caractéristiques des données biomédicale.

la prévision

2.1 Introduction

L'analyse prédictive (AP) est la technologie qui permet de prédire les futures des comportements des individus afin d'aider à la prise de décision, elle recouvre un ensemble de méthodes très diverses qui ont en commun de chercher à réduire l'incertitude liée à la non connaissance du future. Elle est définie aussi comme technique d'exploration, d'extraction et d'analyse de données historiques pour prédire les évènements futurs. La décision est difficile car elle risque d'être coûteuse si les prévisions sur lesquelles elle est fondée sont infirmées.

2.2 Définition

Anticiper, prévoir les fluctuations de l'environnement interne et externe afin d'optimiser la stratégie de l'entreprise pour les années à venir.

La prédiction décisionnelle permet aux entreprises à se projeter dans l'avenir pour prendre des meilleures décisions. [15]

2.3 Les caractéristiques des prévisions

- ✧ **Elles sont généralement fausses** : Une fois les prévisions déterminées, elles sont souvent utilisées comme s'il s'agissait d'informations certaines. Le système de planification doit être suffisamment robuste pour réagir en cas d'erreur de prévision non anticipée.
- ✧ **Une bonne prévision est plus qu'une valeur numérique** : les prévisions sont généralement fausses, une bonne prévision doit également inclure une mesure de l'erreur anticipée pour la prévision. Ceci peut prendre la forme d'une étendue ou d'une mesure de l'erreur telle que la variance de la distribution de l'erreur de prévision.
- ✧ **Les prévisions à long terme sont moins précises** : Cette caractéristique est assez intuitive. On peut prévoir plus exactement les consommations de la prochaine semaine que celles à survenir dans 12 mois.

- ❖ **Les prévisions ne remplacent pas la vraie information** : Une technique de prévision donnée peut donner de bonnes prévisions dans la plupart des cas. Cependant, il peut y avoir une information disponible concernant la demande future qu'on ne peut pas anticiper simplement en regardant les données historiques.

[17]

2.4 Les types de prédiction

Dans la prévision il existe deux types d'opération selon la variable de sortie :

- ✓ Si la variable de sortie est de type discret, la classification aura pour rôle de construire le modèle qui permettra de classer correctement les enregistrements, c'est-à-dire d'assigner des catégories prédéfinies aux données.
- ✓ Si la variable de sortie est de type continu, l'estimation consistera à compléter une Valeur manquante dans un champ particulier en fonction des autres champs de L'enregistrement. Les outils statistiques usuels de régression sont les plus employés. Les réseaux de neurones sont aussi souvent employés dans cette activité.[11]

2.5 Les périodes en avant à prévoir

- ✓ **Long terme** : peut aller jusqu'à cinq ans ou plus.
- ✓ **Moyen terme** : mesuré en semaines ou en mois, peut aller jusqu'à deux ans.
- ✓ **Court terme** : généralement mesuré en jours ou en semaines, peut aller jusqu'à un an.

2.6 Les étapes du processus de prévision

On peut identifier cinq étapes de base dans tout processus de prévision comportant des données quantitatives.

❖ Définition du problème

Cette étape consiste à développer une bonne compréhension de comment les prévisions seront utilisées, qui a besoin des prévisions et comment cette fonction cadre dans l'organisation. Il est habituellement utile de prendre le temps de discuter avec les personnes qui seront impliquées dans la collecte des données, le maintien des bases de données et l'utilisation des prévisions pour la planification.

La personne responsable des prévisions a beaucoup de travail à faire pour bien définir le problème avant que toute prévision soit faite.

❖ Recherche des données :

Il y a toujours au moins deux types d'information disponibles :

- les données statistique.
 - le jugement des experts et du personnel clé. Ces deux types d'information doivent être amassés.
- ❖ **Analyse préliminaire** :Consiste a construire un graphique à partir des données de manière à en avoir un aperçu visuel, des statistiques descriptives peuvent être calculées (moyennes, écart- type ,etc.). L'idée à cette étape est bien de connaitre les données qui doivent être traitées.
 - ❖ **Choix de la méthode de prévision** :Cette étape concerne le choix et l'ajustement des modèle de prévision potentiels qui ont été déterminés à l'étape d'analyse préliminaire.
 - ❖ **Utilisation et évaluation du modèle** :Une fois que le modèle a été sélectionné et que ses paramètres ont été définis, il est possible de l'utiliser pour faire des prévisions.[15]

2.7 Les techniques utilisées pour l'AP

Les prévisions sont la base de tous les système de gestion, une méthodes de prévision peut influencer le choix d'une politique de gestion de stock. La plupart des techniques utilisées pour l'analyse prédictive sont généralement basées sur l'apprentissage automatique des données existantes qui consiste à produire automatiquement des règles à partir des données traitées. Cela permet au différent système d'évoluer dans le temps.

Les méthodes prévisionnelles se divise en deux groupes

- ❖ **Méthodes qualitatives** :sont basées principalement sur le jugement humain, les données historiques, les prévisions sont généralement générées à partir des modèles mathématiques.Sont utilisées pour expliquer .
- ❖ **Méthodes quantitatives** :sont basées sur les variables qui la variation de la demande. Ces méthodes ne nécessitent que l'opinion subjective des personnes sans qu'il soit nécessaire d'exprimer la prévision à l'aide d'équations mathématiques.

2.7.1 Méthodes quantitatives

séries chronologiques(temporelles)

- ❖ **Définitions**
- ✓ Est une séquence de données ordonnées chronologiquement,il s'agit de prévoir les valeurs futures d'une variable grâce aux valeurs observées dans le présent et le passé de cette même variable.
- ✓ Série chronologique ou série temporelle ou chronique : Suite d'observations régulièrement espacées dans le temps.
- ❖ **Les objectifs des séries temporelles**

- Prédire les valeurs futures .
- Comprendre le passé.
- Etudier les liens avec d'autres séries.

[12]

❖ **Les principales étapes de traitement d'une série chronologique sont les suivantes :**

- correction des données.
- observation de la série.
- modélisation (avec un nombre fini de paramètres).
- analyse de la série à partir de ses composantes.
- diagnostic du modèle - ajustement au modèle.
- prédiction (prévision). [13]

❖ **Les composantes d'une série chronologique :**

L'objectif principal de la décomposition d'une série chronologique est de distinguer dans l'évolution de la série . [14]

- ✓ Une tendance générale.
- ✓ Des variations saisonnières qui se répètent chaque année.
- ✓ Des variations accidentelles imprévisibles.

L'intérêt de ceci est :

- De mieux comprendre.
- De mieux décrire l'évolution de la série
- De prévoir son évolution (à partir de la tendance et des variations saisonnières).

a. Tendance : La tendance est une évolution croissante ou décroissante du niveau de base de la demande en fonction du temps. Un historique de données couvrant plusieurs années est souvent nécessaire afin de déterminer les tendances. Le patron d'une tendance peut être représenté par :

- ❖ Une ligne droite pour une tendance linéaire .
- ❖ Une fonction quadratique ou à une courbe exponentielle par exemple pour une tendance non linéaire.

b. Cycle : Un phénomène est soumis à un effet cyclique lorsqu'il se répète après un certain intervalle de temps qui n'est pas forcément régulier. Par exemple, tous les cinq à sept ans.

c. Saisonnalité : Une série chronologique comporte une saisonnalité lorsqu'elle est soumise à des fluctuations qui se répètent périodiquement. Et qui sont reliés à un ou plusieurs facteurs environnementaux. Les ventes d'équipements d'hiver, par exemple, enregistrent des hausses durant les mois d'hiver. On se réfère également au terme de saisonnalité lorsque l'on observe des variations régulières lors de certains jours de la semaine. [7]

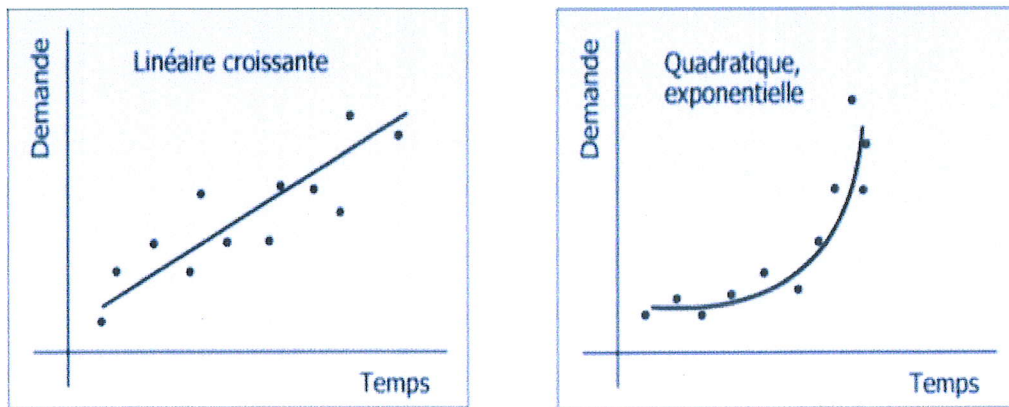


FIGURE 2.1 – Représentation de la tendance.



FIGURE 2.2 – Représentation des cycles.

❖ Les modèle pour série stationnaire :

❖ Moyenne mobile simple :

Ce modèle est l'un des plus simples. Dans Ce cas la demande prévue est simplement la moyenne des n dernières demandes[14].

Cette Méthode est particulièrement adaptée aux produits ayant une demande quasi constante avec une faible variance. Cette méthode permet également d'éliminer les variations erratiques Perturbant les séries. Les moyennes mobiles sont des moyennes mises à jour au fur et à mesure que de nouvelles observations sont disponibles. L'expression de cette Moyenne est la suivante :

$$SMA_{t+1} = \left(\frac{1}{n}\right) \sum_{i=t+1-n}^n D_i$$

SMA_{t+1} : Prévission de la demande qui avait été faite pour la période t.

D_i :Demande réelle à la période i.

n :Le nombre de période prise en compte. Ce nombre peut varier selon les cas d'études, il est

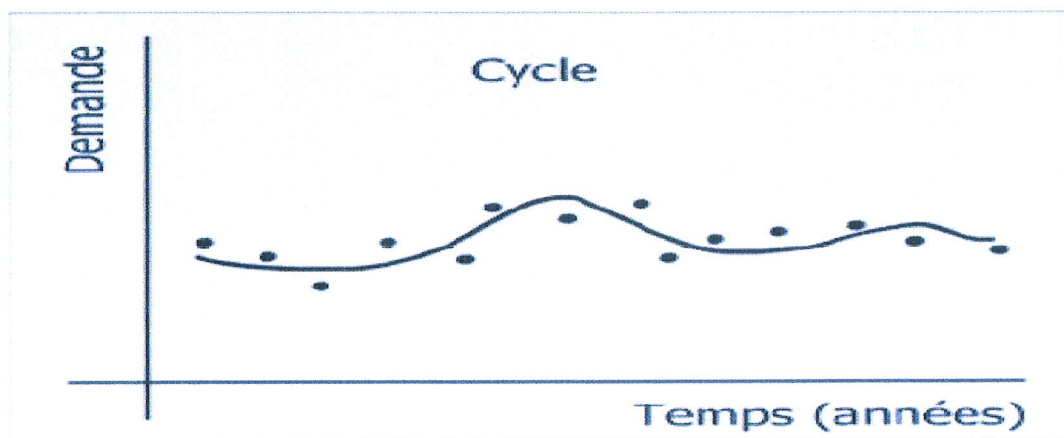


FIGURE 2.3 – Les différents lissages exponentiels.

difficile de donner une valeur standard. Selon la période de prévision ce nombre évoluera.

✧ **Moyenne mobile simple pondérée :**

Elle permet de donner différents poids pour les données utilisées dans le calcul de la moyenne. On peut de cette manière donner plus d'importance aux données plus récentes afin qu'elles influencent davantage la prévision que les données plus anciennes.

$$P_t = \frac{W_1(D_{t-1}) + \dots + W_n(D_{t-n})}{n}$$

P_t :prévision de la demande pour la période t. w_n :les poids . Note :la somme des poids utilisés doit égaler 1.

✧ **Lissage exponentiel simple :**

Le lissage exponentiel simple (LES) s'applique à des séries chronologiques sans saisonnalité et à tendance localement constante. Elle permet une pondération particulière des données passées. Cette méthode se différencie de la méthode de la moyenne mobile par la prise en compte de toutes les données du passé, et pas uniquement des n dernières valeurs. Le coefficient de pondération joue essentiellement le même rôle que n, le nombre de données utilisées dans le calcul de la moyenne mobile simple.

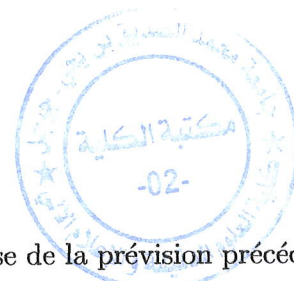
- ✓ Si n est grand, plus de poids est mis sur la plus récente donnée et moins de poids sur les données passées, ce qui résulte en une prévision qui réagit rapidement aux changements (la réaction du modèle à des variations de la demande sera plus rapide).
- ✓ Si n est petit, alors un poids similaire est mis sur les données prise en compte et la prévision est plus stable.

$$P_{t+1} = P_t + \alpha(D_t - P_t)$$

D_t :Demande réelle à la période t.

P_t :Prévision de la demande pour la période t.

α :constante de lissage : tel que α est supérieure de 0 et inférieure DE 1.



Cette formule fournit la prévision pour la période $t+1$ sur la base de la prévision précédente à laquelle on rajoute l'écart de prévision de la période t corrigé par un facteur de lissage α .

- ✓ Si α est égal à 1, la prévision de chaque période correspond simplement à la demande réelle de la période précédente.
- ✓ Si α égal à 0 correspond à une prévision stationnaire : la prévision correspond à la prévision de la période précédente qui elle-même correspond à la prévision précédente et ainsi de suite. Il faut à présent se poser la question de quel facteur il convient de choisir pour être le plus proche possible de la réalité. Ce choix peut résulter d'une décision subjective du gestionnaire qui se base sur son expérience du phénomène ou sur son intuition. Mais il est également possible de déterminer le coefficient optimal en choisissant celui qui minimise la moyenne des erreurs au carré.

✧ Lissage exponentiel doubles et triple :

Le lissage exponentiel double permet une modélisation de séries chronologiques avec tendances contrairement au lissage exponentiel simple qui nécessite des données stationnaires.

- ✓ **Modèle de BROWN** : (lissage exponentiel double à paramètre simple), du modèle du HOLT (lissage exponentiel double à paramètre double)
- ✓ Modèle de WINTERS ou HOLT/WINTERS (lissage exponentiel triple).

Les modèles de Brown et Holt sont les premières approches permettant une modélisation de série chronologique intégrant une tendance sans saisonnalité. Le modèle de Winter, plus récent, intègre à la fois la tendance et la saisonnalité. Ces modèles sont aujourd'hui très répandus dans les progiciels de prévision. Cependant, la recherche sources modèles continues d'être très active.[17]

- ✧ **Régression linéaire** : le but de la Régression linéaire simple (res.multiple) est d'expliquer une variable Y à l'aide d'une variable X (res : plusieurs variables). aussi de chercher une fonction $f \approx f(x_i)$.

pour définir \approx il faut se donner un critère quantifiant la qualité de l'ajustement de la fonction f aux données, quand la fonction f est une droite.

c'est à dire $f \approx b_1(x_i) + b_0$ on parle de régression linéaire. une étude de régression linéaire débute toujours par un trace des observations $(x_i, y_i) \quad i=1 \dots n$. Cette première représentation permet de savoir si le modèle défini "linéaire" est pertinent.

- ✧ **SVM et SMO régression** : Une machine à vecteurs de support (SVM) est un type de modèle qui est optimisé de sorte que l'erreur de prédiction et la complexité du modèle sont simultanément minimisées. En dépit de ses nombreuses qualités admirables, la recherche dans le domaine des SVMs a été entravée par le fait que les solveurs de programmation quadratique (QP) ont fourni le seul algorithme d'entraînement connu pendant des années. Cette dernière fait la résolution des problèmes de SVM à partir de décomposition des grands problème QP en une série de sous problèmes QP plus petit.

L'algorithme d'optimisation minimale séquentielle (SMO) a été introduit (Platt,

1998, 1999b) comme un exemple extrême de décomposition. Parce que SMO utilise un sous-problème de taille deux, chaque sous-problème a une solution analytique. Ainsi, pour la première fois, les SVM pourraient être optimisés sans un solveur QP.

Les problèmes de régression sont particulièrement sensibles à ces problèmes car les entrées sont généralement des nombres réels avec des solutions qui ont de nombreux vecteurs de support mais aujourd'hui avec les évolutions SMO s'est avéré efficace dans ce Domain, est elle prédéfinie dans les plus part les outils de prévision comme WEKA, Tanagra et R.[30]

2.7.2 Méthodes qualitatives

✧ Les panels d'expert :

Lorsque il n'y a pas de données historiques, l'opinion d'expert peut être la seule source d'informations pour préparer les prévisions. La mise en commun des différentes prévisions peut être faite de plusieurs manières.

