



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA

Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie

Département d'informatique

N° d'ordre :SIOD17/M2/2021

Mémoire

Présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : Système d'Information Optimisation et Décision (SIOD)

Titre La réalisation d'un système de data mining pour le domaine pharmaceutique

Par : Yakoub Sana

Soutenu le Juin 2021 devant le jury composé de :

Mme. Djazia CHAMI

MAA

Président

Rapporteur

Examineur

Année universitaire 2020-2021

Dédicaces

Mes parents sont la joie de ma vie,

mes frères,

et toute ma famille,

Et tous mes chers professeurs et amis,

je dédie ce modeste travail.

Remerciements

Tout d'abord je remercie le grand ALLAH de m'avoir donné la force et la puissance pour finir ce travail,

Je souhaite exprimer ma sincère gratitude à *Pr Djazia CHAMI*, ma superviseuse pour son aide, ses conseils et ses encouragements depuis le premier jour jusqu'à maintenant ; sans vous je ne pourrais pas présenter ce travail.

J'en profite pour remercier tous les professeurs qui m'ont accompagné tout au long de mon parcours.

Ma gratitude ultime à ma grande et petite famille, mes parents, mes sœurs et frères, et tous ceux qui m'ont aidé de près ou de loin.

Résumé :

Le Data Mining est une nouvelle technologie récemment apparue qui vise à analyser des données et à extraire des connaissances à partir d'une grande quantité de données mixtes. Par conséquent, nous trouvons de nombreuses entreprises et organisations souffrant d'une surcharge et d'une mauvaise utilisation des données. L'exploration de données est devenue un outil important qui augmente la rentabilité des entreprises, et pour y parvenir, nous avons abordé l'exploitation minière dans le domaine de la pharmacie, car nous voyons au fil du temps une augmentation continue des médicaments et la souffrance du pharmacien dans l'organisation et la classification des médicaments, car nous essayé dans ce projet de classer les médicaments de manière particulière selon leur catégorie thérapeutique .

Mots clés : data mining , domaine pharmaceutique.

الملخص

التنقيب عن البيانات هو تقنية جديدة ظهرت مؤخرًا تهدف إلى تحليل البيانات واستخراج المعرفة من كمية كبيرة من البيانات المختلطة. لذلك نجد العديد من الشركات والمؤسسات تعاني من التحميل الزائد للبيانات وسوء الاستخدام. أصبح التنقيب عن البيانات أداة مهمة تزيد من ربحية الشركات ، ولتحقيق ذلك اقتربنا من التعدين في مجال الصيدلة ، حيث نرى بمرور الوقت زيادة مستمرة في الأدوية ومعاونة الصيدلي في تنظيم وتصنيف الأدوية حيث حاولنا في مشروعنا هذا إلى تصنيف الأدوية بطريقة خاصة على حسب الفئة العلاجية لها.

الكلمات المفتاحية : التنقيب عن البيانات ، المجال الصيدلاني.

Table des matières

Table des figures	IV
Liste des tableaux	VI
Introduction générale	1
1 Généralités Sur Le Data Mining	2
1.1 Introduction	3
1.2 Définition du data mining	3
1.3 Motivations	4
1.4 types de données Data mining	5
1.4.1 Les fichiers plats	5
1.4.2 Les bases de données relationnelles	6
1.4.3 Les Data Warehouses (entrepôt)	6
1.5 Les taches du data mining	6
1.5.1 La Classification	6
1.5.2 L'estimation	7
1.5.3 Le groupement par similitude	7
1.5.4 L'analyse des clusters	7
1.5.5 La description	8
1.6 Processus du data mining	8
1.6.1 Définition et compréhension du problème	9
1.6.2 Collecte des données	9
1.6.3 Prétraitement	9
1.6.4 Estimation du modèle	9
1.6.5 Interprétation du modèle et établissement des conclusions	9
1.7 Techniques du data mining	10
1.7.1 Réseaux de neurones	10
1.7.2 Les arbres de décision	11
1.7.3 Les algorithmes génétiques	11
1.7.4 Les règles associatives	12

1.7.5	L'algorithme des k-Plus proches voisins	13
1.7.6	Algorithme de clustering par K-Means	14
1.7.7	Machines à vecteur support	15
1.8	Catégorisation des systèmes du data mining	18
1.9	Domaines d'application du data mining	18
1.10	Conclusion	20
2	Une étude sur le domaine de la pharmacie	21
2.1	introduction	22
2.2	definition le domaine pharmaceutique	22
2.2.1	Majors en pharmacie	22
2.3	Le médicament	23
2.3.1	Définition	23
2.3.2	Type de Médicament	23
2.3.3	Composition du médicament	24
2.3.4	Les voies d'administration d'un médicament	25
2.3.5	Médicaments dangereux	27
2.3.6	Classification des médicaments	28
2.3.6.1	Classification clinique (therapeutic)	28
2.3.6.2	Classification pharmacologique	29
2.3.6.3	Classification chimique	29
2.3.6.4	Classification légale (listes de médicaments et risque de ter- nissement)	30
2.4	Travaux Connexes	32
2.4.1	Les objectifs ATC	32
2.4.2	La Structure du système ATC	33
2.4.2.1	Le principes de classification du système ATC	33
2.4.3	Classification des antibiotiques	35
2.4.4	Origine	35
2.4.5	Mode d'action	35
2.4.5.1	Spectre d'activité	35
2.4.5.2	Nature chimique	36
2.4.5.3	Expression de l'activité antibactérienne	36
2.5	Conclusion	36
3	Conception	37
3.1	Introduction	37
3.2	Architecture globale	37
3.3	Architecture détaillée	38
3.3.1	Pré-traitement	39

3.3.2	Apprentissage	40
3.4	Fonctionnement du système	42
3.4.1	Diagramme de classes	42
3.4.2	Diagramme de Séquence	44
3.4.3	diagramme de cas d'utilisation	45
3.5	Conclusion	45
4	Implémentation	46
4.1	Introduction	46
4.2	Environnement de développement	46
4.3	Bibliothèques utilisées	48
4.4	Datase	51
4.5	Application	55
4.6	Conclusion	63
	Conclusion Générale	64
	Bibliography	64

Table des figures

1.1	L'évolution de la technologie du système BD	4
1.2	Riche en données mais pauvre en informations.	5
1.3	Architecture du DW	6
1.4	Processus de data mining (CRISP-DM)	8
1.5	Architecture d'un réseau de neurones artificiel.	10
1.6	Un exemple représentant K-PPV	14
1.7	SVMs binaires	16
1.8	SVMs multiclasse	17
2.1	La méthode de développement de médicaments	23
3.1	Architecture globale	38
3.2	Pré-traitement de la basse	39
3.3	Encodage de la source	39
3.4	Classification des médicaments utilisant SVM	40
3.5	Utilisation	42
3.6	Diagramme de classes UML pour l'application	43
3.7	diagramme de Séquence UML pour l'application	44
3.8	diagramme de cas d'utilisation UML pour l'application	45
4.1	Navigateur Anaconda	47
4.2	Anaconda	47
4.3	sklearn avec algorithme	49
4.4	bibliothèque SVM	49
4.5	bibliothèque graphique	50
4.6	bibliothèque Tkinter	51
4.7	Base de Données.csv	51
4.8	diagramme de catégories	52
4.9	Catégorie de médicament	52
4.10	algorithme de codage de base	53
4.11	resultat du codage entraînement	54
4.12	resultat du codage et validation	54

4.13 Précision du modèle	55
4.14 L'interface principale de l'application	55
4.15 Première entrée utilisateur	56
4.16 Première entrée utilisateur	56
4.17 Explication du taux d'erreur	57

Liste des tableaux

2.1	Catégories de produits pharmaceutiques.	28
2.2	Classification therapeutique	29
2.3	Classification pharmacologique	29
2.4	Classification chimique	30
2.5	Classification légale	31
2.6	Classification avec risque de ternissement	32
2.7	Les 5 niveaux différents des médicaments	34
2.8	Antibiotiques bactériostatiques et bactéricides.	36

Introduction générale

L'augmentation et l'accumulation constantes d'informations dans divers domaines sont devenues un obstacle majeur à l'extraction d'informations, et c'est ce qui a incité la technologie à développer des technologies informatiques et de stockage et à trouver une solution grâce à l'exploration de données pour extraire des connaissances utiles.