Un premier méthode consiste à mener de entrevus individuelles et à combiner les resultat. une seconde méthode consiste à réunir les experts dans une salle et leur demander d'en arriver à un consensus.[17]

✓ Avantage :

Cette méthode a l'avantage de donner une prévision dans un laps de temps relativement court de amène une variété de points de vue. Il est cependant plus difficile de séparer la prévision par territoires et produits spécifiques.[18]

✧ La méthode Delphi :

La méthode Delphi est une technique de group en dans laquelle des experts sont questionnés individuellement sur leur perception de la tendance future. Les individus ne se rencontrent pas en groupe pour réduire le risque qu'un consensus soit atteint à cause de personnalités dominantes ou de supériorités hiérarchiques.

Les prévisions et les justifications qui les accompagnent sont résumées par la personne s'occupant de l'enquête puis retournées aux experts avec d'autres questions. ce processus continue jusqu'à ce qu'un consensus soit atteint par le groupe.[17]

2.8 Conclusion

La détermination du bon modèle dépend des caractéristiques de l'historique des données pour des prévisions sans risques. Une fois le modèle choisi, il doit être suivi régulièrement afin d'assurer qu'il est toujours adéquat et que les hypothèses faites sur les données passées sont toujours applicables aux données à venir.

Dans ce chapitre nous avons donné un aperçu sur l'analyse prédictive tel que les caractéristiques, les deux type de méthodes (quantitatives et qualitatives), nous l'avons présenté comme étant une approche très puissante pour l'étude des futurs phénomènes.

les ontologies

3.1 Introduction

Avec l'apparition de l'intelligence artificielle la notion « connaissance » a été élargie dans l'informatique. Le passage d'« information » à « connaissance » est lié à l'expérience de l'action. Donc, la connaissance par définition c'est l'information et le mode d'emploi de cette information. Pour que cette connaissance peut être exploitée par l'ordinateur il faut la représenter.

Les ontologies sont très utilisées dans la représentation des connaissances sur le Web vu que la capacité qu'elle possède pour représenter et manipuler la richesse sémantique de l'information, aujourd'hui les ontologies ne sont considérées que relatives aux différents domaines de connaissances, elles répondent aux problèmes de représentation et de manipulation et aussi d'organisation, plus ou moins formellement, un ensemble de connaissances.

Dans ce chapitre nous représentons la notion d'ontologie en général, en mentionnant ces différents types, ces composants, les étapes de construction ainsi que les outils utilisés pour développer une ontologie.[19]

3.2 Origine de notion d'ontologie

Le terme « ontologie », construit à partir des racines grecques onto (ce qu'existe, l'existant) et logos (le discours, l'étude), En philosophie, l'Ontologie est une branche fondamentale de la Métaphysique, qui s'intéresse à la notion d'existence, aux catégories fondamentales de l'existant et étudie les propriétés les plus générales de l'être.

Par analogie, le terme est repris en informatique et en science de l'information, où une ontologie est l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations, que ce soit par les métadonnées d'un espace de noms, ou les éléments d'un domaine de connaissances.

L'émergence de la notion d'ontologie en informatique résulte en partie des travaux de recherches en intelligence artificielle au début des années 1990 où le mot ontologie est emprunté par l'informatique à la philosophie, Comment formaliser des connaissances pour qu'elles puissent être interprétées par les machines ? ; c'est une problématique centrale de ce domaine

qui a motivé de nombreux travaux sur les représentations par objets. En suite les ontologies formelles sont constituées après l'utilisation conjointe des systèmes à bases d'objets et de logiques. Depuis 2001, l'émulation autour du Web Sémantique participe au développement de standards pour les ontologies et à la multiplication de leur utilisation. [20]

3.2.1 Définition

En informatique, une ontologie est un ensemble structuré de concepts. Les concepts sont organisés dans un graphe dont les relations peuvent être des relations sémantiques ou des relations de composition et d'héritage (au sens objet), c'est un outil qui permet précisément de représenter un corpus de connaissances sous une forme utilisable par une machine.

D'après Gruber et son équipe à Stanford « une ontologie est une spécification formelle et explicite d'une conceptualisation partagée » on observe dans cette définition quatre caractéristiques qui nous détaillons ci-dessous :[21]

- ✧ **Spécification explicite** : signifie que les concepts, les propriétés, les relations, les fonctions, les restrictions et les axiomes de l'ontologie sont définis de façon déclarative.
- ✧ **Formelle** : exprimée dans un langage de syntaxe et de sémantique formalisées qui la rendre compréhensible par la machine.
- ✧ **Conceptualisation** : réfère à un modèle abstrait d'un phénomène du monde en identifiant les concepts appropriés à ce domaine.
- ✧ **Partagé** : admise par l'ensemble des membres (et des systèmes) d'une communauté.

3.3 Pourquoi développer une ontologie

Une ontologie définit un vocabulaire commun pour les chercheurs qui ont besoin de partager l'information dans un domaine. Elle inclue des définitions lisibles en machine des concepts de base de ce domaine et de leurs relations.

Donc une ontologie est développée Pour les raisons suivantes :

- ✧ **Permettre la réutilisation du savoir sur un domaine.**
 - ✓ Créer et conserver des bases de connaissance réutilisables.
 - ✓ Assembler des bases de connaissances à partir de modules réutilisables .
- ✧ **Analyser le savoir sur un domaine.**
 - ✓ Partage de la connaissance et communication .
 - ✓ Assurer l'interopérabilité entre systèmes.
 - ✓ Permettre l'échange de connaissances entre systèmes.
- ✧ **Distinguer le savoir sur un domaine du savoir opérationnel.**
- ✧ **Expliciter ce qui est considéré comme implicite sur un domaine.**[22]

3.4 Domaines d'utilisation

Les ontologies ont des applications diverses et qui couvrent plusieurs domaines et activités comme ils sont montrés dans la figures suivante, et d'autre qui ne montre pas comme l'ingénierie des bases de données, traduction, datamining.

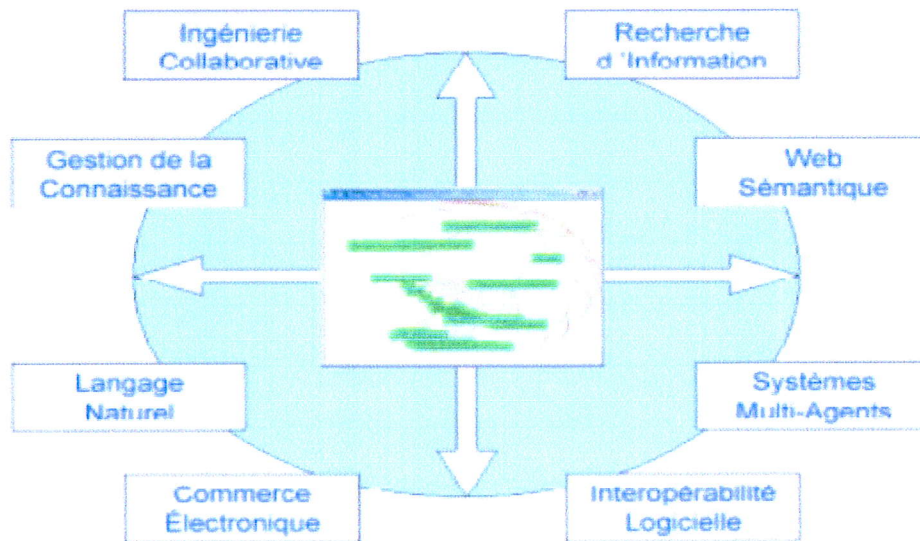


FIGURE 3.1 – les domaines d'utilisation des ontologies.

[23]

3.5 Les composants d'une ontologie

Une ontologie peut être vue comme un ensemble structuré de concepts et de relations entre ces concepts destinés à représenter les objets du monde sous une forme compréhensible aussi bien par les hommes que par les machines. Les composants d'une ontologie sont :

- ✧ **Concept** :ou classe, définissant un ensemble d'objet, abstrait ou concret, que l'on souhaite modéliser pour un domaine donné. Les connaissances portent sur des objets auxquels on se réfère à travers des concepts.
- ✧ Expliciter ce qui est considéré comme implicite sur un domaine

Un concept peut représenter un objet matériel, une notion, une idée. Il est caractérisé par un ensemble de propriétés :

- ✓ Concept :Un concept est générique s'il n'admet pas d'extension. La vérité, par exemple, n'a pas d'extension..
- ✓ Un concept porte une propriété d'identité si cette propriété permet de différencier deux instances de ce concept.

- ✓ Un concept est rigide s'il ne peut pas être une instance d'autres concepts.
 - ✓ Un concept est anti-rigide s'il peut être une instance pour d'autres concepts.
- ◇ **Les instances** :Ou individus, constituent la définition extensionnelle de l'ontologie (pour représenter les éléments spécifiques).
 - ◇ **Les relations** :Une relation permet de lier des instances de concepts ou des concepts génériques.Elles sont caractérisées par un terme ou plusieurs, et une signature qui précise le nombre d'instances de concepts que la relation lie, leurs types et l'ordre des concepts, c'est a dire la façon dont la relation doit être lue.
 - ◇ **Fonctions** :Les fonctions sont des cas particuliers de relations dans lesquelles un élément de la relation est défini à partir des autres éléments.
 - ◇ **Les axiomes** :Une ontologie est en outre composée d'axiomes qui forment des contraintes sémantiques pour le raisonnement et donnent un acompte d'une conceptualisation. Ils prennent la forme d'une théorie logique. Les axiomes formels sont utilisés pour vérifier la consistance de l'ontologie.

3.6 Type d'ontologie

◇ selon le niveau de granularité

On peut distinguer les ontologies selon le niveau de détail utilisé lors de la conceptualisation de l'ontologie en fonction de l'objectif opérationnel envisagé pour l'ontologie, deux catégories peuvent être identifiées : granularité fine et granularité large [21]

- ✓ **Granularité fine** :elle correspond à une ontologie très détaillée, possédant ainsi un vocabulaire plus riche capable d'assurer une description détaillée des concepts pertinents d'un domaine ou d'une tâche.
- ✓ **Granularité large** :elle correspond à un vocabulaire moins détaillé. Les ontologies génériques possèdent une granularité large, compte-tenu du fait que les notions sur lesquelles elles portent peuvent être raffinées par des notions plus spécifiques.

◇ Selon l'objet de conceptualisation

Il existe plusieurs sortes d'ontologies, destinées à des utilisations très variées. Elles se divisent en plusieurs catégories que nous détaillerons brièvement.

- ✓ **Ontologie supérieure** :Dites aussi de haut niveau « top level ontologies ou Upper Level ontologies »,ontologies « génériques ou noyaux d'ontologies », « méta ontologie», «ontologies de sens commun/général »; elles sont universelles, réutilisables et

référencables à partir des concepts des autres niveaux d'ontologies. Elles comportent des concepts abstraits (généraux) subsumant les concepts existant dans les différents domaines. Une ontologie de haut niveau est généralement conçue afin de réduire les incohérences des termes définis plus bas dans la hiérarchie. Il n'existe pas pour le moment d'ontologies de haut niveau unifiées[19]

- ✓ **Ontologie des tâches** : Une ontologie de tâches sert à gérer un certain type de tâches dans les systèmes : diagnostic, planification, conception, configuration, tutorat, etc. Ce type d'ontologie est lié à la résolution de problèmes et décrit un ensemble de vocabulaires concernant une tâche, notamment en spécialisant les concepts d'une ontologie supérieure.[21]
- ✓ **Ontologie générique** : Il s'agit d'une méta-ontologie, une ontologie de l'ontologie, définit un ensemble de primitives de représentation d'une ontologie. Elle véhicule des connaissances génériques, moins abstraites que celles que l'on retrouve dans les ontologies supérieures, mais quand même suffisamment générales pour qu'elles soient réutilisées à travers différents domaines.[21]
- ✓ **Ontologie de domaine** : Ce type d'ontologie régit un ensemble de vocabulaires et de concepts qui décrivent un domaine d'application ou monde cible. Il s'agit d'une méta-description d'une représentation des connaissances, une sorte de méta-modèle de connaissances dont les concepts et les propriétés sont de type déclaratif. La plupart des ontologies existantes sont des ontologies de domaine.[21]
- ✓ **Ontologie d'application** : Ce sont les ontologies les plus spécifiques. Contrairement à l'ontologie de domaine, l'ontologie d'une application donnée ne peut pas être réutilisée pour d'autres applications. Elle contient les connaissances requises pour une application particulière. Ce type d'ontologie décrit des concepts qui dépendent à la fois d'un domaine particulier et d'une tâche particulière. Par conséquent, elle spécialise souvent des ontologies de domaine et des ontologies de tâches pour une application donnée.[21]
- ✓ **Ontologies de représentation** : Ce type d'ontologies est un cas particulier d'ontologies supérieures qui regroupe des concepts déjà utilisés pour formaliser les connaissances. Indépendamment des domaines puisqu'elles décrivent des primitives cognitives communes. Parmi les ontologies de représentation, on trouve la «Frame-Ontology» qui définit de manière formelle, les concepts utilisés principalement dans des langages à base de frames : classes, sous classes, attributs, valeurs, relations et axiomes .[19]
- ✓ **Les ontologies de raisonnement** : regroupe les processus de raisonnement appliqués aux connaissances qui forment eux-mêmes un domaine de connaissance. On parle particulièrement d'ontologies développées pour représenter des connaissances génériques mises en œuvre lors de la résolution automatique de problèmes.[19]
- ✓ **Ontologie informatique** : Une ontologie informatique est une représentation de propriétés générales de ce qui existe dans un formalisme supportant un traitement rationnel. C'est le résultat d'une formulation exhaustive et rigoureuse de la conceptualisation

d'un domaine. Cette conceptualisation est souvent qualifiée de partielle car, en l'état de l'art, il est illusoire de croire pouvoir capturer dans un formalisme toute la complexité d'un domaine. Notons aussi que le degré de formalisation d'une ontologie varie avec l'usage qui en est envisagé.

C'est pour cet aspect de description de l'existant et de ses catégories que les ontologies informatiques ont emprunté leur nom à l'Ontologie philosophique. De ce rapprochement vient aussi la possibilité d'adapter des méthodes de la philosophie pour proposer des méthodes d'ingénierie d'ontologies.

Considérons l'exemple de la figure suivante. On y voit une certaine « scène du monde ». La description de cette scène demande deux choses : un vocabulaire non ambigu, aussi appelé vocabulaire conceptuel ou ontologie ; une énonciation des faits de la scène, reposant sur l'utilisation du vocabulaire de l'ontologie.

D'un point de vue pratique, une ontologie informatique permet, en particulier grâce aux travaux de l'intelligence artificielle symbolique sur les systèmes à base de connaissances et les moteurs d'inférence, d'implanter des mécanismes de raisonnement déductif, de classification automatique, de recherche d'information, et d'assurer l'interopérabilité entre plusieurs systèmes de ce type[24]

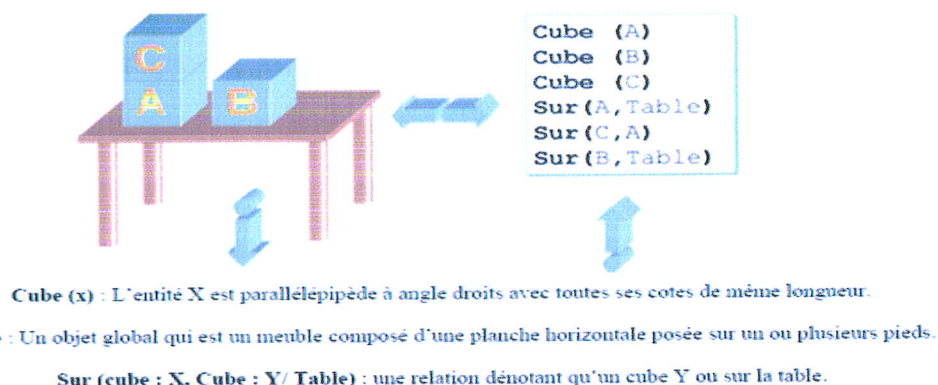


FIGURE 3.2 – exemple d'une ontologie informatique.

3.7 Les étapes de construction des ontologies

Pendant le pré-développement, une étude de l'environnement est effectuée pour connaître l'environnement dans lequel l'ontologie sera utilisée, les applications dans lesquelles l'ontologie sera intégrée, etc. Aussi pendant le pré-développement, l'étude de faisabilité répond aux questions telles que : est-il possible ou est-il approprié de construire l'ontologie ?, etc.

Pour développer une ontologie on doit être pris en considération les points généraux qui nous abordons dessous.[24]

❖ **Etape1. Spécification :**

On commence par la définition du domaine et de sa portée. C'est-à-dire, préciser les réponses des questions de base tel suivent :

- Qu'elle est le domaine que l'ontologie couvrira ?
- Quels sont les buts de l'utilisation de l'ontologie ?
- A quel type de question l'information contenue dans l'ontologie devra-t-elle fournir des réponses ?
- Qui va utiliser et maintenir une ontologie ?

❖ **Etape 2. Envisager une éventuelle réutilisation des ontologies existantes :**

Réutiliser des ontologies existantes peut même constituer une exigence si le système a besoin d'interagir avec d'autres applications qui utilisent déjà des ontologies spécifiques.

❖ **Etape3. Enumérer les termes importants dans l'ontologie :**

En définissant les points suivants :

- ✓ Définir les termes.
- ✓ Définir les propriétés de ces termes.
- ✓ Définir la sémantique de chaque terme.