À la fin des années 1990, le terme de data mining est apparu, qui vise à extraire en grand nombre des connaissances cachées dans l'entrepôt de données d'une institution ou d'un domaine spécifique (science, médecine, pharmacie, santé publique ...). [1]

Au fil du temps, des maladies nouvelles et différentes apparaissent qui conduisent à l'émergence de nouveaux médicaments dont nous parlerons dans le deuxième chapitre, et c'est ce qui rend le pharmacien difficile dans la gestion, la gestion et la régulation des médicaments, et dans une partie de cette étude nous allons essayer de trouver une solution à ce problème en extrayant des bases de données pharmaceutiques.

L'objectif d'un projet est de préparer un système qui permet l'exploration de données en classant les médicaments selon leur catégorie thérapeutique. c'est pourquoi nous avons commencé notre thèse avec le premier chapitre qui parle du concept d'exploration de données en présentant les motivations les plus importantes qui ont conduit à son émergence et clarifier ses tâches et types de data mining. Ensuite, nous avons abordé les étapes du processus de data mining, avec les avantages et les inconvénients des algorithmes utilisés dans le data mining et ses classifications, et à la fin de ce chapitre nous avons clarifié les différents domaines de l'application d'exploration de données.

Le deuxième chapitre est consacré à parler du domaine de la pharmacie et se compose de deux parties. la première partie est consacrée à présenter quelques concepts et à expliquer les méthodes désignées pour classer les médicaments, et la deuxième partie nous avons parlé de quelques travaux utilisant la classification des médicaments pour extraire Les données.

Dans le troisième chapitre, nous avons donné une description générale de notre système, puis nous avons fourni une explication détaillée de chaque étape, en clarifiant les principales techniques et diagrammes qui expliquent notre système.

Le quatrième et dernier chapitre dans lequel nous avons présenté les outils, programmes, langages, bibliothèques et données utilisés dans notre système, puis nous avons fourni un exemple détaillé de l'application que nous avons obtenue .

Chapitre 1

Généralités Sur Le Data Mining

Sommaire

1.1	Introduction	3
1.2	Définition du data mining	3
1.3	Motivations	4
1.4	types de données Data mining	5
1.4.1	Les fichiers plats	5
1.4.2	Les bases de données relationnelles	6
1.4.3	Les Data Warehouses (entrepôt)	6
1.5	Les taches du data mining	6
1.5.1	La Classification	6
1.5.2	L'estimation	7
1.5.3	Le groupement par similitude	7
1.5.4	L'analyse des clusters	7
1.5.5	La description	8
1.6	Processus du data mining	8
1.6.1	Définition et compréhension du problème	9
1.6.2	Collecte des données	9
1.6.3	Prétraitement	9
1.6.4	Estimation du modèle	9
1.6.5	Interprétation du modèle et établissement des conclusions	9
1.7	Techniques du data mining	10
1.7.1	Réseaux de neurones	10
1.7.2	Les arbres de décision	11

1.7.3	Les algorithmes génétiques	11
1.7.4	Les règles associatives	12
1.7.5	L'algorithme des k-Plus proches voisins	13
1.7.6	Algorithme de clustering par K-Means	14
1.7.7	Machines à vecteur support	15
1.8	Catégorisation des systèmes du data mining	18
1.9	Domaines d'application du data mining	18
1.10	Conclusion	20

1.1 Introduction

Quand on pense à la prospection (mines), la première chose qui vient à l'esprit est la prospection d'or ou d'argent, dans le monde des données la question n'est pas très différente. Vous pouvez rechercher des données pour découvrir et extraire des ressources cachées et précieuses, mais les minéraux sont une ressource finie, alors que les données au contraire sont abondantes, en croissance constante et réutilisables, ce qui rend le processus plus difficile, Alors qu'est-ce que la data mining ?

1.2 Définition du data mining

La fouille de données est un domaine qui est apparu avec l'explosion des quantités d'informations stockées, avec le progrès important des vitesses de traitement et des supports de stockage. La fouille de données vise à découvrir, dans les grandes quantités de données, les informations précieuses qui peuvent aider à comprendre les données ou à prédire le comportement des données futures. [1]

Ils existent d'autres définitions :

- L'exploration de données est utilisée pour identifier des moyens d'extraire des informations à partir de très grandes bases de données, grâce à la reconnaissance de formes, aux statistiques, aux bases de données et à la visualisation . [2]
- Le Data Mining est un processus inductif, itératif et interactif dont l'objectif est la découverte de modèles de données valides, nouveaux, et compréhensibles dans de larges Bases de Données. [3]

1.3 Motivations

La surcharge d'informations est due à sa diversité croissante et continue et à la nécessité de les convertir en données significatives dans plusieurs domaines divers (santé, éducation, commerce, découverte scientifique ...) et c'est ce qui a conduit les chercheurs à s'intéresser au data mining .

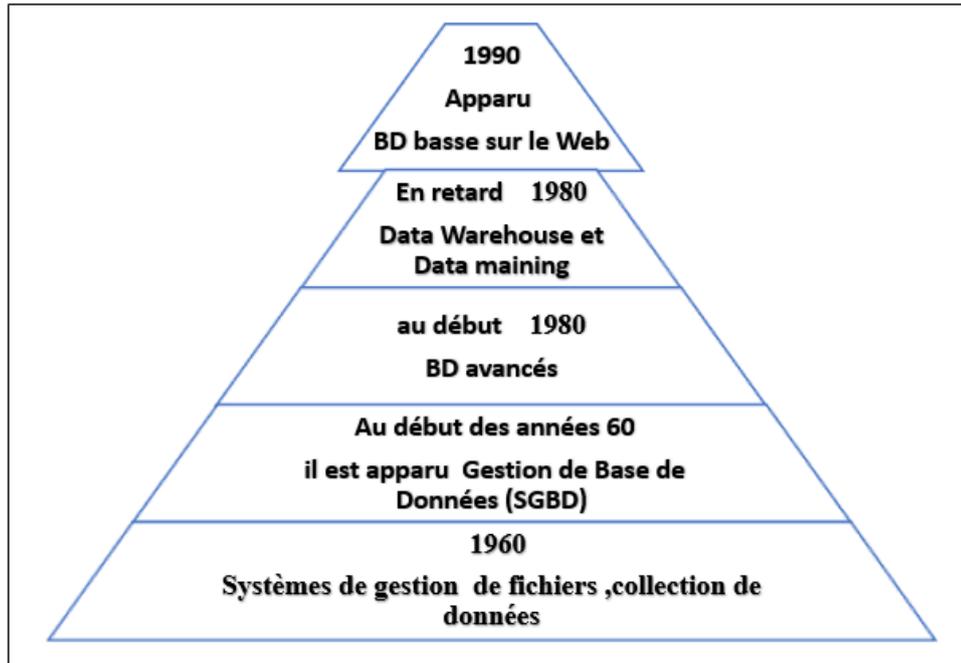


FIGURE 1.1 – L'évolution de la technologie du système BD

Le terme «data mining» est apparu pour la première fois au début des années 1960 et avait un sens péjoratif à l'époque. Certains chercheurs ont commencé à traiter des tableaux de données sur les expériences à leur disposition sans prioriser. Constatant que les résultats obtenus, loin d'être biaisés, étaient encourageants, ils ont été encouragés à réglementer cette approche. [4]

Vers la fin des années 1980, des chercheurs et des personnes intéressées par les bases de données, comme Rakesh Agrawal, ont commencé à travailler pour exploiter le contenu de grandes bases de données telles que les reçus de vente des supermarchés, convaincus qu'ils pouvaient ajouter de la valeur à ces blocs de données passives, et ce C'est ce qui a conduit à l'émergence des systèmes BD Les systèmes d'information hétérogènes et basés sur Internet tels que le World Wide Web (WWW) jouent un rôle essentiel dans l'industrie de l'information. [4]

Au fil du temps, des équipements informatiques se sont développés (stockage, ...) qui ont facilité la gestion, la récupération et l'analyse des informations pour les très grandes bases de données. La base de données la plus récente pour stocker des informations est l'entrepôt de données qui combine de nombreuses sources de données hétérogènes pour faciliter leur

gestion en un seul endroit grâce à un ensemble de techniques d'analyse (consolidation et agrégation) via OLAP (traitement analytique en ligne).[4]

mentation continue des données, le besoin et le désir de l'humanité de les comprendre et de les analyser sans utiliser d'outils auxiliaires figure 1.2 ont augmenté, mais cela a rendu la tâche plus difficile, de sorte qu'ils sont rarement exploités ou reconsidérés. De nombreux chercheurs ont recours à l'utilisation bases de données comme outils pour connaître les attentes d'un incident particulier à l'avenir.[4]



FIGURE 1.2 – Riche en données mais pauvre en informations.

1.4 types de données Data mining

IL peut être utilisé le data mining Dans plusieurs types de données, Cela ne dépend pas du type spécifique, Appliqué à data mining incluant les Bases de Données relationnelles et les Bases de Données Orientées-Objets, les data warehouses, les Bases de Données transactionnelles ... ect.

1.4.1 Les fichiers plats

Ce sont des fichiers de données au format texte ou binaire avec une structure qui peut être facilement extraite par des algorithmes d'exploration de données qui n'ont pas de relations les uns avec les autres et qui sont notamment utilisés dans le transfert de données entre différents serveurs.[4]

1.4.2 Les bases de données relationnelles

bases de données relationnelles c'est fondamentalement collecte de données organisées en tableaux avec des lignes et des colonnes, Et compter algorithmes du Data Mining appliqués sur des Bases de Données relationnelles sont plus polyvalents Et c'est ce qui fait Bénéficiez de SQL pour la sélection, la transformation et la consolidation.[5]

1.4.3 Les Data Warehouses (entrepôt)

Un entrepôt de données est une collection de données thématiques(orientées sujet), intégrées, non volatiles et historisées, conçu pour l'aide à une prise de décisions, Il existe trois types d'entrepôts de données : Entrepôt de données d'entreprise, Data Mart et Entrepôt virtuel.[6]

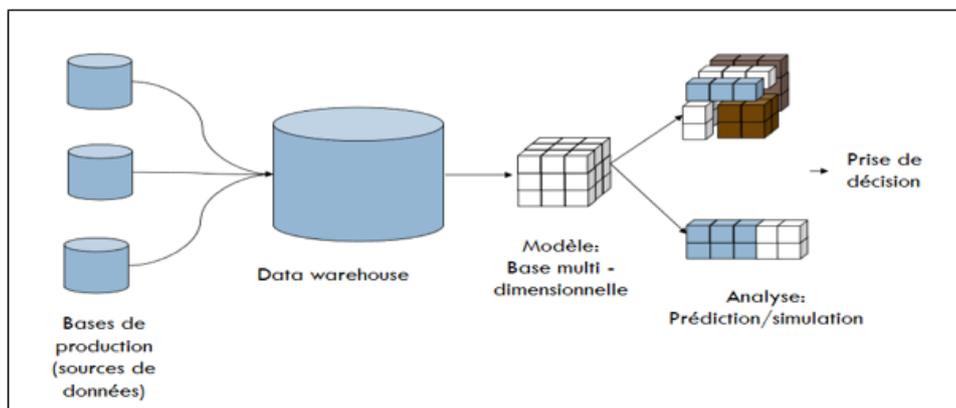


FIGURE 1.3 – Architecture du DW

La figure 1.3 montre la structure de l'entrepôt et c'est à travers pour extraction des données et transféré à DW Ce que nous concevons à travers modèle multi dimensionnelle (Forme de cube) Qui représentent les déterminants des données .

1.5 Les taches du data mining

Nous utilisons l'exploration de données pour résoudre de nombreux problèmes scientifiques et pratiques, et ces problèmes sont classés comme suit :

1.5.1 La Classification

La classification est une tâche très importante et plus utilisées dans l'analyse des bases de données, Elle permet d'apprendre des modèles de décision qui permettent de prédire le comportement des exemples futurs.[1]

On utilise classification pour classe des donne non classe pour créer un modèle, les objets

à classifiés stocké dans une base de données.

Les exemples des domaines recherche qui permette nous utilisation La classification : [4]

- classifiés la population qui peut bénéficier d'un logement social.
- éterminer les étudiants qui peuvent s'inscrire au master.

1.5.2 L'estimation

L'estimation est similaire à la classification à part que la variable de sortie est numérique plutôt que catégorique. En fonction des autres champs de l'enregistrement l'estimation consiste à compléter une valeur manquante dans un champ particulier. [1]

Les exemples des domaines recherche qui permette nous utilisation La L'estimation : [4]

- Tenter d'estimer l'étendue de l'apparition d'une maladie particulière chez un individu en particulier.
- Tenter d'estimer la mesure dans laquelle une covid 19 particulière apparaît chez un individu.

1.5.3 Le groupement par similitude

Le groupement par similitude (Analyse des associations et de motifs séquentiels) Regroupez les attributs similaires "vont ensemble". La tâche la plus Le plus fréquent dans le monde du business, est celle appelée l'analyse d'affinité ou l'analyse du panier du marché. Cela nous permet d'étudier la relation entre deux ou plusieurs attributs. [1]

Les exemples des domaines recherche qui permette nous utilisation La Le groupement par similitude :

- Découvrez quels étudiants peuvent s'inscrire à un doctorat.

1.5.4 L'analyse des clusters

Le clustering (ou la segmentation) est le regroupement d'enregistrements ou des observations en classes d'objets similaires Un cluster est une collection d'enregistrements similaires l'un à l'autre différence entre le clustering et la classification est que dans le clustering il n'ya pas de variables sortantes, La tâche de clustering ne classe pas, n'estime pas, ne prévoit pas la valeur d'une variable sortantes. Au lieu de ce la, les algorithmes de clustering visent à segmenter la totalité de données en des sous groupes relativement homogènes. Augmenter l'homogénéité au sein du groupe. [1]

Les exemples des domaines recherche qui permette nous utilisation La L'analyse des clusters :[4]

- Distinguer les élèves qui ont la même pensée.
- Regrouper les zones les plus vulnérables aux phénomènes naturels.
- Classification des aliments sains et malsains en fonction de leurs composants.