❖ **Etape 4.définir les classes et la hiérarchie des classes :**

pour développer une hiérarchie de classes ,Il existe un certain nombre d'approche possible :

- ✓ Un procédé de développement de haut en dans commence par une définition des concepts les plus généraux de domaine et se poursuit et se poursuit par la spécialisation des concepts.
- ✓ procédé de développement de bas en haut commence par la définition des classes les plus spécifiques, les feuilles d'une hiérarchie, et se poursuit avec le regroupement de ces classes en concept généraux.
- ✓ Un procédé combiné de développement est une combinaison des deux approches, de haut en bas et de bas en haut. Au tout début, les concepts les plus saillants sont définis, ensuite ils sont généralisés ou spécialisés, selon le cas.

❖ **Etape 5. Définir les propriétés des classes :**

Les classes seules ne fourniront pas assez d'informations pour l'expression de connaissance sur un domaine. Après avoir défini quelque classe, nous devons décrire la structure interne des concepts.

❖ **Etape 6. Définir les facettes des propriétés :**

Les attributs peuvent avoir plusieurs facettes qui peuvent décrire le type des valeurs, les valeurs autorisées, le nombre de valeurs (cardinalité), et d'autres caractéristiques de valeurs que les attributs peuvent avoir.

✓ **La cardinalité des attributs :**

- Cardinalité unique.
- Cardinalité multiple.
- Cardinalité minimale facette.

✓ **Le type de valeur des attributs :**

La facette décrit les types de valeurs pouvant être affectés à l'attribut. Voici une liste de type de valeurs le plus typiques :

- Chaine de caractères.
- Nombre.
- Booléens.
- Enumérés.
- Instances.

✓ **Le domaine des attributs .**

✓ **Les classes auxquelles un attribut est rattaché ou les classes dont l'attribut décrit les propriétés, sont appelées le domaine d'un attribut.**

❖ **Etape 7. Créer les instances :**

La dernière étape de construction de l'ontologie consiste à créer les instances des classes dans la hiérarchie. Définir une instance individuelle d'une classe exige :

- Choisir une classe.
- Créer une instance individuelle de cette classe.
- La renseigner avec les valeurs des attributs.

3.8 Le graphe d'ontologie

Le graphe d'ontologie est un graphe orienté et étiqueté. Ses noeuds représentent des entités de l'ontologie et possèdent deux étiquettes. L'une des étiquettes de noeud est le

représentée alors que l'autre identifie la catégorie de ladite entité. Comme catégorie d'entité on peut avoir CLASSE, OBJET, RELATION, PROPRIÉTÉ, etc. Deux noeuds du graphe d'ontologie sont reliés par un arc si et seulement s'il existe un li qu'ils représentent. Chaque arc est étiqueté par le nom du lien qu'il encode.

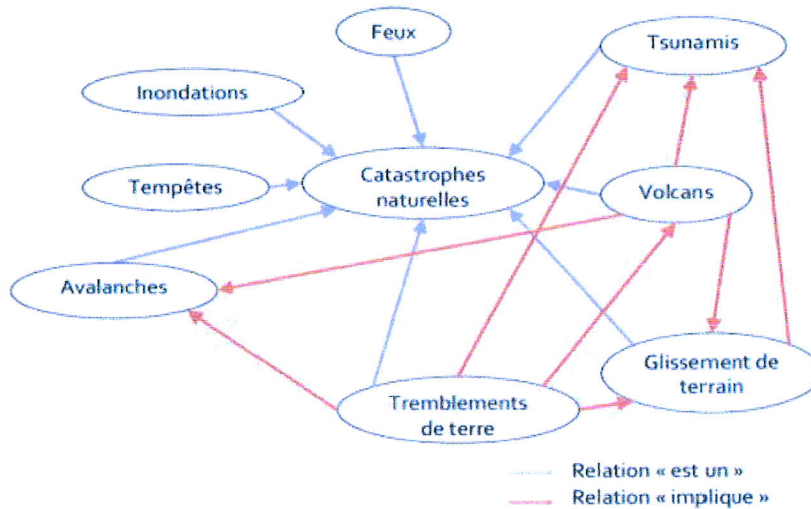


FIGURE 3.3 – exemple d'un graphe d'ontologie.

3.9 Cycle de vie

Les ontologies étant destinées à être utilisées comme des composants logiciels dans des systèmes répondant à des objectifs opérationnels différents, leur développement doit s'appuyer sur les mêmes principes que ceux appliqués en génie logiciel. Ainsi, les ontologies doivent être considérées comme des objets techniques évolutifs et possédant un cycle de vie qui nécessite d'être spécifié. Dans ce contexte, les activités liées aux ontologies sont d'une part des activités de gestion incluant la planification, le contrôle, et la garantie de qualité, et d'autre part des activités orientées développement regroupant les activités de pré-développement, de développement et de post-développement. Un cycle de vie inspiré du génie logiciel (voir figure), comprend une étape initiale d'évaluation des besoins, qui se transforme en idée, la concrétisation de l'idée qui se traduit par la conception qui est diffusée pour son utilisation. Vient ensuite l'étape de l'évaluation qui donne naissance, le plus souvent à une étape d'évolution et de maintenance du modèle. Une réévaluation de l'ontologie et des besoins devra se faire après chaque utilisation significative. L'ontologie peut être étendue et, si nécessaire, en partie reconstruite.

La validation du modèle de connaissances est au centre du processus et se fait de manière itérative.

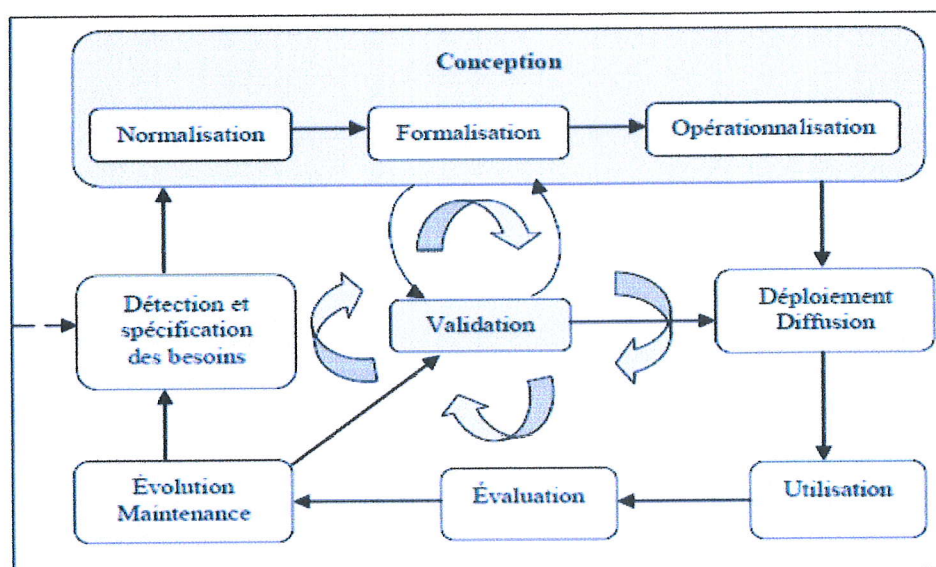


FIGURE 3.4 – cycle de vie d’une ontologie.

3.10 Les problèmes de construction des ontologies

En générale la construction d’une ontologie pose deux types de problèmes :

- Un problème de modélisation ou d’abstraction qui consiste à déterminer le contenu de l’ontologie, à trouver les concepts et les relations à représenter.
- Le problème de représentation qui consiste à déterminer le formalisme à adopter et à représenter les connaissances dans le formalisme retenu.

[25]

3.11 Les critères d’évaluation des ontologies

Les critères pertinents pour évaluer une ressource ontologique sont ceux qui doivent être respectés lors de la construction de cette même ressource. Ainsi, il est important de respecter un certain nombre de contraintes pour obtenir une modélisation de qualité, à la fois adaptable et maintenable. Nous présentons ci-dessous une série de principes, qui nous semblent particulièrement pertinents .[26]

- La clarté (objectivité) :La définition (complète et documentée en langage naturel) d’un concept doit faire passer le sens voulu du terme de manière objective.
- La perfection :Une définition exprimée par des conditions nécessaires et suffisantes est préférable à une définition partielle.
- La cohérence et l’extensibilité :Une ontologie doit être cohérente pour permettre des inférences conformes aux définitions. Les extensions doivent être anticipées (l’ajout de nouveaux concepts ne doit pas conduire à toucher aux fondations de l’ontologie).
- La déformation d’encodage minimale :La déformation d’encodage (lorsque la spécification

influe sur la conceptualisation) doit être minimale.

-L'engagement ontologique minimal :L'ontologie doit définir uniquement des termes nécessaires pour partager les connaissances consistantes de la théorie (le reste est superflu).

3.12 Niveau de formalisme de représentation

D'autre part, selon le niveau du formalisme de représentation du langage utilisé pour décrire l'ontologie, proposent une classification comprenant quatre catégories :

- ❖ **Informelles** : Ontologies opérationnelles dans un langage naturel (sémantique ouverte).
- ❖ **Semi-formelles** :Langage artificiel défini formellement.
- ❖ **Formelles** :Utilisation d'un langage artificiel contenant une sémantique formelle, ainsi que des théorèmes et des preuves de propriétés telles la robustesse et l'exhaustivité.

3.13 Outils de développement d'ontologies

De nombreuses plateformes logicielles utilisant des formalismes variés et offrant différentes fonctionnalités ont été développées pour supporter les ontologistes dans les différentes activités du cycle de vie d'une ontologie. Nous présentons ici brièvement les principaux outils de construction d'ontologies.

3.13.1 Les outils dépendants du formalisme de représentation

- ❖ **Ontolingua** :Le serveur Ontolingua est le plus connu des environnements de construction d'ontologies en langage Ontolingua. Il consiste en un ensemble d'environnements et de services qui supportent la construction en coopération d'ontologies, entre des groupes séparés géographiquement. Il supporte plusieurs langages et dispose de traducteurs permettant de passer de l'un à l'autre. Il y a trois différentes possibilités d'intégrer les ontologies Ontolingua :
 - **Inclusion** :Une ontologie inclut et utilise les définitions d'autres ontologies.
 - Restriction** :l'ontologie importe les définitions depuis d'autres ontologies et les rend plus spécifiques .
 - Raffinement polymorphe** : on redéfinit une définition importée depuis n'importe quelle ontologie.
- ❖ **OilEd** :L'éditeur OILEd10 a été développé en 1991 sous la responsabilité de l'université de Manchester. OILEd a été conçu pour éditer des ontologies dans le langage de représentation OIL, il est souvent considéré comme une simple interface. Cet éditeur offre également les services d'un raisonneur, FaCT(un moteur d'inférences bâti sur OIL)

qui permet de tester la satisfiabilité des définitions de classes et de découvrir des sub-somptions restées implicites dans l'ontologie. L'outil dispose de mécanismes pour la classification et le contrôle de la cohérence des ontologies.

- ❖ **OntoSaurus** :OntoSaurus est composé de deux modules : un serveur utilisant LOOM comme langage de représentation des connaissances, et en un serveur de navigation créant dynamiquement des pages HTML qui affichent la hiérarchie de l'ontologie ; le serveur utilise des formulaires HTML pour permettre à l'utilisateur d'éditer l'ontologie. Il utilise LOOM comme langage de représentation des connaissances. On peut représenter les concepts, la taxonomie des concepts, les relations entre les concepts, les fonctions, les axiomes et les instances.
- ❖ **OilEd** :OILed a été conçu pour éditer des ontologies dans le langage de représentation OIL, il est souvent considéré comme une simple interface. Cet éditeur offre également les services d'un raisonneur, FaCT(un moteur d'inférences bâti sur OIL) qui permet de tester la satisfiabilité des définitions de classes et de découvrir des sub-somptions restées implicites dans l'ontologie. L'outil dispose de mécanismes pour la classification et le contrôle de la cohérence des ontologies.

3.14 Les outils indépendants de formalisme de représentation

- ❖ **ODE et WebOde** :ODE et WebOde L'outil ODE permet de construire des ontologies au niveau connaissance, comme le préconise la méthodologie METHONTOLOGY. L'utilisateur construit son ontologie dans un modèle de type frame, en spécifiant les concepts du domaine, les termes associés, les attributs et leurs valeurs, les relations de sub-somption. L'ontologie opérationnelle est alors générée en utilisant les formalismes ONTOLINGUA.
- ❖ **OntoEdit** :OntoEdit (Ontology Editor) est également un environnement de construction d'ontologies indépendant de tout formalisme mis au point par l'institut AIFB de l'université de Karlsruhe et qui est maintenant commercialisé par la société Ontoprise GmbH. Il permet l'édition des hiérarchies de concepts et de relations et l'expression d'axiomes algébriques portant sur les relations, et de propriétés telles que la généralité d'un concept. Des outils graphiques dédiés à la visualisation d'ontologies sont inclus dans l'environnement. Ontoedit intègre un serveur destiné à l'édition d'une ontologie par plusieurs utilisateurs. Un contrôle de la cohérence de l'ontologie est assuré à travers la gestion des ordres d'édition.
- ❖ **Protégé** :PROTÉGÉ a été développé par le Stanford Medical Informatics de l'Université de Stanford.

Protégé est une plate-forme Open Source autonome, qui fournit un environnement graphique permettant l'édition, la visualisation et le contrôle (vérification des

contraintes) d'ontologies. Le modèle de représentation de connaissances de PROTÉGÉ, est issu du modèle des frames. Ce dernier contient des classes (pour modéliser les concepts), des slots (pour modéliser les attributs des concepts) et des facettes (pour définir les valeurs des propriétés et des contraintes sur ces valeurs), ainsi que des instances des classes.

protégé introduit la notion de métaclasse, dont les instances sont des classes. L'interface très complète ainsi que l'architecture logicielle extensible permettant l'insertion de plusieurs pluggins offrant de nouvelles fonctionnalités, notamment des pluggins pour gérer les représentations sous forme graphique, par exemple OWLViz et la prise en charge de nouveaux langages. Toutes ces caractéristiques ont participé à son succès et le rendent l'éditeur d'ontologie jouissant de la plus grande renommée à l'heure actuelle. Protégé est un outil employé par les développeurs et des experts de domaine pour développer des systèmes basés sur les connaissances (Ontologies).

Des applications développées avec Protégé sont employées dans la résolution des problèmes et la prise de décision dans un domaine particulier.

-Protégé permet aussi de créer ou d'importer des ontologies écrites dans les différents langages d'ontologies tel que : RDF-Schéma, OWL, DAML, OIL, ?etc

3.15 Langages pour les ontologies

Plusieurs langages de spécification d'ontologies (ou langages d'ontologies) ont été développés pendant les dernières années. La plupart de ces langages se basent ou sont proches sur la logique du premier ordre, et représentent donc les connaissances sous forme d'assertion (sujet, prédicat, objet). Ces langages sont typiquement conçus pour s'abstraire des structures de données et se concentrer sur la sémantique. Parmi les formalismes les plus employés se basant sur la logique des prédicats, on retrouve des langages comme N3 ou N-Triple. On peut aussi évoquer le langage DEF.

Par ailleurs, dans le cadre de ses travaux sur le Web sémantique, le W3C a mis en place en 2002 un groupe de travail dédié au développement de langages standards pour modéliser des ontologies utilisables et échangeables sur le Web. S'inspirant de langages précédents comme DAML+OIL et des fondements théoriques des logiques de description, ce groupe a publié en 2004 une recommandation définissant le langage OWL fondé sur le standard RDF et en spécifiant une syntaxe XML. Plus expressif que son prédécesseur RDFS, OWL a rapidement pris une place prépondérante dans le paysage des ontologies et est désormais, de facto, le standard le plus utilisé.[27]

- ❖ **RDF** :La standard RDF est un langage formel qui permet d'affirmer l'existence de relations entre les objets. Elles sont donc adaptées à l'expression des annotations que l'on veut associer aux ressources du Web.Le modèle RDF définit trois types d'objets :

-**Ressources** :Les ressources sont tous les objets décrits par RDF. Généralement, ces ressources peuvent être aussi bien des pages Web que tout objet ou personne du monde réel. Les ressources sont alors identifiées par leur URI.

-**Propriétés** :Une propriété est un attribut, un aspect, une caractéristique qui s'applique à une ressource. Il peut également s'agir d'une mise en relation avec une autre ressource.

-**Valeurs** :Les valeurs en question sont les valeurs particulières que prennent les propriétés. Ces trois types d'objets peuvent être mis en relation par des assertions, c'est-à-dire des triplets (ressource, propriété, valeur), ou encore (sujet, prédicat, objet), une collection des triplets construit un graphe RDF. Dans la figure suivante on donne des URIs sous le forma préfixe : nom où le préfixe designe un espace de nommage :

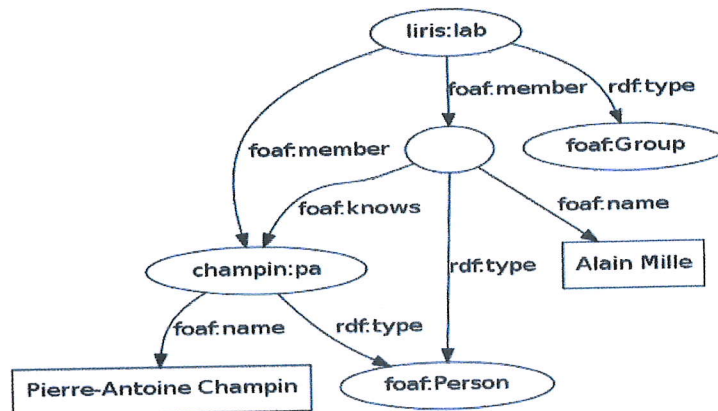


FIGURE 3.5 – exemple d'un graphe RDF.