1.5.5 La description

C'est une description de ce qui se passe dans la base de données et l'étude de la relation qui existe entre eux. [1]

Les exemples des domaines recherche qui permette nous utilisation La description :

- les femmes supportent le changement plus que les hommes.

1.6 Processus du data mining

L'exploration de données ne dépend pas seulement de la résolution du problème et de la création du modèle, mais fait plutôt partie d'un ensemble d'étapes développées par les chercheurs, qui se compose de cinq étapes selon la norme CRISP-DM (Cross-Industry Standard Process for Data Mining) .[1]

Comme indiqué dans Figure 1.4 :

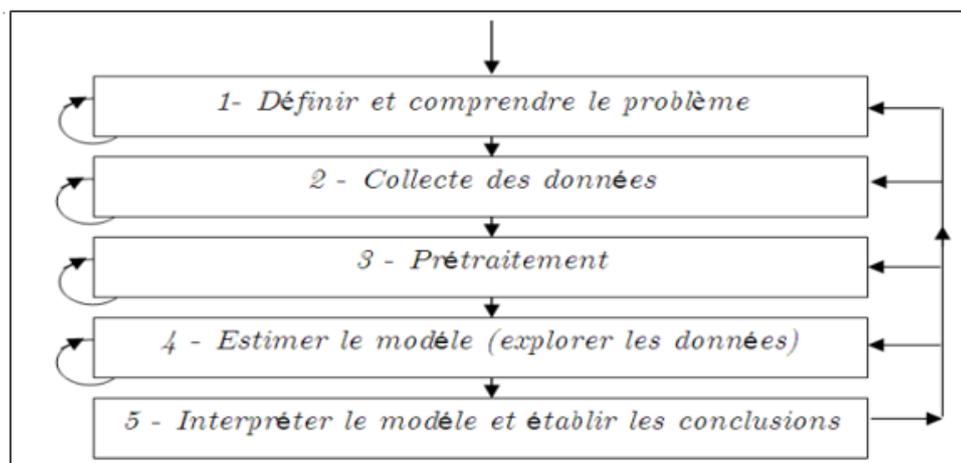


FIGURE 1.4 – Processus de data mining (CRISP-DM)

1.6.1 Définition et compréhension du problème

Il est impératif de comprendre le problème et de l'expliquer plus clairement afin d'obtenir un algorithme et des résultats fiables , Généralement, le data mining est effectué dans un domaine particulier (banques, médecine, biologie, marketing, ...etc) Comme cela nous permet de mieux évaluer les résultats et de convaincre l'utilisateur de sa rentabili. [1]

1.6.2 Collecte des données

Grâce à l'étape précédente définissant le problème et les objectifs de définition des données, nous pouvons définir des données utilisables, et ces données n'ont pas la même structure (word , excel , web...) Habituellement, les données sont divisées en deux parties, la plus grande partie (80%) qui est utilisée pour créer un modèle qui prédit les résultats futurs et le reste (20%) que nous validons ce modèle. [1]

1.6.3 Prétraitement

Une fois que nous avons rassemblé les données dont nous avons besoin, nous devons les nettoyer :

- Suppression de données en raison d'erreurs d'écriture ou d'erreurs causées par le système lui-même.
- Les données peuvent être incohérentes, c'est-à-dire qu'en dehors des périodes autorisées, elles doivent être rejetées ou calibrées.
- La sténographie est la sélection et l'élimination des caractéristiques de faible importance dans la caractérisation des données, en mesurant leurs variances.

Les courbes, les graphiques, les graphiques, etc. peuvent nous aider à sélectionner et à nettoyer les données importantes. C'est ce qu'on appelle un entrepôt de données.

1.6.4 Estimation du modèle

L'étape la plus importante est de sélectionner soigneusement la technologie appropriée pour extraire les connaissances des technologies de données telles que les réseaux de neurones, les arbres de décision, les réseaux bayésiens, le clustering, ... où plus d'une technique peut être appliquée et la bonne peut être choisie .

1.6.5 Interprétation du modèle et établissement des conclusions

Cette étape nous oblige à fournir une interprétation précise du modèle résultant, car les chercheurs conviennent que les modèles simples sont plus compréhensibles mais moins

précis, tandis que les modèles complexes sont plus précis mais difficiles à interpréter. Et à partir de là, nous sommes parvenus à aider les utilisateurs à prendre des décisions et à leur fournir des modèles compréhensibles.

1.7 Techniques du data mining

Pour réaliser des modèles de données, nous devons implémenter des tâches d'exploration de données, que nous apprendrons comme suit :

1.7.1 Réseaux de neurones

Il s'inspire de la pensée de l'esprit humain qui dépend de la méthode parallèle et simplifiée, qui est complètement différente de la façon dont l'ordinateur fonctionne. Réseaux de neurones C'est une structure de réseau constituée d'un groupe de nœuds interconnectés entre eux par des liens directionnels où chaque nœud représente une unité de traitement et les liens sont la relation entre ces nœuds . [1]

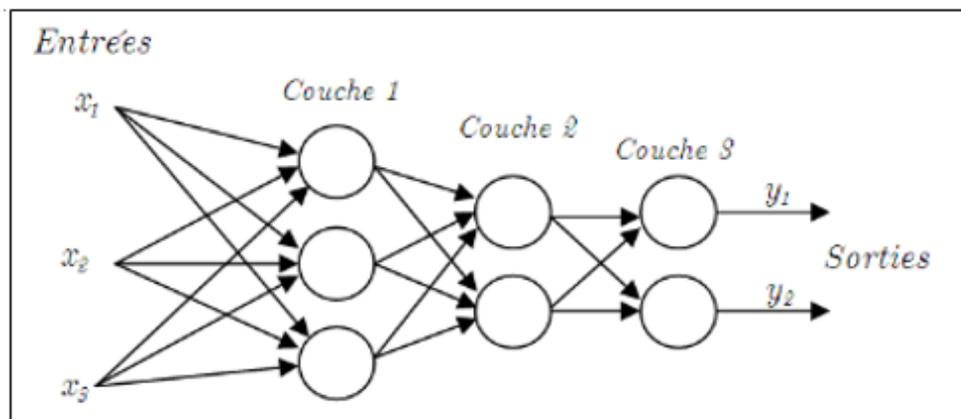


FIGURE 1.5 – Architecture d'un réseau de neurones artificiel.

Dans la figure 1.5 Les couches 1 et 2 s'appellent des couches cachées tandis que la couche 3 est la couche de sortie.

La tâche principale des réseaux de neurones artificiels est l'apprentissage pour la classification, qui est réalisée par un processus itératif d'adaptation des poids pour arriver à la meilleure fonction.[1]

■ Avantages et inconvénients

□ Avantages

- Sur monter le problème d'analyse d'un système donné pour le modéliser.
- On peut simuler son comportement uniquement à partir d'un certain nombre d'exemples observés.

- inconvénients
 - Très difficile (impossible) d'analyser et comprendre le fonctionnement en face d'un problème donné.
 - Difficulté de choisir la structure (type, nombre de nœuds, organisation, connexions,...etc) la mieux adaptée au problème.[4]

1.7.2 Les arbres de décision

C'est l'une des méthodes les plus efficaces pour traiter d'apprentissage supervisé. Ce la dépend de la division des données en groupes homogènes.

Nous prenons comme entrées un ensemble de données et de sorties et produisons un arbre qui est très similaire à un diagramme de d'orientation, Pour construire un arbre, il existe plusieurs algorithmes : ID3, CART, C4.5, ... etc, Habituellement, nous commençons par choisir un petit groupe et ensuite tester un plus grand groupe.

L'algorithme continue de manière itérative jusqu'à l'obtention des nœuds liés aux données pour chaque classe.[1]

■ Avantages et inconvénients

- Avantages
 - Bonne vitesse d'entraînement par rapport aux autres méthodes.
 - Transférable à des règles simples et compréhensibles.
 - La possibilité d'utiliser des requêtes SQL pour accéder aux bases de données.
- inconvénients
 - Plus les arbres sont complexes, moins le résultat est fiable.
 - Cela prend plus de temps pendant la construction.[4]

1.7.3 Les algorithmes génétiques

L'algorithme génétique (AG) est fondé sur une analogie avec le monde biologique. recherche dans l'espace de configurations : faire évoluer (recombinaison et modification) un ensemble de solutions du problème, et en sélectionner la meilleure à la fin.[7]

1) Principe de base des algorithmes génétiques

- (a) Génération de la population initiale : un nombre fini d'individus n choisis généralement par tirage aléatoire uniforme dans D (espace de recherche) forment la population initiale P_0 ,
- (b) Evaluation de la population initiale (calcul de la performance),
- (c) Sélection certains individus (les plus performants : les parents),

- (d) Création d'un nouvel ensemble d'individus, appelé « population d'enfants » (c'est une étape est toujours stochastique ,
- (e) Création d'une nouvelle génération P (c'est une étape de remplacement) ,[7]

2) Codage d'un algorithme génétique

Pour résoudre le problème, les informations doivent être codées selon plusieurs méthodes, comme suit :

- Codage binaire : Sous la forme d'une chaîne binaire 0 et 1 .
- Codage réel : représentation directe des valeurs réelles de la variable .[7]

Et ses avantages et inconvénients sont les suivants :

■ Avantages et inconvénients

□ Avantages

- L'attention est davantage la fonction objective et évite sa nature et est donc plus applicable.
- Pour éviter que l'idéalisme local ne passe à l'amélioration globale, nous sommes autorisés à utiliser les règles du croisement probabiliste et de la transmission des mutations. [4]

□ inconvénients

- Le calcul prend plus de temps
- Avec le temps, nous nous assurons que nous avons seulement approché la solution,etqu'elle ne peut pas être séparée en ce que c'est la meilleure solution.[4]

1.7.4 Les règles associatives

Les règles associatives sont des règles extraites d'une base de données transactionnelles et qui décrivent des associations entre certains éléments.[4]

règles d'association basées sur tous les items fréquents trouvés . [1]

■ Avantages et inconvénients

□ Avantages

- Simplicité d'interprétation des résultats.
- Faciliter la méthode de calcul.
- Aucune information préalable (Apprentissage non supervisé).[4]

□ inconvénients

- Le calcul prend plus de temps.
- Ce la donne un grand nombre de règles dont nous n'avons pas besoin.[4]

1.7.5 L'algorithme des k-Plus proches voisins

L'algorithme des k plus proches voisins (K-PPV) Il est considéré comme l'un des algorithmes de classification les plus simples en termes d'application, car il dépend du calcul de la distance entre les éléments à classer, à partir de la distance simple habituelle (la distance entre la maison et l'école) jusqu'à des procédures non numériques plus complexes si nécessaire (couleurs , sexuelles). [1]

1) Algorithme de classification par k-PPV

pour appliquer l'algorithme k-PPV :

- Paramètre : nombre de k voisins.
- **Données** : échantillon de m exemples et leurs catégories.
 - La classe d'un exemple \mathbf{X} est $\mathbf{C}(\mathbf{X})$.
- **Entrée** : Enregistrement \mathbf{Y} .
 - Déterminer les \mathbf{K} exemples les plus proches de \mathbf{Y} en calculant les distances (Euclide, masse de la ville, ...).
 - Combiner les classes de ces \mathbf{K} exemples en une classe \mathbf{C}
- **Sortie** : la classe de \mathbf{Y} est $\mathbf{C}(\mathbf{Y})=\mathbf{C}$.

2) Fonctionnement

Supposons que nous ayons un ensemble d'éléments dont la classe est connue. Lorsque nous voulons classer un élément, nous calculons sa distance par rapport au reste des éléments. Si cette règle contient éléments, nous calculons une distance de 100 et obtenons 100 nombres réels si $k = 25$ Par exemple, nous recherchons les 25 plus petits nombres entre ces nombres, ces nombres sont 25 C'est le plus proche de l'élément que nous voulons pour classer on peut changer k en fonction de ce qu'on veut faire et l'idée reste la même .[1]

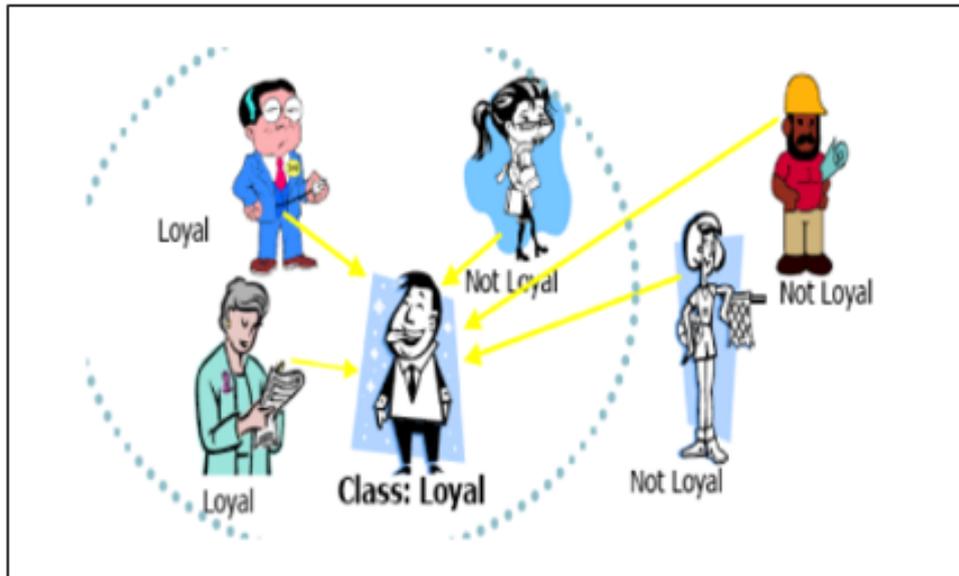


FIGURE 1.6 – Un exemple représentant K-PPV

3) Avantages et inconvénients

■ Avantages

- Au fur et à mesure que nous entrons de nouvelles données, la qualité du modèle s'améliore sans qu'il soit nécessaire de reconstruire un nouveau modèle .
- Les résultats sont plus significatifs.
- Facilité de traitement des données, aussi complexe soit-elle (son, image ...) par rapport aux autres méthodes.[4]

■ inconvénients

- Cette méthode n'a pas besoin d'être apprise par rapport aux autres méthodes.
- Si le nombre d'attributs pertinents est faible par rapport au nombre total d'attributs, cela entraînera un manque de crédibilité.[4]