- Une ressource identifier par une URI est représentée par une ellipse.
- Sur l'arc la propriété est identifiée par l'URI de préfixe `rdf` et le nom type.
- Le noeud rectangulaire représente la valeur prise par la propriété, ici l'objet est un littéral non typé, ce dernier peut être typé.
- L'ellipse vide représente une ressource anonyme dont on ne connaît pas l'URI.
- Le sujet peut être un URI ou un noeud vide.
- Le prédicat toujours un URI.
- L'objet peut être un URI, un noeud vide ou un littéral.

❖ **RDFS (RDF Schéma)** :RDF Schéma est le vocabulaire de structuration proposé pour la description du savoir.

Répandu aussi sous l'acronyme RDFS, il répond aux principes du RDF (collection de triplets organisés en graphes) et propose des noms de balises et attributs XML standardisés (tant dans leur forme que dans leur organisation) pour l'échange et le partage des données décrivant les savoirs. Le RDF Schéma permet de :

- Définir un « schéma » de méta donné.

-Qualifier les relations décrites en RDF.

❖ **XML** :Le XML est un Langage informatique orienté texte et formé de balise qui permet d'organiser les données de manière structurée.

LeXML est un standard utiliser pour le stockage du document et pour la transmission de donnée entre application.Le XML répond à certains principes rigoureux dont la validation nécessite le plus souvent une DTD ou autre schéma. Syntaxique, mais également à une logique informationnelle, une rigueur dans la formulation (ce qui lui donne ce degré sémantique).

❖ **OWL** :OWL est un langage d'ontologies Web. Les premiers langages utilisés pour le développement d'outils et d'ontologies pour des communautés d'utilisateurs spécifiques n'ont pas été définis pour être compatibles avec l'architecture du Web en général, et du Web Sémantique en particulier. OWL fournit des mécanismes pour créer tous les composants d'une ontologie : classes, instances, propriétés et axiomes. OWL repose également sur la syntaxe des triplets RDF et réutilise certaines des constructions RDFS.OWL a pour objectif principal de publier et partager les ontologies sur le web sémantique. [28]

❖ **Les niveaux d'OWL** :Il existe trois sous-langages se présentant avec une expressivité croissante, les voici respectivement :

-OWL Lite :Est un langage le plus simple syntaxiquement, il permet de créer des hiérarchies de concepts sur lesquelles des contraintes simples peuvent être définies.

-OWL-DL : Est une extension d'OWL Lite. Il permet d'avoir une expressivité plus importante que celui-ci. Il permet de garantir la complétude des raisonnements et leur décidabilité (leur calcul se fait en une durée finie).

OWL-Ful :Est une extension d'OWL-DL. Entre le sous-langage d'OWL, il est le langage le plus complexe. Néanmoins, il ne garantit pas la complétude des raisonnements ni leur décidabilité.[29]



FIGURE 3.6 – la relation entre les 3 sous-langages.

3.16 Conclusion

Dans le présent chapitre, nous avons donné un aperçu global sur la notion d'ontologie en général, on a vu l'ontologie informatique qui est l'outil qui permet de présenter précisément un corpus de connaissance sous une forme utilisable par une machine, et enfin on a présenté les outils de développement d'ontologie tel que protégé avec les langages comme OWL que nous allons utiliser pour développer notre ontologie.

Conception du système

4.1 Introduction

Cette partie qui est consacrée à la conception de notre projet, consiste : Premièrement à concevoir une ontologie qui servira à faciliter et à guider l'exploitation des données relatives aux médicaments par classe. En effet, pour une meilleure exploitation des données pharmaceutiques intégrées et pour bien exprimer les besoins des utilisateurs et faire apparaître l'aspect sémantique des données, on va concevoir une ontologie qui va refléter exactement la vision des utilisateurs envers les données.

Et deuxièmement à tester un ensemble de méthodes de prévision sur les anciennes données relatives à la consommation des produits pharmaceutiques pour faire dégager les meilleures méthodes qui s'adapteront à la prévision des besoins de la période future répondant ainsi aux besoins des utilisateurs dans le domaine de la prédiction.

4.2 Présentation du domaine

Les établissements de santé de proximité (EPSP) ont été créés par décret exécutif n° 07-140 du 19 mai 2007. Un établissement public de santé de proximité comme celui de Jijel est constitué d'un ensemble de polycliniques et de salles de soins couvrant un bassin de population. Il a pour mission de prendre en charge de manière intégrée et hiérarchisée :

- ✓ La prévention et les soins de base.
- ✓ Le diagnostic .
- ✓ Les soins de proximité .
- ✓ Les consultations de médecine générale et les consultations de médecine spécialisée de base .
- ✓ Les activités liées à la santé reproductive et à la planification familiale.
- ✓ La mise en oeuvre des programmes nationaux de santé et de population.

La pharmacie est une spécialité médicale mais aussi une science. C'est l'étude de la recherche, de l'élaboration, de la conception et des effets des médicaments. Elle joue un rôle tyrannique dans le fonctionnement d'un EPSP.

4.2.1 Contact avec le champ d'étude

Suite aux visites successives au service pharmacie de l'EPSP de Jijel et les rencontres effectuées avec le personnel chargé de la gestion de la pharmacie, nous avons pu collecter l'ensemble des données relatives à la consommation des produits pharmaceutique classées par classes et englobant les quantités et les prix unitaires.

Nous avons aussi pu recenser la totalité des besoins du personnel en matière d'un éventuel système exploitant ses données.

4.2.2 Préparation de données

Les données initiales sont représentées sous différente formes :

- ✓ Format papier qui contient des états de consommation mensuelle et trimestrielle et qui concernent une période de sept (07) années allant de l'année 2009 jusqu'à l'année 2015. Le schéma suivant illustre, comme exemple, la forme initiale des données :

Médicaments	Conssommations			Quantité	Prix Unitaire	Prix Total
	OCT	NOV	DEC			
VITAMINE C	400	0	0	400	409,72	1638,55
VOLTARENE 75 MG	1	0	0	1	633,64	633,64
RANITIDINE 50 MG	2	0	5	7	2778,75	19451,25
DICLOFENAC 75MG	0	0	12	12	2448,31	29385,72
TOTAL					51109,49	

FIGURE 4.1 – format initiale de données .

Format CSV qui contient des états de consommation mensuelle qui concernent les deux années dernier (2016 et 2017). Le schéma suivant illustre la forma CSV :

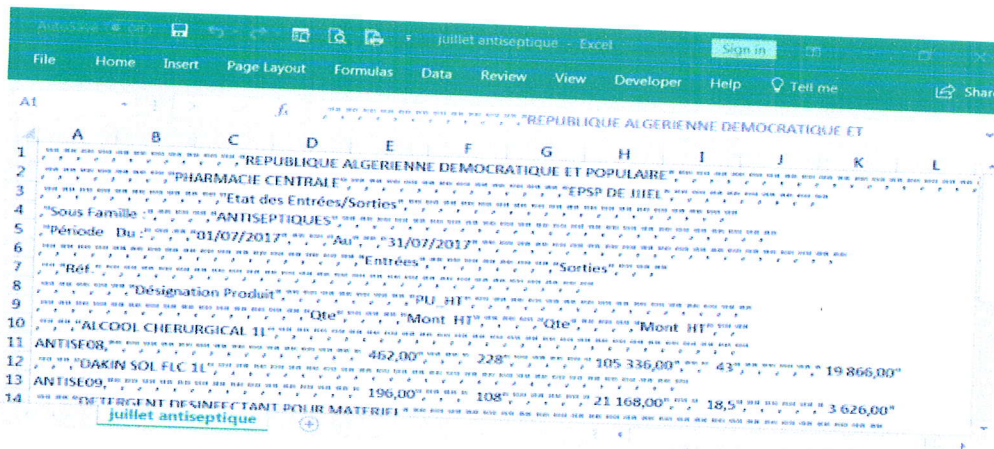


FIGURE 4.2 – les données au format csv.

On a transformé le format CSV au forma XLSX comme illustre la figure suivante :

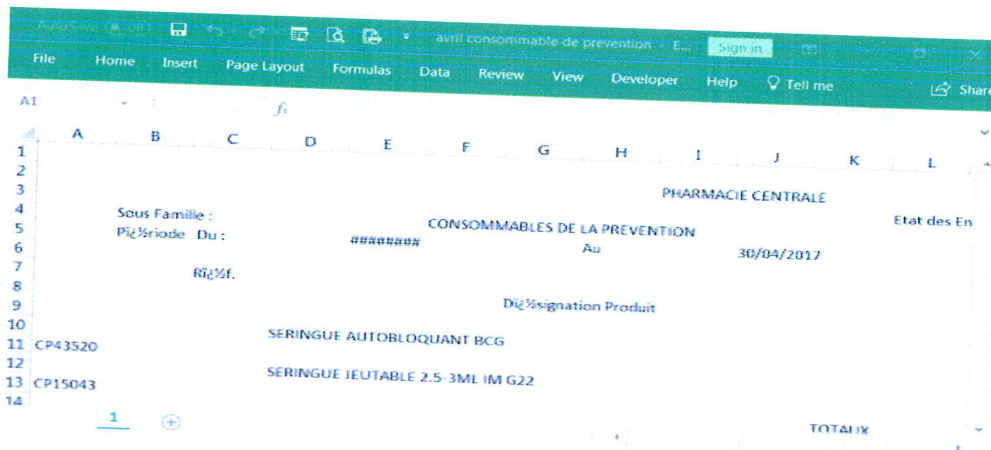


FIGURE 4.3 – Les données au format xlsx

4.3 Les produits pharmaceutiques

4.3.1 Définition

Le terme « produit pharmaceutique » désigne un élément fondamental tant dans la médecine moderne que la médecine traditionnelle, qui a pour but d'améliorer la qualité de vie et la santé des gens, et qui permette de faciliter la mission du personnel de santé dans l'administration des traitements et l'établissement du diagnostic Ce terme général englobe tous les produits retrouvés dans les pharmacies (médicaments, matériels médicaux), et mise sous la responsabilité des pharmaciens.

4.3.2 Classification des produits pharmaceutiques

Dans le cas de l'EPSP de Jijel c'est la pharmacie qui se charge de la gestion de stock des produits pharmaceutiques pour les différents services et les différentes unités de soins qui lui sont affiliés.

Donc Pour la mise en oeuvre de notre projet, plusieurs produits ont été pris en considération à savoir :

- ❖ Produits dentaires.
- ❖ Médicaments.
- ❖ Solutés.
- ❖ Films.
- ❖ Pansement.
- ❖ Gaz médicaux.

Transformation de donnée : Pour garantir une bonne transformation à la base de données, on a traité les données en suivant les étapes ci-dessous :

- ✓ Élimination des doublons.
- ✓ Traitement des champs vides et nuls.
- ✓ Normalisation de la casse de quelques chaînes de caractères.
- ✓ Corrections des valeurs non significatives.



4.4 Les données après traitement

Les données sont prêtes pour intégration dans une base de donnée, elle prennent le format suivant :

id_prod	id_mois	qte	prix_u	prix_t
A04	042009	54,00	368,75	19912,50
A06	042009	1,00	403,71	403,71
A10	042009	65,00	58,50	3802,50
A17	042009	0,00	3276,00	0,00
A26	042009	30,00	156,24	4687,14
A31	042009	0,00	410,00	0,00
A38	042009	32,00	350,00	11200,00
A04	052009	33,00	368,75	12168,75
A06	052009	3,00	403,71	1211,13
A10	052009	30,00	58,50	1755,00
A17	052009	1,00	3276,00	3276,00
A26	052009	13,00	156,24	2031,09
A31	052009	0,00	410,00	0,00
A38	052009	0,00	350,00	0,00
A04	062009	36,50	368,75	13459,38

FIGURE 4.4 – Les données collectées

4.5 Analyse des besoins des utilisateurs

Suite aux entretiens réalisés avec les gens de la pharmacie de l'EPSP de Jijel et suite aux besoins exprimés par leurs soins. La liste des besoins peut se résumer dans les points suivants :

- ✓ L'affichage des informations sur le nombre des produits pharmaceutiques ainsi que leurs caractéristiques : dosage, unité, prix.
- ✓ Regroupement des différents produits pharmaceutiques selon les quantités consommées.
- ✓ Des analyses comparatives sur le nombre de produits consommés (les plus et les moins consommés).
- ✓ Une analyse pour prédire les besoins futurs en produits pharmaceutiques.
- ✓ Les informations sur l'historique de consommation.
- ✓ L'affichage des recherches effectuées sous format tabulaire ou graphique pour faciliter la prise de décision.
- ✓ La possibilité d'imprimer ou d'exporter sous format EXCEL les différents résultats que ce soit ceux de la recherche, du regroupement ou de la prédiction.

Ces besoins sont demandés pour être effectués que ce soit pour :

- Toutes les catégories.
- Tous les produits d'une catégorie donnée.
- Un produit donné.
- Un produit d'une catégorie donnée.

Et cela selon différentes périodes voulues à savoir :

- Un mois ou un trimestre d'une année donnée.
- Un mois ou un trimestre de toutes les années.
- Une année donnée.
- Toutes les années (de 2009 à 2017).
- Entre deux mois ou deux trimestres d'une année donnée.
- Entre deux mois ou deux trimestres de toutes les années.
- Entre deux années.

4.6 Conception de l'ontologie

4.6.1 Présentation d'ontologie

Dans cette partie, notre travail consiste à développer une application qui sert à faciliter l'exploitation des données relatives aux médicaments par classe en concevant une ontologie répondant aux besoins des utilisateurs dans le domaine pharmaceutique.

Les parties composant notre ontologie sont :

- ✓ Médicament.
- ✓ produits Dentaire.
- ✓ Film.
- ✓ Soluté.
- ✓ Pansement.
- ✓ Gaz médicaux.

La figure suivante présente une partie d'ontologie des médicaments, elle contient trois classes principales :

- ✧ La classe « Thérapeutique » : c'est la classe globale, qui est composée des différentes « Sous classes ».
- ✧ La classe « sous-classe-thérapeutique » comprenant la classe « Médicament »
- ✧ La classe « sous-classe-médicament » contient les différents types des médicaments.
- ✧ La relation « EST UN » : PARACETAMOL AD INJ EST UN Thérapeutique.
- ✧ La relation « SUB CLASS OF » Les antibiotiques SUB CLASS OF médicament
- ✧ La relation « SAME AS » ASPEGIC 0,9 MG SAME AS ASPEGIC 900 MG IN

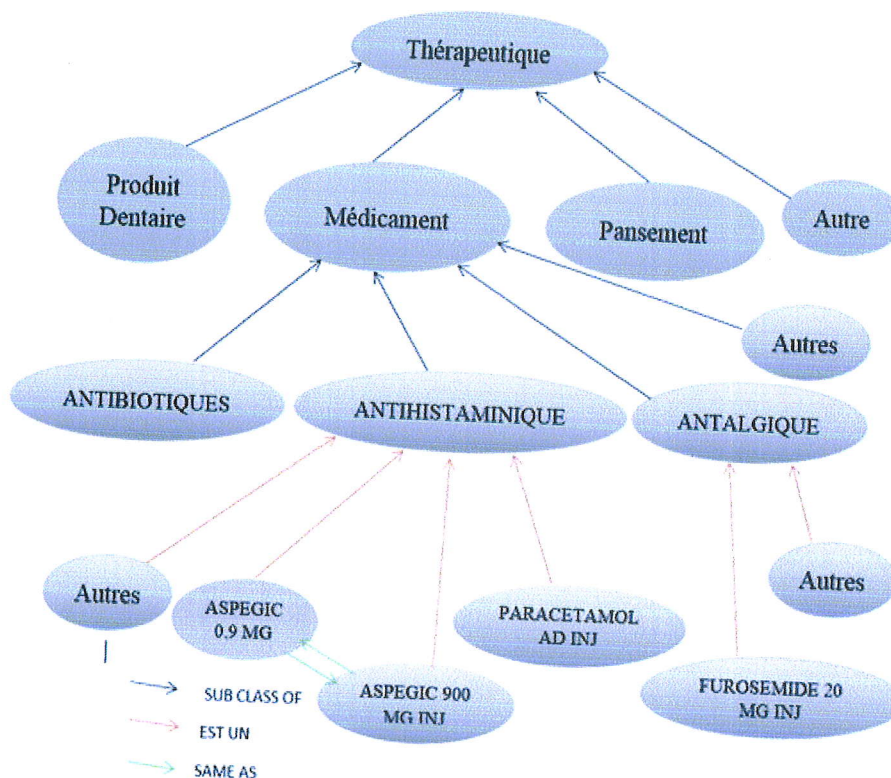


FIGURE 4.5 – Une partie d'ontologie

4.7 Classification des médicaments

La conception de l'ontologie se base sur la classification des médicaments illustrée dans le tableau suivant :

Classe	ID_MED	DESIGNATION
ANTISEPTIQUES	M08	ALCOOL IODE
	M13	BETADINE SOL
	M14	BETADINE SOL 1%
	M15	BETADINE SOL 10% (POLYVIDONE)
	M25	DAKIN
	M31	EAU OXYGENEE
	M32	EOSINE ACQUEUSE 2%
	M44	HEXOMIDINE 1% sol
ANTALGIQUES	M10	ASPEGIC 0,9 MG INJ
	M11	ASPEGIC 900 MG INJ
	M64	PARACETAMOL 10MG INJ ADULTE
	M65	PARACETAMOL 10MG INJ ENFANT