1.7.6 Algorithme de clustering par K-Means

Des algorithmes les plus simples à utiliser Clustering partitionnel L'algorithme est le suivant :

1. Sélectionner une partition initiale contenant des enregistrements choisis arbitrairement, puis calculer les centres des clusters,
2. Calculer une liste des distances interclusters et la trier dans l'ordre croissant ;
3. Générer une nouvelle partition en affectant chaque enregistrement au cluster du centre le plus proche ;

4. Calculer les nouveaux centres des clusters ;
5. Répéter 2 et 3 jusqu'à ce que les enregistrements se stabilisent dans leurs clusters ;
[1]

1) **Evaluation**

K-Means est utilisé pour désigner chaque groupe de manière similaire, et à la fin nous obtenons une section. Nous essayons d'améliorer la distance moyenne entre les données de chaque groupe et son centre, mais il existe des cas particuliers, qui sont les suivants :

- Si le cluster est trop bas dans le kit, il vaut mieux le combiner avec notre cluster (non supervisée).
- S'il y a un cluster dominant, il est préférable de diviser le groupe en deux. [1]

2) **Avantages et inconvénients**

■ Avantages

- Vous n'avez pas besoin de connaissance préalable des données.
- Facile à appliquer à tout type de données .[4]

■ inconvénients

- Mauvais choix en raison de mauvais résultats.
- La difficulté de trouver une bonne fonction de distance.[4]

1.7.7 Machines à vecteur support

Les machines à vecteurs de support sont au centre du développement de la recherche humaine dans les technologies d'apprentissage. Les SVM sont une classe de techniques d'apprentissage introduites par Vladimir Vapnik au début des années 1990, qui sont basées sur une théorie mathématique solide par opposition aux méthodes de réseau de neurones. Développés dans le sens inverse du développement des réseaux de neurones, ces derniers ont suivi un chemin heuristique allant de l'application et de l'expérimentation à la théorie ; Les SVM sont donc passés de la théorie du son à l'application.[1]

Et on le trouve en deux types SVMs binaires et SVMs multiclasse.

1) **SVMs binaires**

La classification binaire est lorsque les données d'entraînement proviennent uniquement de deux classes différentes (+1 ou -1) où trouver un superplan (une ligne droite dans le cas de deux dimensions) qui sépare mieux ces deux classes, c'est-à-dire

si les données sont linéairement séparables et l'hyperplan de séparation est représenté par l'équation suivante : [1]

$$H(x) = w^T x + b \quad (1.1)$$

Où w est un vecteur de m dimensions et b est un terme. La fonction de décision, pour un exemple x , peut être exprimée comme suit : [1]

$$\begin{cases} \text{Classe} = 1 & \text{Si } H(x) > 0 \\ \text{Classe} = -1 & \text{Si } H(x) < 0 \end{cases} \quad (1.2)$$

Et c'est ce que montre la Figure 1.7 :

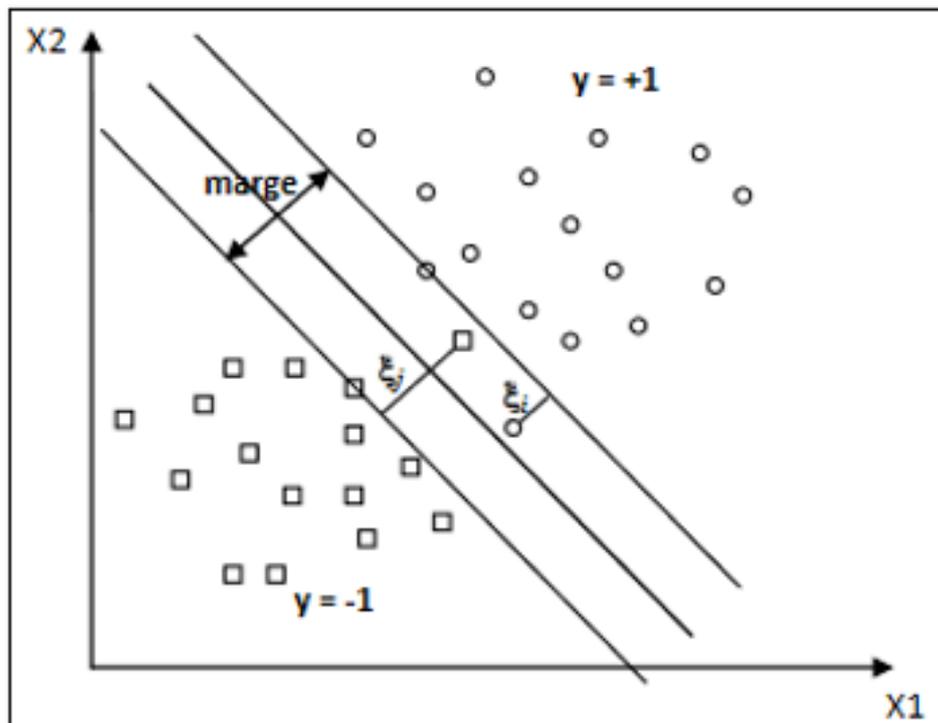


FIGURE 1.7 – SVMs binaires

2) SVMs multiclasse

Les machines à vecteurs de support sont à l'origine binaires. Avec des problèmes nous abordons des multi-couches et l'exemple le plus simple est la Reconnaissance Optique de Caractères (OCR) On cherche à attribuer un nouvel exemple à l'une des deux classes mais à l'une de plusieurs classes.[1]

Et la figure montre l'idée 2.1 :

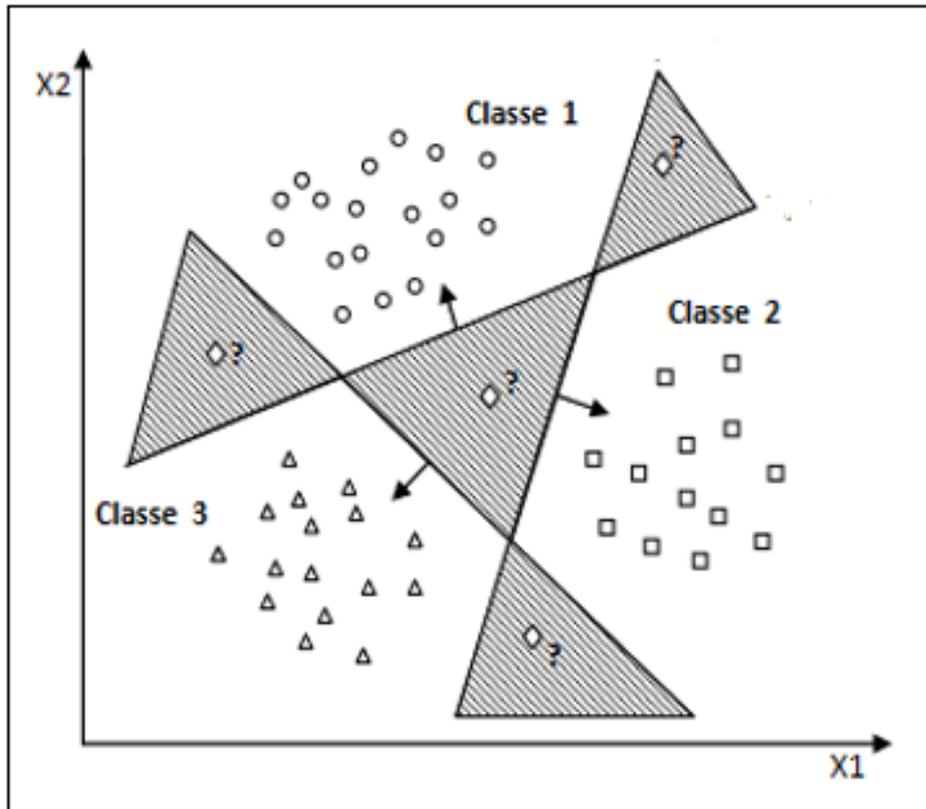


FIGURE 1.8 – SVMs multiclasse

3) Avantages et inconvénients

■ Avantages

- Ce même modèle peut être utilisé pour résoudre des problèmes de régression et de classification.
- C'est souvent une meilleure alternative aux réseaux de neurones, en particulier dans le cas de petits ensembles de données.
- très précis avec une marge de séparation nette.[8]