	M66	PARACETAMOL 10MG INJ NOURISSON
	M67	PARACETAMOL 10MG/ML
	M68	PARACETAMOL 25MG INJ ADULTE
	M69	PARACETAMOL AD INJ
ANTIASTHMATIQUES	M09	ANYL INJ 0.5 MG
	M12	ASTHALIN 5 MG INNHALATION
	M16	BRICANYL 0,5 mg
	M17	BRICANYL INHALATION
	M117	SALBUTAMOL 0,5 MG
	M118	SALBUTAMOL 0.5MG IV INJ
	M119	SALBUTAMOL AEROSOL
	M120	SALBUTAMOL INHALATION
	M121	SALBUTAMOL INHALATION 5MG
	M122	SALBUTAMOL INHALATION 5MG (ASTHALIN)
TRAQUILISANTS	M26	DIAZEPAM 10MG INJ (VALIUM)
	M27	DIAZEPAM 10MG/2ML INJ
	M150	VALIUM INJ 10 MG
ANTIINFLAMMATOIRES	M28	DICLOFENAC 75 MG INJ (VOLTARENE)
	M29	DICLOFENAC 75MG INJ
	M45	HYDROCORTISONE 100 MG INJ
	M46	HYDROCORTISONE HEMISUCINATE 100MG
	M55	METHLPRDNISOLONE 20 MG INJ
	M56	METHLPRDNISOLONE 40 MG INJ
	M57	METHYLPREDNISOLONE
	M125	SOLUMEDROL 20 MG
	M126	SOLUMEDROL 20 MG INJ
	M127	SOLUMEDROL 40 MG
	M128	SOLUMEDROL 40 MG INJ
	M153	VOLTARENE 75 mg
	M154	VOLTARENE 75 mg INJ
	M155	VOLTARENE 75MG
CARDIOLOGIE	M37	FUROSEMIDE 20 MG INJ
	M38	FUROSEMIDE 20MG INJ (LASILIX)
	M61	NICARDIPINE 10MG/10ML IV INJ
	M60	NICARDIPINE 10MG INJ (LOXEN)
ANTIPARASITAIRES	M39	GLUCANTIME 1,5 / 5 ml
	M40	GLUCANTIME 1,5 GR INJ

	M41	GLUCANTIME 1,5 MG INJ	
	M42	GLUCANTIME 1.5 GR INJ	
ANTIARYTMIQUES	M50	LIDOCAINE 2%	
	M51	LIDOCAINE 2% 10ML INJ	
	M52	LIDOCAINE 2% INJ	
	M53	LIDOCAINE INJ	
	M156	XYLOCAINE 2%	
	M157	XYLOCAINE INJ 2%	
	M158	XYLOCAINE VISQ 2%	
	ANTIEMETIQUES	M58	METOCLOPRAMIDE 10MG INJ
M59		METOCLOPRAMIDE 10MG INJ(PRIMPERAN)	
M75		PRIMPERAN	
M76		PRIMPERAN 10MG INJ	
M77		PRIMPERAN INJ	
ANTIHISTAMINIQUES	M78	PROMETHAZINE 25 MG INJ	
	M81	RANITIDINE 50 MG INJ	
ANTISPASMODIQUES	M70	PHLOROGLUCINOL 10MG INJ	
	M85	RIABAL 15 MG INJ	
	M129	SPASFON 10 MG	
	M130	SPASFON 10 MG INJ	
	M144	TIEMONIUM	
	M145	TIEMONIUM 5MG INJ (VISCERALGINE)	
	M146	TIEMONIUM METHYLSULFATE 5MG INJ	
	M151	VISCERALGINE 5 MG INJ	
	ANTIBIOTIQUES	M133	STRECHTOMYCINE 1 MG INJ
		M134	STRECHTOMYCINE 1GR INJ
M135		STREPTOMYCINE 1GR INJ	
ANTIHISTAMINIQUES	M141	TERBUTALINE 0,5 MG INJ(BRICANYL)	
	M142	TERBUTALINE 0.5MG INJ	
	M143	TERBUTALINE SOL	
DERMATOLOGIE	M147	TULLE GRAS	
	M148	TULLE GRAS GM	
	M149	TULLE GRAS PM	
AUTRES	M152	VITAMINE C	
	M01	ACETYL SALICYLATE DE LESYNE 900MG INJ	
	M02	ACETYLSALICYLATE DE LYSINE 900MG A	

M03	ACIDE SULPERIQUE
M04	ADRENALINE 1 MG INJ
M05	ADRINALINE 0,25 MG INJ
M06	ADRINALINE 1 MG INJ
M07	ALCOOL CHIRUGICAL
M18	CALCIUM CHLORURE 10% 10ML INJ
M19	CHLORURE DE POTASSIUM INJ
M20	CHLORURE DE SODIUM 5% 500 mg
M21	CHLORURE DE SODIUM 9% 250 mg
M22	COMPRESSES IMBEBEES
	COMPRESSES IMBEBEES DE
M23	VASELINE G.M
	COMPRESSES IMBEBEES DE
M24	VASELINE P.M
M30	DICYNONE 250MG INJ
M33	ETAMSYLATE 250MG INJ(DICYNONE)
M34	ETAMSYLATE 250MG/2ML INJ
M35	ETHAMBUTOL 400 MG CP
M36	ETHAMBUTOL 40MG CP
M43	GLUCOSE 5% 500 mg
M47	IZONIAZIDE 100 MG CP
M48	IZONIAZIDE 300 MG CP
M49	LARGACTIL INJ
M54	LOXEN 10 MG INJ
M62	OXYDE DE ZINC pde
M63	PANITIDINE 50MG INJ
M71	POLYVIDONE IODEE 10%
M72	POLYVIDONE IODEE 10% SOL DERM
M73	PRIFINIUM BROMURE 15MG
M74	PRIFINIUM BROMURE 15MG INJ
M79	PUROSEMIDE 20MG INJ
M80	PYRAZINAMIDE 400 MG CP
M82	RH 150 RIFAMPICINE 150/75 mg
M83	RHZ RIFAMPICINE 150/75/400 mg
M84	RHZ RIFAMPICINE 150/75/400/275 mg
M86	RIFAMP/ISON/PYRA
	RIFAMP/ISON/PYRA 60/30/150MG
M87	COMP
M88	RIFAMP/ISON/PYRA.150/75/4
	RIFAMP/ISON/PYRA.150/75/400MG
M89	COMP



M90	RIFAMP/ISON/PYRA/ETHAM
M91	RIFAMPICINE 150 MG CP
M92	RIFAMPICINE 150 MG GLES
M93	RIFAMPICINE 150/75 MG
M94	RIFAMPICINE 150/75 MG CP (RH)
M95	RIFAMPICINE 150/75 MG RH
M96	RIFAMPICINE 150/75 MG RH CP
M97	RIFAMPICINE 150/75/400
M98	RIFAMPICINE 150/75/400
M99	RIFAMPICINE 150/75/400 /250MG INJ
M100	RIFAMPICINE 150/75/400 MG CP
M101	RIFAMPICINE 150/75/400 MG CP (R.H.Z)
M102	RIFAMPICINE 150/75/400 MG RHZ
M103	RIFAMPICINE 150/75/400/275
M104	RIFAMPICINE 150/75/400/275
M105	RIFAMPICINE 150/75/400/275 MG
M106	RIFAMPICINE 150/75/400/275 MG CP RHZE
M107	RIFAMPICINE 150/75/400/275MG INJ
M108	RIFAMPICINE 300 MG CP
M109	RIFAMPICINE 300 MG GLES
M110	RIFAMPICINE 60/30 MG
M111	RIFAMPICINE 60/30 MG CP
M112	RIFAMPICINE 60/30/150 MG CP
M113	RIFAMPICINE/ISONIAZIDE
M114	RIFAMPICINE/ISONIAZIDE 150/75MG
M115	RIFAMPICINE/ISONIAZIDE 60/30MG COMP
M116	RINFAMPICINE150/75 MG RH
M123	SGI 15% 500
M124	SGI 5% 500
M131	SSI 9% 250
M132	SSI 9% 500
M136	SULFADIAZINE
M137	SULFADIAZINE ARGENTIQUE 25MG INJ
M138	SULFADIAZINE ARGENTIQUE CREME
M139	SULFADIAZINE ARGENTIQUE+FCEH CREME
M140	SULFADIAZINE CREME

TABLE 4.1- Classification des Medicaments

4.8 Méthode d'exploitation pour la prédiction

L'objectif principal de la prédiction est d'aider les gestionnaire de l'Établissement Public de Santé de Proximité lors de la prise de décision en ce qui concerne le réapprovisionnement en différents produits pharmaceutiques.

4.8.1 Définition du problème

L'un des principaux soucis des gestionnaires de stocks des établissements de santé est la précision dans la quantité des produits.leurs préoccupations est de se réapprovisionner en quantités suffisantes pour ne pas tomber dans le problème de manque ou de gaspillage.

4.8.2 La collecte des données

Afin de bien appréhender les besoins, la méthode principale de collecte de données et de l'étude de ces données est le planning des entretiens.En collaboration avec la pharmacienne de l'EPSP de Jijel , nous avons pu collecter les différentes informations nécessaires sur les produits pharmaceutiques consommés au cours des années précédentes (de 2009 à 2017)

4.8.3 Analyse préliminaire

Notre prévision est basée sur l'outil WEKA qui contient les quatre méthodes suivantes applicables sur les données quantitatives sous forme de série temporal :

- ✓ SMO reg.
- ✓ Multilayer Perceptron.
- ✓ Linéaire régression.
- ✓ Gaussian Processes.

Pour choisir les méthodes les plus adéquates avec ce type de données, une stratégie d'analyse a été élaborée et qui se résume dans les points suivants :

- 1.choisir des parties de données par exemple «la consommation des produit dentaire entre les années 2009 et 2013» comme illustre la figure suivante :

ANNEE	QTE
2009	684180
2010	170940
2011	218960
2012	647640
2013	385000

FIGURE 4.6 – la consommation des produit dentaire entre les 2009 et 2013



2. tester les quatre méthodes sur les données de période 2009-2012 pour prédire la consommation de l'année 2013, les quatre figures suivant montrent cette étape :

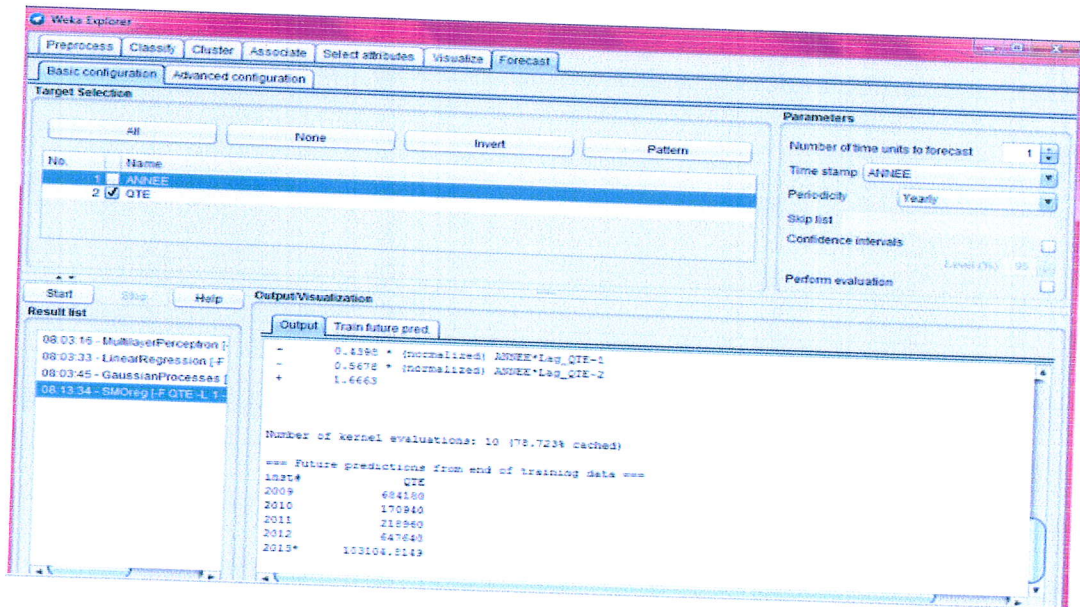


FIGURE 4.7 – résultats de la méthode SMO reg

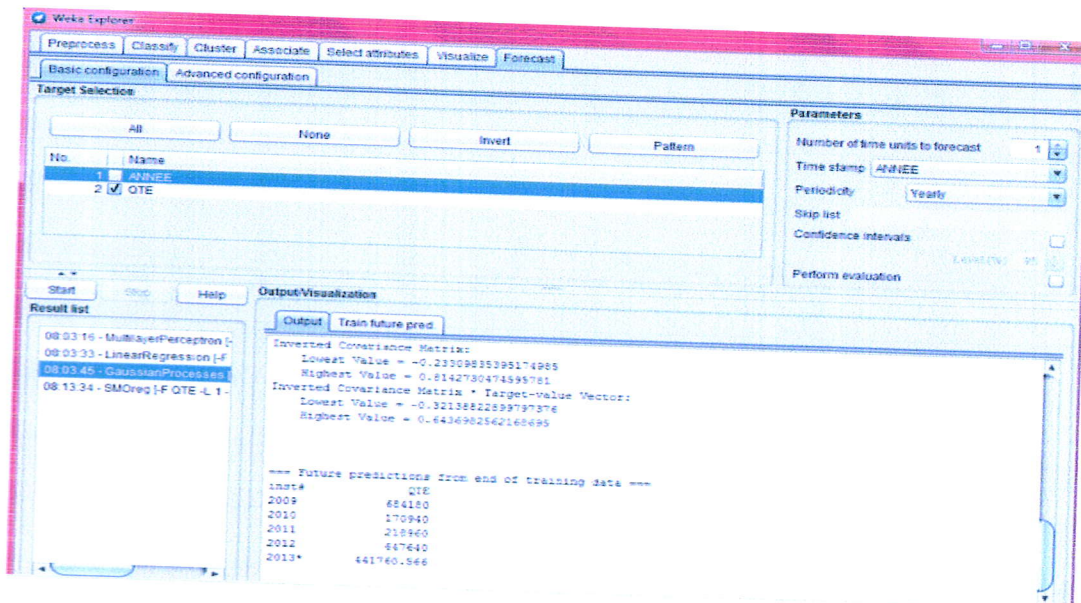


FIGURE 4.8 – résultats d'analyse par la méthode Gaussian Processes

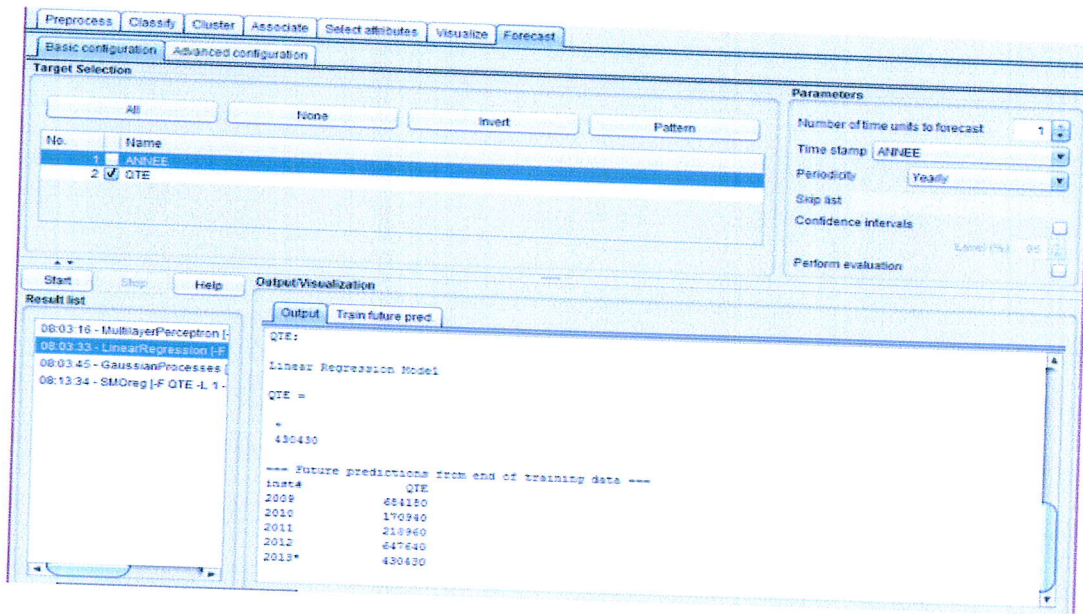


FIGURE 4.9 – résultats d'analyse par la méthode Linéaire Régression

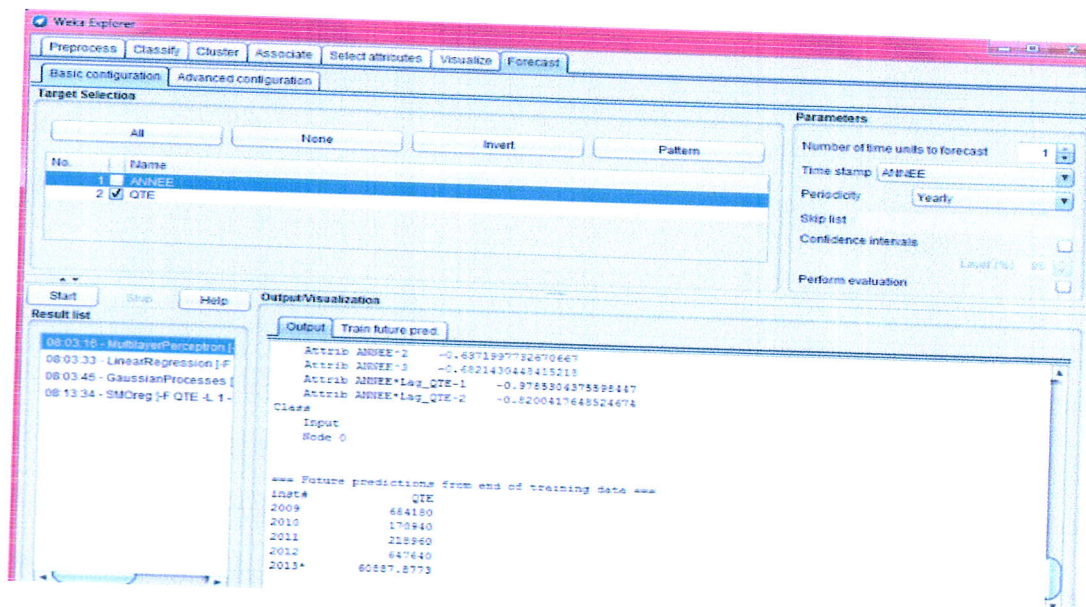


FIGURE 4.10 – résultats d'analyse par la méthode Multilayer Perceptron

3. Comparer les résultats avec la consommation réelle de l'année 2013 :

D'après la comparaison on a trouvé que les résultats des méthodes Linéaire régression et SMO reg sont les plus proches aux consommations réelles.