■ inconvénients

- la mémoire nécessaire augmente avec le nombre de support vectors.
- les modèles sont difficilement interprétables par des humains, contrairement par exemple aux arbres de décisions.
- La difficulté de trouver une bonne fonction de distance.[8]

1.8 Catégorisation des systèmes du data mining

Le processus de classification est effectué sur la base d'un ensemble de principes, dont les suivants :

- Classification selon le type de données : Les données sont classées et regroupées selon le type de données dont nous traitons (texte, nombres, ...)
- Classification selon les modèles de données avancés : La classification prend en charge les modèles de données avancés dans ce processus les bases de données relationnelles, les bases de données orientées objets, les data warehouses, les bases de données transactionnelle, etc.
- Classification selon le type de connaissance à découvrir : Les données sont classées en fonction des connaissances à accéder (classification, l'estimation, la prédiction, ..)
- Classification selon les techniques d'exploration utilisées : Dans ce type, la classification est effectuée selon l'approche analytique utilisée (les réseaux neurones, les algorithmes génétiques ...).[4]

1.9 Domaines d'application du data mining

L'objectif principal de l'exploration de données reste d'extraire des connaissances ciblées qui peuvent être appliquées. C'est pourquoi de nombreuses personnes et chercheurs intéressés s'efforcent de l'appliquer dans tous les domaines de la vie, et pour connaître ces domaines, nous les mentionnons :

1) Le data mining dans le secteur bancaire

Le secteur bancaire est l'un des secteurs les plus importants qui utilisent l'exploration de données, et c'est ce qui en a fait le chef de file des autres secteurs. L'application de l'exploration de données est la suivante :

- Attentes d'augmentation ou de diminution des taux d'intérêt.
- Identifier les clients qui sont en retard dans les remboursements de prêts.
- Identifiez les clients fidèles.
- Déterminez qui peut changer de banque.[4]

2) Le data mining dans la bio-informatique

Un champ large qui comprend plusieurs disciplines (informatique, mathématiques, microbiologie, écologie ...) et les données y sont en constante augmentation et en augmentation. L'application de l'exploration de données est la suivante :

- Prédire l'émergence d'un nouveau type de protéine.

- Prédire les symptômes de certains médicaments. [4]

3) Le data mining dans le marketing direct et le collecte de fonds

Le marketing est l'une des stratégies les plus importantes qu'un vendeur adopte pour promouvoir certains produits toujours variés, ce qui nous pousse à appliquer le data mining.

L'application de l'exploration de données est la suivante :

- Prédire une série de produits pouvant être achetés avec les clients.

- Prédire les résultats des campagnes électorales.

- Identifier les clients intéressés par un produit spécifique.

4) Le data mining dans la détection de fraude

L'application de l'exploration de données est la suivante :

- Détection de la fraude au passeport grâce à l'identification personnelle

- Détecter la fraude par carte de crédit. [4]

- Détecter le plagiat scientifique.[4]

5) Le data mining dans la gestion de données scientifiques

L'application de l'exploration de données est la suivante :

- Prévision des conditions météorologiques.

- Prédire les phénomènes naturels.

- Déterminez quels engrais seront présentés par les plantes. [4]

6) Le data mining dans le secteur des assurances

Les compagnies d'assurance visent à fidéliser leurs clients tout en augmentant leurs profits et c'est ce qui les motive Pour utiliser des méthodes d'exploration de données.

L'application de l'exploration de données est la suivante :

- Identifiez les personnes qui utilisent le plus la sécurité .

- Concevoir des modèles pour prédire le risque d'assurance. [4]

7) Le data mining dans la télécommunication

L'application de l'exploration de données est la suivante :

- Prévoyez davantage les heures des appels téléphoniques.
- Identifiez les personnes qui utilisent le plus d'appels téléphoniques. [4]

8) **Le data mining dans la médecine et la pharmacie**

L'application de l'exploration de données est la suivante :

- Anticiper certaines maladies.
- Déterminer les personnes les plus sensibles à une maladie spécifique.
- Choix des antibiotiques pour des infections. [4]

9) **Le data mining dans le commerce au détail**

L'application de l'exploration de données est la suivante :

- Identifiez les personnes les plus intéressées par le commerce de détail .
- L'évaluation des risques pour chaque transaction.
- Déterminez les produits qui intéressent le plus les clients.
- Prédire les prix des parts de marché. [4]

10) **Le data mining dans le e-commerce et le World Wide Web**

L'application de l'exploration de données est la suivante :

- Identifiez les personnes les plus intéressées par le e-commerce.
- En formulant des tactiques du marché dans les opérations de business.
- Déterminer les produits les plus réceptifs au client. [4]

1.10 Conclusion

La technologie est l'élément principal du développement des données et des informations dans plusieurs domaines, et c'est ce qui pousse de nombreux chercheurs à accorder plus d'attention à l'exploration de données, qui sont les nouvelles technologies de ces derniers temps et qui ont donné des résultats corrects sur lesquels on peut compter.

Le chapitre suivant Utiliser l'exploration de données pour découvrir de nouvelles associations, modèles et tendances dans le domaine de la pharmacie.

Chapitre 2

Une étude sur le domaine de la pharmacie

Sommaire

2.1	introduction	22
2.2	definition le domaine pharmaceutique	22
2.2.1	Majors en pharmacie	22
2.3	Le médicament	23
2.3.1	Définition	23
2.3.2	Type de Médicament	23
2.3.3	Composition du médicament	24
2.3.4	Les voies d'administration d'un médicament	25
2.3.5	Médicaments dangereux	27
2.3.6	Classification des médicaments	28
2.4	Travaux Connexes	32
2.4.1	Les objectifs ATC	32
2.4.2	La Structure du système ATC	33
2.4.3	Classification des antibiotiques	35
2.4.4	Origine	35
2.4.5	Mode d'action	35
2.5	Conclusion	36

2.1 introduction

Le domaine de la santé est l'un des domaines les plus importants qui occupent la majeure partie de la population mondiale, mais la présence de certains problèmes pousse de nombreux chercheurs en pharmacie à appliquer les dernières techniques d'exploration de données afin d'extraire et d'appliquer les connaissances.

Dans cette partie, nous discuterons de la définition du domaine de la pharmacie et de certains des éléments qui le combinent avec l'exploration de données.

2.2 définition le domaine pharmaceutique

C'est une science qui étudie les propriétés du médicament et de ses composés et comment le fabriquer et à travers les étapes qu'il traverse, il est utilisé comme traitement pour des maladies spécifiques.