4.8.4 Choix de la méthode

Comme démontré dans l'analyse préliminaire, les méthodes les plus adéquates à utiliser sont celles Linéaire régression SMO reg prédéfinie sous WEKA.

4.9 Conclusion

Dans ce chapitre, nous avons tout d'abord présenté le domaine d'application du projet. Ensuite, nous avons énuméré les étapes de conception du système décisionnel : conception de l'ontologie, la préparation des données à classées dans cette dernière ainsi que l'analyse des besoins des utilisateurs. Enfin nous avons adopté les méthodes les plus appropriées pour la réalisation de la prédiction. Nos choix se sont portés respectivement sur les méthodes Linéaire régression et le SMO reg reflétant ainsi la nature de nos données.

Implémentation et réalisation

5.1 Introduction

Dans ce chapitre, nous allons détaillés l'implémentation de notre système en commençant par l'implémentation de l'ontologie, par la suite nous énumérons la succession d'étapes nécessaires à l'intégration du WEKA et api Jena dans NetBeans.

Enfin nous présentons l'interface qui permet de visualiser les différents résultats que ce soit ceux de la prédiction ou de la recherche et cela sous un format tabulaire ou graphique.

5.2 Les outils d'implémentations

Pour réaliser notre système nous somme basé sur les outils suivants : :

- ✓ Langage de programmation Java.
- ✓ NetBeans comme IDE de programmation Java.
- ✓ JfreeChart comme outil de reporting.
- ✓ Protégé pour les ontologies.
- ✓ Jena pour intégrer l'ontologie sous NetBeans.
- ✓ Weka pour la prévision.

5.3 Le schéma général du système

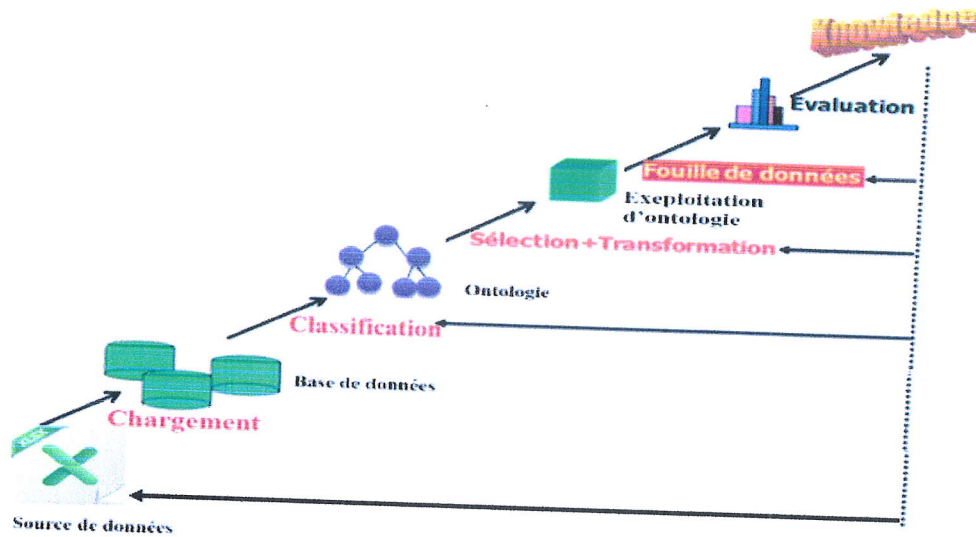


FIGURE 5.1 – Le schéma général du système .

5.4 Réalisation de notre ontologie

5.4.1 Pour l'édition de l'ontologie

On a choisi protégé 4.3, plusieurs raisons ont motivé notre choix :

- ✧ Protégé 4.3 permet d'importer et d'exporter des ontologies dans les différents langages d'implémentation d'ontologies (RDF-Schéma, OWL, DAML, OIL,...etc).
- ✧ Protégé 4.3 possède une interface modulaire, ce qui permet son enrichissement par des modules additionnel (plugins).
- ✧ Protégé 4.3 permet l'édition et la visualisation d'ontologies.
- ✧ Protégé 4.3 permet le contrôle de la cohérence de l'ontologie par des vérifications des contraintes.
- ✧ Protégé 4.3 est fourni avec une API écrite en JAVA, qui permet de développer des applications pouvant accéder aux ontologies de Protégé et de les manipuler.

5.4.2 Création d'une ontologie

Pour créer une ontologie, il faut tout d'abord créer un projet en choisissant le langage (ou format) désiré (RDF/XML, OWL/XML, OWL FonctionalSyntax, ?, etc). La fenêtre principale est comme suit :

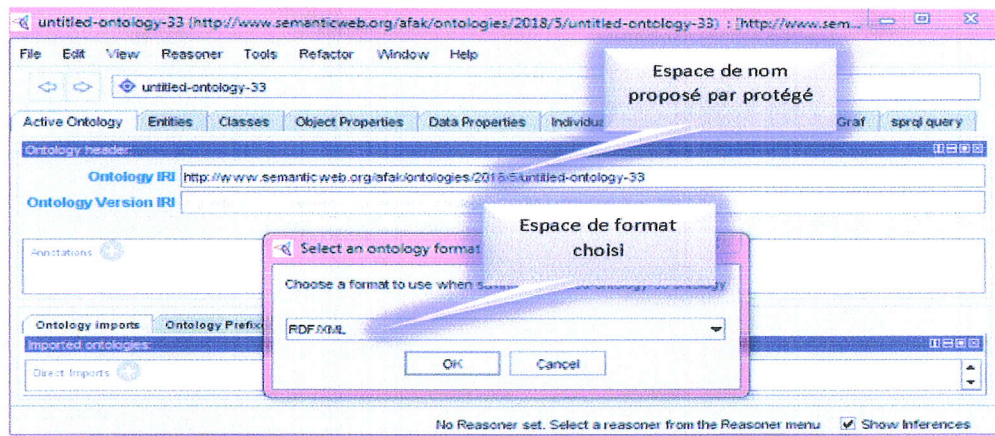


FIGURE 5.2 – Création d'une ontologie .

On peut parcourir les ontologies en utilisant les onglets suivants :

- L'onglet « OWLClasses » nous permet de voir et d'éditer des classes et leurs propriétés.
- L'onglet « objet propriété » nous permet d'éditer les relations entre les classes et les sous-classe.
- L'onglet « Data Properties » nous permet d'éditer toutes les propriétés.
- L'onglet « Individuals » nous permet d'éditer les instances des classes.

5.4.3 Création des classes

Une classe ou un concept représente un type d'objet dans l'univers. Les classes sont habituellement organisées en taxinomies auxquelles on applique des mécanismes héritage comme nous terminons la création de projet, tout suit il faut commencer par la création des classes de l'ontologie, nous allons créer les classes et les sous-classes de notre ontologie comme suit :

- Activer l'onglet « Classes ».
- Dans la fenêtre qui s'affichera, sélectionner la classe mère « Thing » dans la vue « class hierarchy » qui est la racine de toutes les classes que nous allons créer.
- Cliquer sur le Bouton « AddSubClasse » pour créer les sous classes de la classe globale

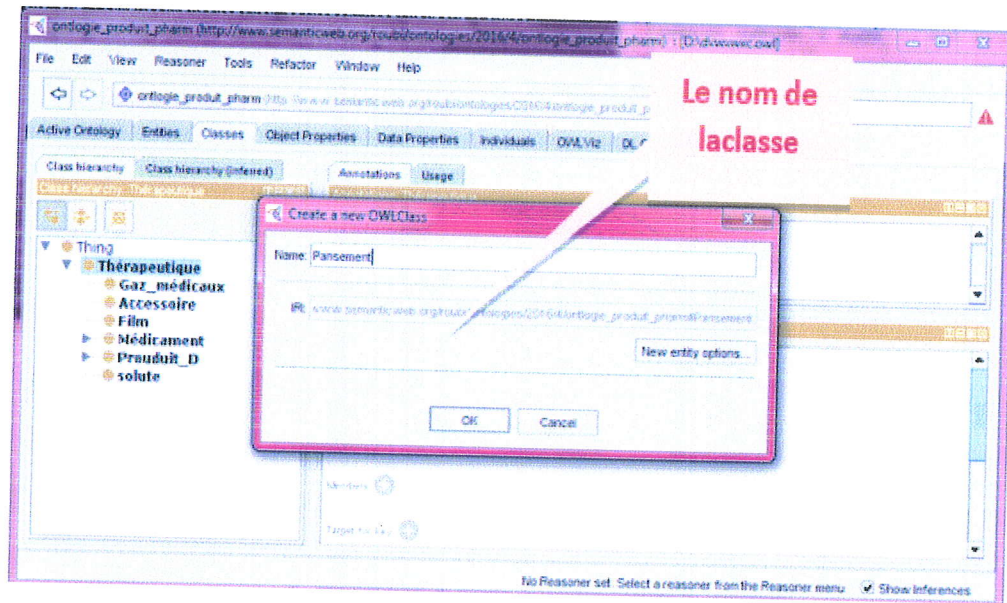


FIGURE 5.3 – Création des classes

- La création de la classe / Thérapeutique est SubClassde « Thing »
- La création des classes : Médicament, Accessoire, Produit_D, Pansement, Solute, Filme, Gaz médicaux sont SubClass de thérapeutique
- La même chose pour la création des sous classe de Médicament et de Produit_D Mettre les classes "disjoint" les unes des autres comme le montre la figure suivante pour éviter les chevauchements entre les classes et les sous-classes (par exemple les individus d'une classe ne peuvent être une instance d'une classe afin que nous puissions éviter l'héritage multiple)

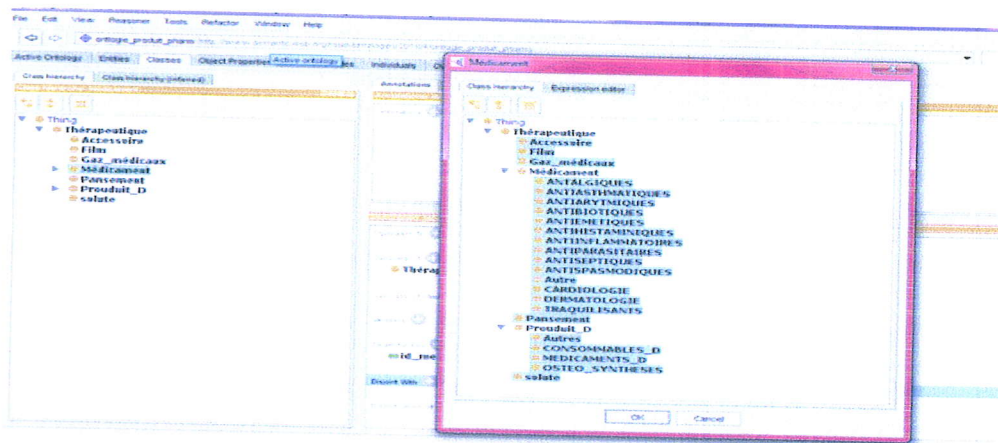


FIGURE 5.4 – Mettre les classes "disjoint" les unes des autres

5.4.4 Création des relations

Comme nous le savons, les Object Properties décrivent les relations entre deux instances (individus). Ils relient les individus d'un Domaine aux individus d'un Range. Une fois les classes créées, il faut définir leurs propriétés (ou Slots).

Nous allons créer les relations pour notre ontologie comme suit :

- Activer l'onglet " Object Properties".
- Cliquer sur le bouton "addsubproperty" pour créer une nouvelle propriété de données nommée par est_un (Noter que tous les Object Properties sont des sous-propriétés de topDataProperty), comme il est montré dans la figure suivante :

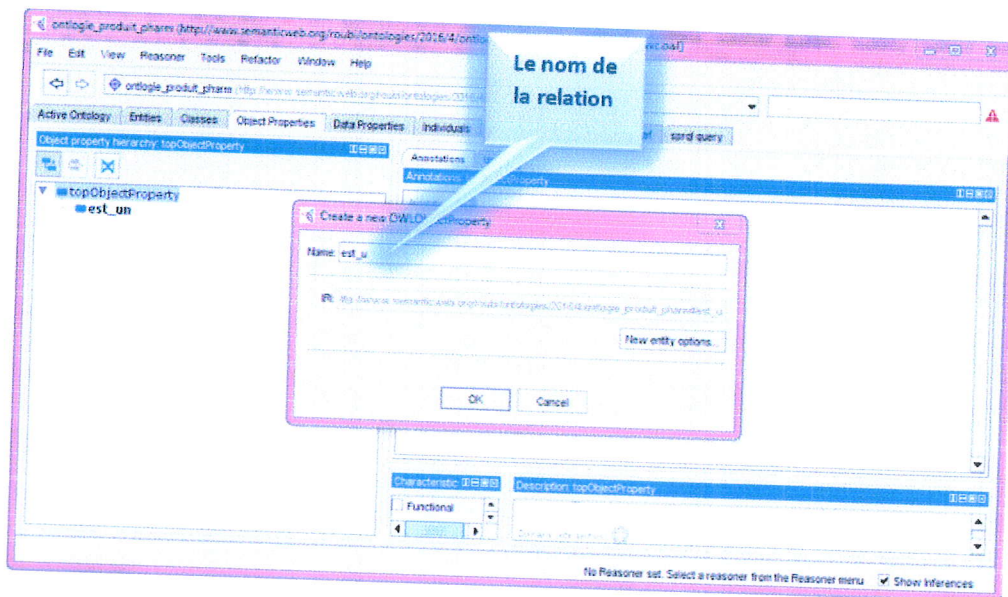


FIGURE 5.5 – Création des relations

Lorsque nous créons le nouveau «objectproperty » est_un nous devons sélectionner Médicament, Produit_D, Solute, Réactif, Antiseptique, Film, accessoire comme Domain et Thérapeutique comme Range. La figure suivante résume tout ça :

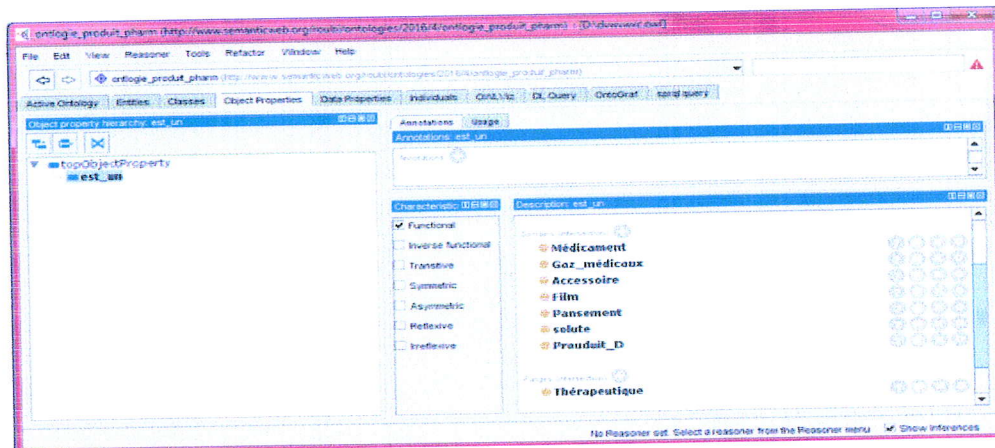


FIGURE 5.6 – Création de nouveau «objectproperty » est_un

5.4.5 Création des Propriétés

Précédemment nous avons mentionné que les «ObjectProperties » décrivent les relations entre deux instances (individus) et les «DataProperties » décrivent les relations entre deux instances et les valeurs de données. Nous allons ajouter «DataProperties » à notre ontologie comme suite :

- Activer l'onglet « Data Properties ».
- Utiliser le Botton « addsubproperty » pour créer une nouvelle propriété de données (Noter que tous DataPropertie sont subproperties de topDataProperty)
- Créer une nouvelle propriété de données (Data Properties) id_med et sélectionner Médicament comme Domain, String comme Ranges,comme il est montré dans la figure suivante :

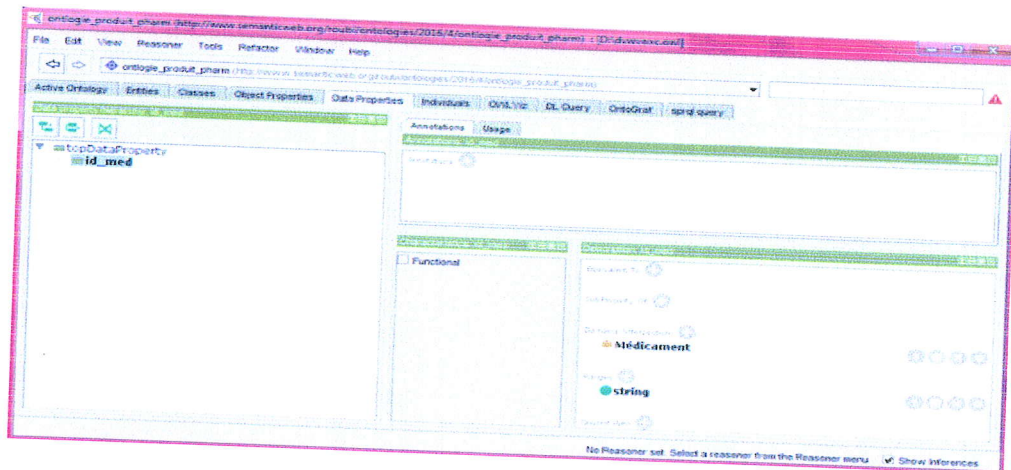


FIGURE 5.7 – ajouter «DataProperties » à notre ontologie

5.4.6 Saisie des individus (Instances)

Les instances sont utilisées pour représenter des éléments dans un domaine.

- Passer à l'onglet « individuals »
- Sélectionner la classe dans le volet gauche du projet
- Utiliser le Bouton « addindividual » pour créer un nouvel individu
- Cliquer dans le volet droit sur Type et choisir la classe qui inclut cet individu, comme il est montré dans la figure suivante :

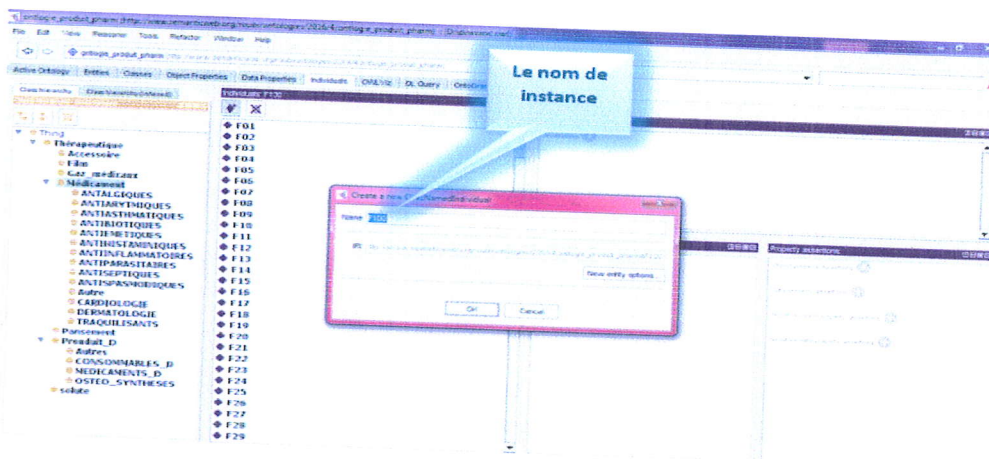


FIGURE 5.8 – insertion des individus

Après avoir ajouté chaque individu (instance), nous l'associons à la classe à laquelle il appartient on utilise la relation est_un D'une part et avec les individus similaires on utilise la relation SameIndividual As d'autre part, la figure suivante illustre ça : Après avoir fini la réalisation, on peut afficher le graphe ontologie, la figure suivante représente une partie de ce dernier :

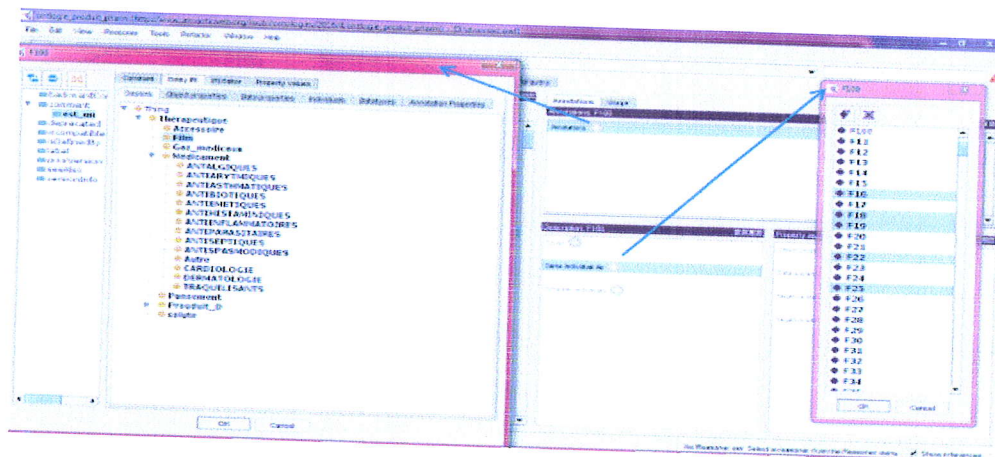


FIGURE 5.9 – association des individus à la classe à laquelle il appartient

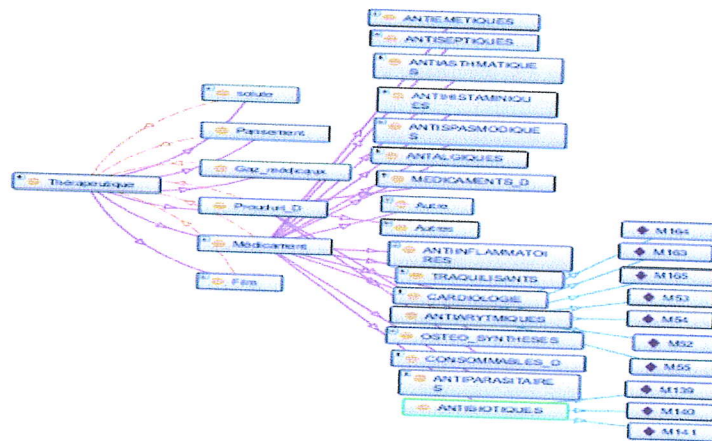


FIGURE 5.10 – graphe ontologie

5.5 Présentation des outils utilisés pour l'interface et le reporting

5.5.1 Développement des interfaces

-Choix du langage de programmation

Pour le choix du langage de programmation de notre système, nous avons choisi les langages :

JAVA : pour de nombreuses raisons : JAVA est un langage orienté objet simple, qui réduit le risque des erreurs et d'incohérences. Il est indépendant de toute plate-forme, il est possible d'exécuter des programmes JAVA sur tous les environnements qui possèdent une Java Virtual Machine (JVM). Il est doté d'une riche bibliothèque de classes, comprenant la gestion des interfaces graphiques (fenêtres, menus, graphismes, boîtes de dialogue, contrôles) Simple, Robuste, Dynamique et Sécurisé. Il permet d'accéder d'une manière simple aux fichiers et aux réseaux (notamment Internet). Il est caractérisé

aussi par la réutilisation de son code ainsi que la simplicité de sa mise en oeuvre. Il est compatible avec L'API JENA, ce qui nous permet la manipulation, le parcours et la modélisation des documents OWL.

-NetBeans

NetBeans est un environnement de développement permet aux programmeurs d'écrire, compiler, déboguer et déployer des programmes. Il est écrit en Java - mais peut supporter n'importe quel langage de programmation.

-**JENA** est un environnement de travail open source en Java, pour la construction d'application web sémantique. JENA permet de manipuler des documents RDF, RDFS, OWL et SPARQL. Il fournit un moteur d'inférences permettant des raisonnements sur les ontologies. JENA est maintenant sous Apache Software Licence.

- **OJDBC7** Afin que NetBeans puisse accéder aux données de base de données oracle, il faut faire une connexion entre les deux on utilise le jar OJDBC7.

-**JfreeChart** Est une bibliothèque open source disponible pour Java qui permet aux utilisateurs de générer facilement des graphiques.

-**WEKA** L'intégration du Weka dans NetBeans nécessite l'ajout de la bibliothèque weka.jar au projet NetBeans afin d'utiliser l'un de ses algorithmes; dans notre cas les algorithmes du linéaire régression et du SMO reg.

5.5.2 Exécution de l'application

Le but général de notre application est de visualiser la consommation (quantité consommée) des produits pharmaceutiques au niveau de EPSP de Jijel dans une période spécifique et de prédire la consommation au futur. Dans cette section nous présentons quelques captures d'écran d'exécution de notre application

◆ Interface Principale

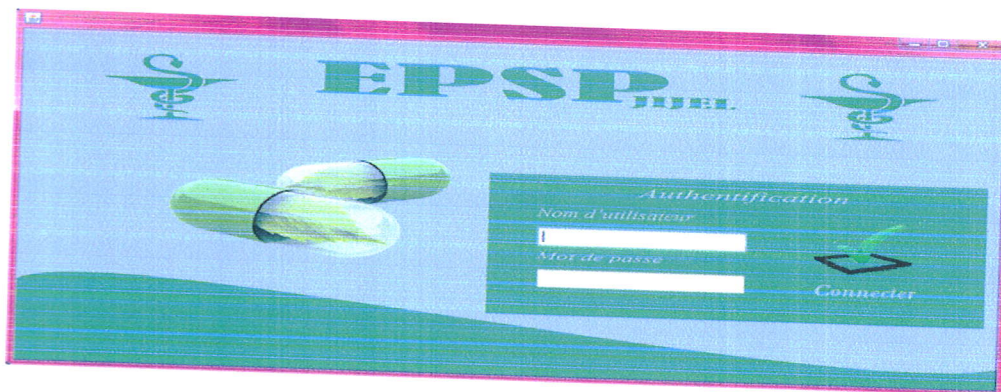


FIGURE 5.11 – Fenêtre d'Authentification

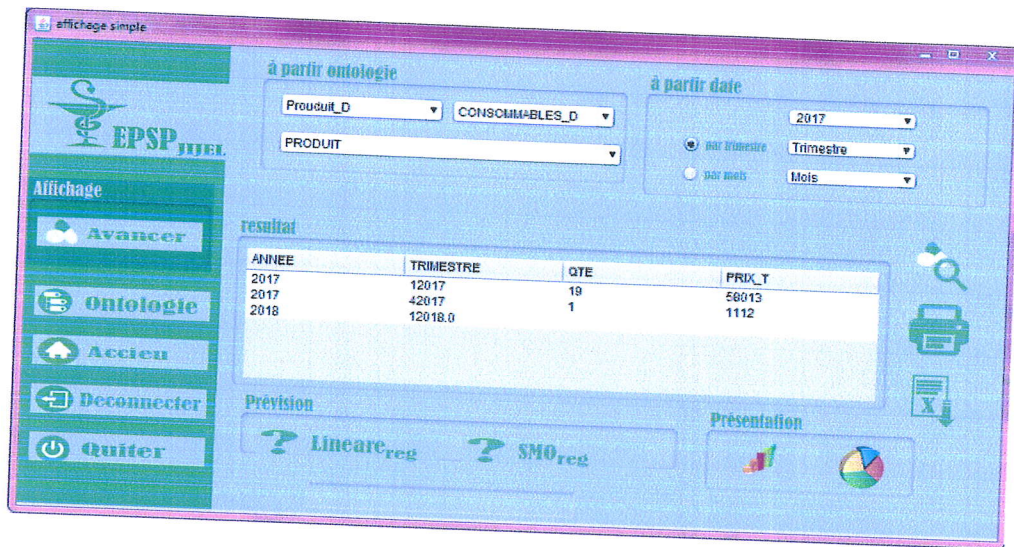


FIGURE 5.14 – consommation en 2017 par mois/classe

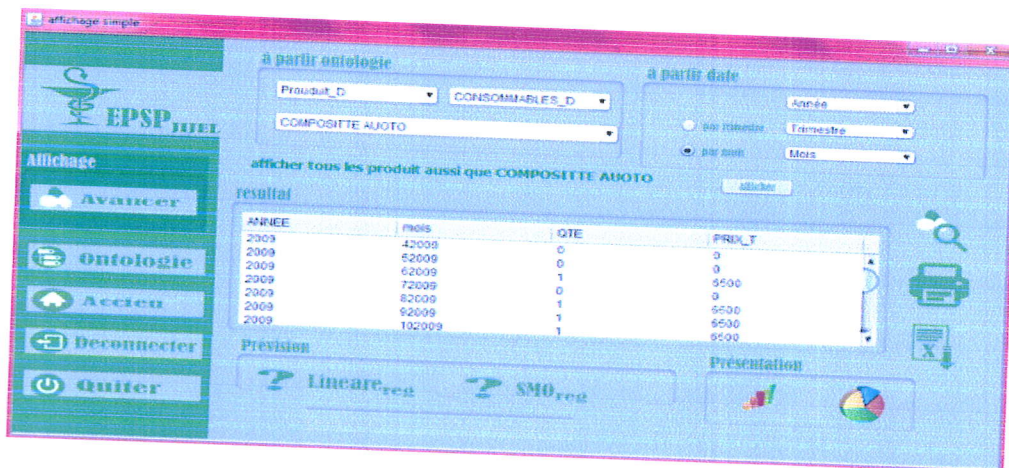


FIGURE 5.15 – consommation en 2017 par mois/produit

Noter que : la recherche dans le dernier exemple n'est pas simple tel que la quantité affichée et la somme des quantités de tous les produits similaires au produit composite auto, cette opération se fait à partir de la relation sameindividual_{as} d'ontologie. Si l'utilisateur veut voir ces produit il faut juste cliquer sur le button afficher, la fenêtre Suivante s'affiche :

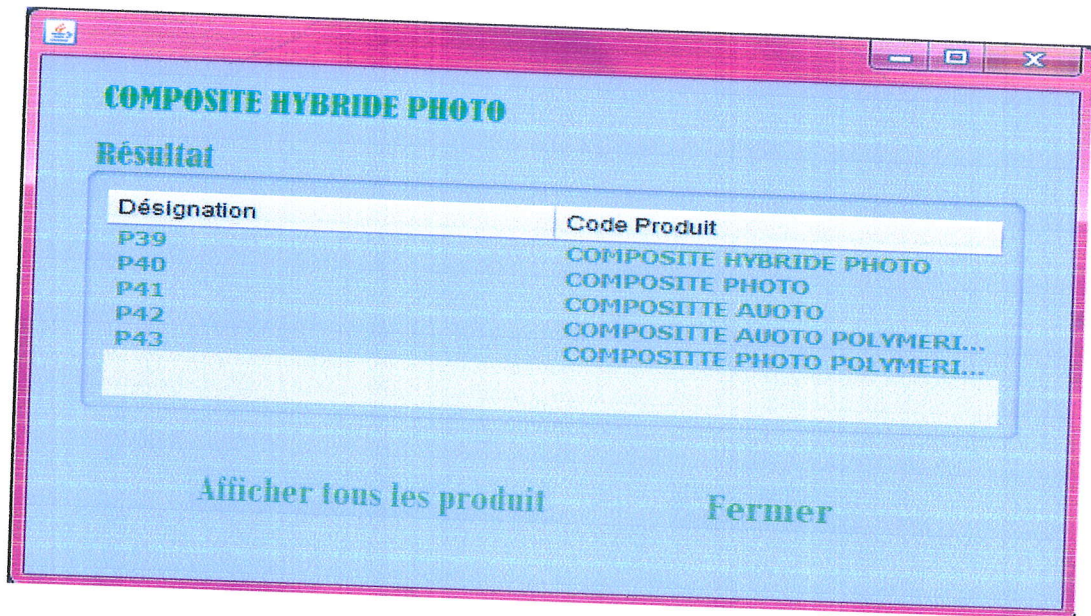


FIGURE 5.16 – Les produits composite auto

❖ Affichage avancer

Cette interface est aussi composée de trois onglets pour la recherche : l'onglet « à partir ontologie » et l'onglet « à partir période » et l'onglet « résultat ».

- Le premier onglet permet à l'utilisateur de choisir la classe, la sous classe et/ou le produit sur lequel il veut effectuer sa recherche.

- Le deuxième onglet permet à l'utilisateur de faire sa recherche dans une période précisée, entre deux années, deux trimestres ou entre deux mois.

-Le troisième onglet permet à l'utilisateur de visualiser le résultat de la recherche.

La fenêtre suivante illustre un exemple de consommation du sous class (Consommable_D) entre Les deux trimestres : trimestre 1(2010) et trimestre 2 (2014).

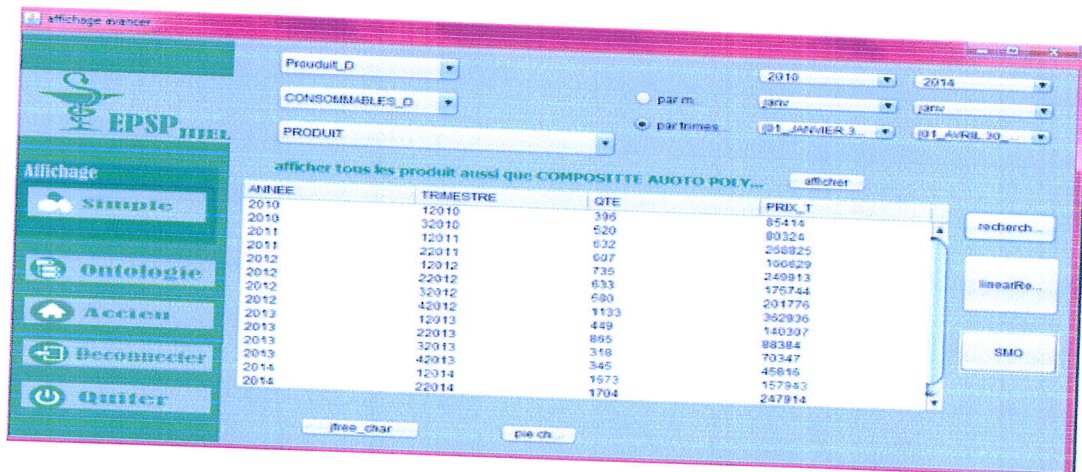


FIGURE 5.17 – consommation du produits de sous class (Consommable_D) entre trimestre 1(2010) et trimestre 2 (2014).

- Les deux fenêtres, affichage simple et affichage avancé contiennent aussi :
- Un bouton qui permet d'imprimer le résultat de la recherche.
 - Un bouton qui permet d'exporter le résultat de la recherche sous un format EXCEL.
 - Deux boutons qui permettent l'affichage sous forme graphique du résultat obtenu pour la quantité (histogramme, diagramme de secteur)

5.6 La prévision :

Les deux buttons de l'onglet prévision permettent aux utilisateurs de faire une prévision en utilisant les deux méthodes linéaire régression et smo reg à base des données affichées dans l'onglet résultat.

Les deux figures suivantes illustrent les résultats d'application des deux méthodes sur l'exemple précédent.

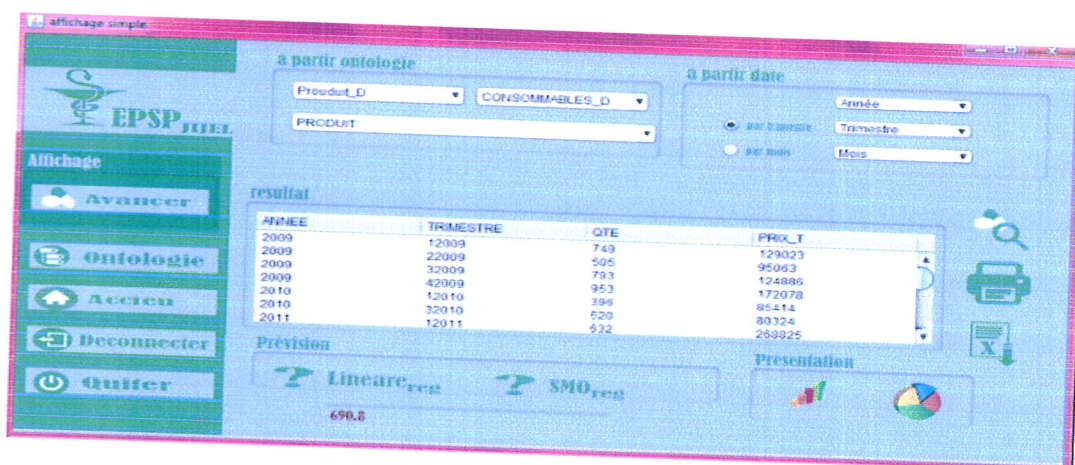


FIGURE 5.18 – la méthode Linéaire Régression .

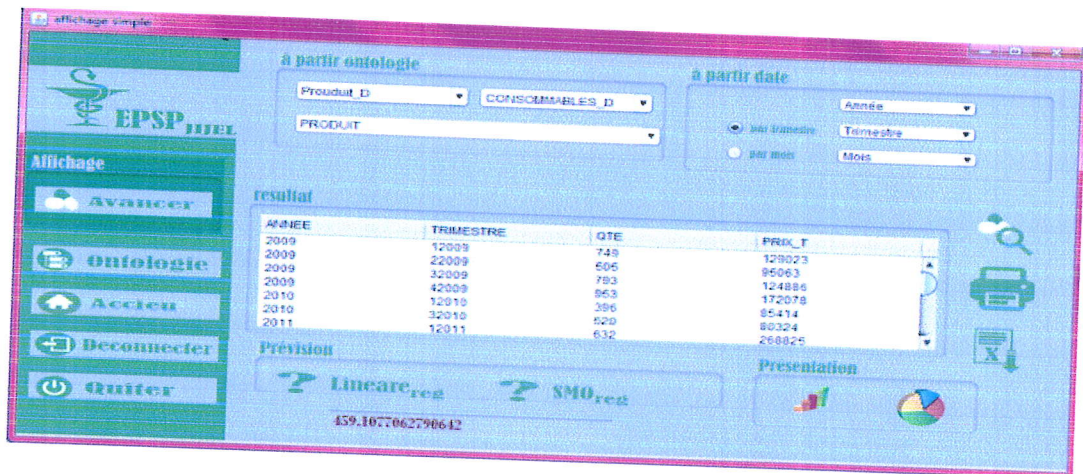


FIGURE 5.19 – Smo Reg.

Ce résultat peut être affiché sous l'une des formes graphiques suivantes :

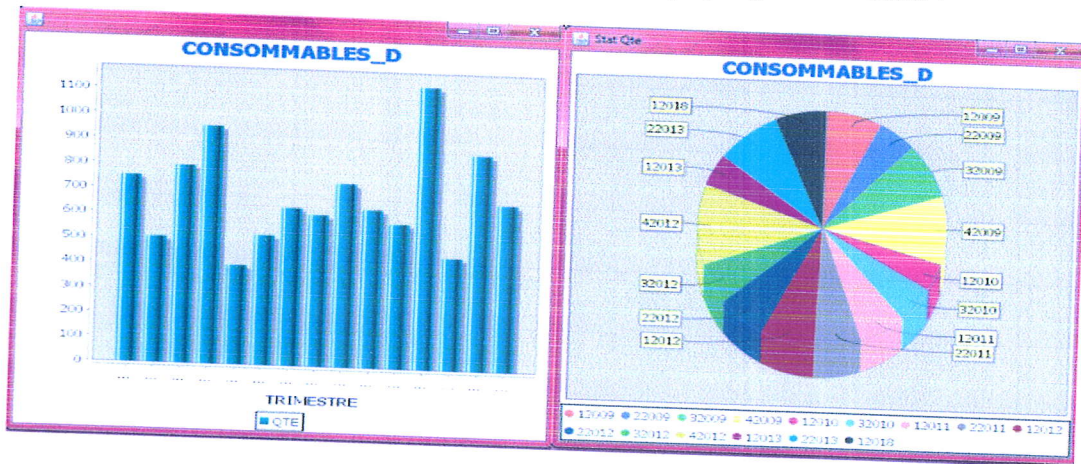


FIGURE 5.20 – Représentation des résultats sous forme histogramme et diagramme de secteur .

5.7 Discussion des résultats de prevision

Les deux techniques; la régression linéaire et SMO reg ont été appliquées sur les 7 dernières années de consommation des produits pharmaceutiques, que ce soit par mois, par trimestre ou par année.

Les résultats obtenus des deux techniques sont pratiquement les même dans le cas des données complètes, mais ils sont différents dans le cas des données incomplètes ou nulles (un mois qui maque par exemple). Dans ce cas de données incomplètes ou nulles, la technique de la régression linéaire donne le résultat de la prévision sans prendre en considération ces données manquantes ou nulles, par contre la technique du SMO REG fait des estimations des données manquantes avant de faire la prévision. Du moment que les données sont aléatoires et les résultats diffèrent d'une technique à l'autre, on ne peut pas dire qu'une technique

est meilleure par rapport à l'autre d'une manière absolue pour cela l'utilisateur est libre de choisir la méthode ou la technique la plus adéquate.

5.8 Conclusion

Ce chapitre a fait le point sur les éléments de base de l'implémentation de notre système, en commençant par l'implémentation d'une ontologie de classification des médicaments sous Protégé. Nous avons développé par la suite une application en Java avec NetBeans comme environnement de développement, Jena comme environnement d'intégration, le langage OWL avec protégé pour la représentation des ontologies, WEKA comme outil de prévision et Jfree-char pour l'édition des rapports.

L'application est sous forme d'une interface d'exécution qui permet de visualiser les données à travers des requêtes et de rendre les résultats sous forme graphique et tabulaire. Cette interface permet aussi le calcul de valeurs prévisionnelles concernant les besoins en médicaments par exécution d'une des deux techniques de prévision retenues, la régression linéaire et SMO Reg en se basant sur les données historiques déjà préparées par notre soin.

Conclusion générale

L'objectif principal de ce mémoire était de concevoir un système décisionnel d'aide à la prévision des besoins en produits pharmaceutiques au niveau d'un établissement de santé.

Nous avons ciblé d'une façon précise l'établissement public de santé de proximité (EPSP) de Jijel. Ce système représente un ensemble de moyens, outils et méthodes qui permettent de collecter et présenter les données des différents services opérationnels en vue d'offrir une aide à la décision.

Notre thème de recherche concerne donc la contribution dans le monde de la santé publique à travers deux volets complémentaires :

- ✓ Le premier est un volet d'intégration de données pharmaceutiques avec la collection des données par un étude de ce qui existe sur le terrain et la préparation de ces données par la conception et la construction d'une ontologie reflétant parfaitement les besoins exprimés par les utilisateurs.
- ✓ Le deuxième est un volet d'exploitation guidé principalement par les besoins exprimés par les utilisateurs de l'organisme concerné. Cette exploitation a pris dans notre système deux formes distinguées :
 - Une exploitation sous forme de rapports tabulaires ou graphiques.
 - Une exploitation par application de la technique de prédiction, linéaire régression et SMO REG afin de prédire les besoins futurs en produits pharmaceutiques selon l'historique de consommation.

Nous avons procédé au niveau de notre travail de la manière suivante :

- ✧ La collection des données concernant la consommation des produits pharmaceutiques des années précédentes au niveau de l'Établissement Public de Santé de Proximité de Jijel.
- ✧ La préparation des données collectées par la conception d'une ontologie : On a créé une ontologie des médicaments à travers Protégé. Cette ontologie nous a aidée à classer les médicaments selon une vision sémantique des données reflétant les besoins des utilisateurs. Ceci nous a facilité énormément la recherche et l'analyse des médicaments par classe.

- ❖ L'exploitation des données préparées dans un but de reporting par mise à disposition de l'utilisateur d'un ensemble varié et paramétré de requêtes qui permettent la visualisation des données chargées dans une base de données sous format tabulaire ou graphique afin de présenter une meilleure vision des données pour les utilisateurs.
- ❖ L'exploitation des données préparées dans un but de prévision par application des techniques les plus adéquates de fouille de données pour la prédiction des besoins en produits pharmaceutique selon la nature des données disponibles. Dans notre cas nous avons opté les deux techniques ; la régression linéaire et SMO REG qui donnent pratiquement les mêmes résultats dans le cas des données complètes. Dans le cas des données incomplètes ou nulles (un mois qui manque par exemple), la technique de la régression linéaire donne le résultat de la prévision sans prendre en considération ces données manquantes ou nulles, par contre la technique du SMO REG fait des estimations des données manquantes avant de faire la prévision.

Comme perspectives à ce projet nous proposons la généralisation de la notion d'ontologie à tous les produits pharmaceutiques en plus des médicaments que nous avons traités.

Bibliographie

- [1] S. Al-Osimi M. Al-saleememail, n. al-kathiry. mining educational data to predict students academic performance.
- [2] R.Rakotomalala D.A.Zighed. Extraction des connaissance a partir des données (ecd). 2002.
- [3] Stéphane Turffy. Cours de data mining,master 2 ingenierie economique et financiere,7 fevrier 2014.
- [4] lamiche Chaabane. fusion et fouille de données guidées par les connaissance : application à l'analyse d'image ,thèse doctorat en sciences.
- [5] CHAMI Djazia. Une plate forme orientée agent pour le data mining, 2009-2010.
- [6] Odilon Irakoze. Que ce que le data mining ? ,2016.
- [7] S.Mohammed. Systeme daide à la distribution des produits pharmaceutiques en algerie en utilisant le data mining, 2010.
- [8] Jérôme Azé. Extraction de connaissances à partir de donnees numériques et textuelles,thèse de l'université parissud.
- [9] . Mme bouchekouf Asma. Perception du comportement de l'apprenant dans un environnement d'apprentissage.mémoire : Présenté en vue de l'obtention du diplôme de magister en informatique, université badji mokhtar-annaba 2013.
- [10] État de l'art sur les aspects méthodologiques et processus en Knowledge Discovery in Database . Cristina oprean.
- [11] Andrew Kusiak Bruno Agard. Exploration des bases de données industrielles à l à aide du data mining â“ perspectives. in 9ème colloque national aip primeca.
- [12] C.Duby S.Robin J.Daudin. Analyse de séries chronologiques. 1996.[18] o.grimaud j.pommier. les fonctions essentielles de santé publique : Histoire, définition et applications possibles. 2007.
- [13] S. Robin P. Trécourt J.J. Daudin, C. Duby. Analyse de séries chronologiques.

- [14] Amirouche Amine and MR.Achouche Abdessalam. Conception et réalisation dâ™un système d'aide à la prévision des besoins en produits pharmaceutique au niveau d'un établissement de sante, 2015.
- [15] <https://fr.wikiversity.org/wiki/Introduction>.
- [16] Catherine Pardoux Bernard GOLDFARB. Prvision à court terme : méthodes de lissage exponentiel.
- [17] <https://WWW.etudier.com/dissertations/445670.html>.
- [18] <https://tel.archives.ouvertes.fr/tel-00879853/Document>.
- [19] https://thesis.univ-biskra.dz/843/1/Inf_m1_2011.pdf.
- [20] https://WWW.academia.edu/4388491/Cours_Types_Ontologies.
- [21] Souheila KHALFI. construction d'une ontology pour la prise en charge des patients à domicile,mémoire de magister.
- [22] Natalya F. Noy et Deborah L. McGuinness. Développement d'une ontology 101 : Guide pour la création de votre première ontology.
- [23] Belabbes fatima. Conception et implémentation d'une ontology médicale cas : insuffisance cardiaquesss ,mémoire de fin d études pour l'obtention du diplôme de master en informatique 2014-2015.
- [24] Fatima Zohra BERRADANE et Meriem KARAOUZENE. La création d'une ontology pharmaceutique.
- [25] https://WWW.technololange.net/imprimer.php3?id_article=280.
- [26] https://WWW.irit.fr/journal-i3/serie/annee2006:REVUE_i3_01_07;PDF.
- [27] Pr. Z.Boufaïda. construction d'une ontology pour la notation des cvc offre d'emploi.
- [28] <https://WWW.lsis.org./espinaseb/these/1.elsarraj-2014.pdf>.
- [29] <https://WWW.bu.ume.edu.dz/these/informatique/DJA5685.pdf>.
- [30] Flake lawrence. Article.efficience svm regression training 2002.pdf.



Résumé

L'objectif principal de ce travail est de concevoir un système d'aide à la prévision des besoins en produit pharmaceutique au niveau de l'établissement de santé de proximité (EPSP) de Jijel. Il s'inscrit donc dans le cadre de la prédiction des données pharmaceutiques guidée par une ontologie.

Il consiste à préparer les données avec la classification des médicaments selon une vision sémantique des données reflétant les besoins d'utilisateur à travers une ontologie. L'utilisateur peut procéder à une interrogation de cette ontologie par un ensemble varié et paramétré de requêtes qui permettent la visualisation des données chargées dans une base de données et ceci sous forme tabulaire ou graphique. L'utilisateur peut aussi lancer des calculs de prévision des besoins futurs en produits pharmaceutiques sur la base de l'historique de consommation avec application des techniques LINEARE REGRESSION et SMO REG.

Tout cela est réalisé avec une interface développée en java avec l'utilisation de protégé pour la création d'ontologie et WEKA pour la prévision.

ملخص

الهدف الرئيسي لهذا العمل هو تصميم نظام للمساعدة على التنبؤ من اجل تلبية مختلف الاحتياجات الصيدلانية للمؤسسة العمومية للصحة الجوارية جيجل, وذلك بالاعتماد على علم الانثولوجيا الذي يعمل على تصنيف المنتجات الصيدلانية وفق مجال استعمالها ,

المستعمل يمكنه رؤية مختلف المعطيات المخزنة في قاعدة البيانات و عرضها اما في جداول او على شكل بيانات عن طريق الاستعلام من الانثولوجيا , كما يمكنه القيام بعمليات حسابية من اجل التنبؤ بالاحتياجات المستقبلية للمنتجات الصيدلانية عن طريق تطبيق

تقنيتي *Régression Linéaire* , *SMO reg*

وقد حقق هذا عن طريق واجهة التي قمنا بتطويرها باستعمال *Protégé* و *Java*

من اجل انشاء الانثولوجيا و *Weka* للتنبؤ