2.2.1 Majors en pharmacie

Avec la diversité illimitée du domaine de la pharmacie (objectifs, agencement ...), elle est répandue et présente dans plusieurs disciplines. Il peut être classé en quatre majors comme suit :

- Aspect santé publique

Il comprend des organisations gouvernementales et non gouvernementales qui cherchent à améliorer et à réglementer le domaine de la santé.

- Aspect scientifique

Comprend des chercheurs et des développeurs en pharmacologie

- L'aspect de l'intérêt national et de l'économie

Il s'agit notamment des Ministères de l'économie et de la sécurité sociale, du contrôle des coûts et de la sécurité personnelle ...

- Le côté commercial

Elle concerne les producteurs ou fabricants de produits pharmaceutiques (importateurs, grossistes, pharmacies ...). Impact sur les nouveaux produits, la gestion des stocks et la génération de bénéfices.[9]

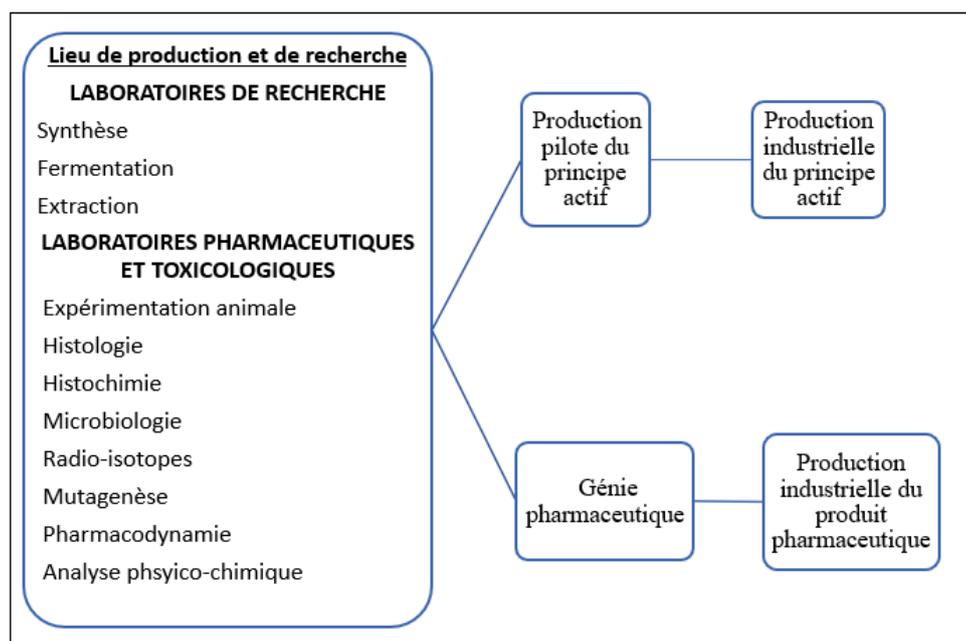


FIGURE 2.1 – La méthode de développement de médicaments

La recherche et le développement sont l'un des facteurs les plus importants qui contribuent à améliorer les médicaments, comme l'illustre Figure 2.1, car il existe des sociétés pharmaceutiques multinationales qui s'y intéressent en fonction des données du marché. Nous pouvons également trouver une coopération entre les hôpitaux et les centres de recherche pharmaceutique en développant des médicaments.[10]

2.3 Le médicament

2.3.1 Définition

Les médicaments désignent toute substance ou combinaison présentée comme ayant des propriétés thérapeutiques ou protectrices ou dans le but de diagnostiquer une maladie chez l'homme ou l'animal.

Ou restaurer, corriger ou modifier leurs fonctions physiologiques par la pratique d'actions pharmacologiques, immunologiques ou métaboliques . [11]

2.3.2 Type de Médicament

1) Spécialité pharmaceutique

Ce nom correspond à tout médicament préalablement préparé dans une entreprise pharmaceutique et mis sur le marché (ou distribué sous forme d'échantillon médical) sous un nom commercial qui lui est donné par l'entreprise qui le produit en accord avec le ministère de la Santé.[11]

Exemple :

Parol, Asepegic, Neuroféne ...

2) **Préparation magistrale**

Cette dénomination correspond à tout médicament préparé à l'officine par le pharmacien à la demande du patient, et dont la formule est composée par le médecin prescripteur.

La boîte contenant la préparation magistrale doit mentionner le nom du patient, le nom du prescripteur, la date, la posologie, la composition, et le numéro d'ordre sous lequel elle est enregistrée dans l'ordonnancier.

Exemple :

sirop contre la toux, pommade soufrée

3) **Préparation officinale**

Il comprend les médicaments préparés en pharmacie selon les normes pharmaceutiques et avec des principes actifs non soumis à prescription et acceptés comme médicaments de consultation.

Exemple :

sirop contre la toux, pommade soufrée ...

2.3.3 Composition du médicament

Les médicaments se composent de plusieurs substances qui aident et donnent de l'efficacité au patient, comme suit :

1) **Le principe actif (PA)**

C'est la molécule utilisée pour créer le médicament et lui donner ses propriétés curatives. Le médicament contient un ou plusieurs ingrédients actifs inclus dans l'excipient.

2) **Les excipients (EX)**

Les excipients viennent du mot latin «excipere», qui signifie recevoir, en d'autres termes, la personne recevant l'AP. C'est une substance inactive en soi, Son objectif :

- Présenter le médicament sous une forme adaptée pour la voie d'administration souhaitée : (comprimé, solution buvable, gélules, suppositoires...).
- Modifier le goût et l'odeur du médicament.
- Améliorer la conservation du médicament.

Exemple :

Eau, amidon, silice, glycine, arôme mandarine, bicarbonate de sodium, acide citrique...[11]

3) **Les articles de conditionnement (ADC)**

L'emballage primaire est un composant essentiel des médicaments il est donc nécessaire d'obtenir l'AMM et la fabrication de la forme pharmaceutique, le choix de l'emballage primaire (emballage primaire) est le résultat de longs mois de tests de stabilité sous différentes températures et humidité relative y compris en temps réel .

Exemple :

Blisters, pilulier, flacons, sachet...[11]

2.3.4 Les voies d'administration d'un médicament

Il existe plusieurs façons de prendre des médicaments, selon les besoins du patient, et chacune a ses propres avantages et inconvénients.

1) **Voie orale ou voie buccale (par os)**

Plus de 70% des médicaments sont pris par voie orale, et c'est le plus courant. Après administration orale, le médicament traverse la barrière intestinale puis le foie avant d'atteindre la circulation générale, et de là, il traverse les organes pour son action thérapeutique.[11]

Il contient des avantages et des inconvénients, qui sont les suivants :

■ **Avantages**

- Il est facile d'administrer des médicaments de cette façon.
- Des doses élevées peuvent être prises en une seule fois sans risque.

■ **inconvénients**

- La voie orale est difficile d'utilisation en cas de nausées et de vomissements.
- Elle ne peut pas être utilisée chez un patient inconscient..
- Il y a irritation du tube digestif par certains médicaments (anti-inflammatoires, corticostéroïdes...).[11]

2) **Voie parentérale ou voie injectable**

C'est la voie la plus directe, car elle met le médicament en contact direct avec le sang ou les liquides interstitiels et contourne le système digestif. Les médicaments parentéraux sont des préparations liquides injectables (solutions, émulsions, suspensions) ou solides (implants).

Il contient des avantages et des inconvénients, qui sont les suivants :

■ Avantages

- Voie utile pour un malade inconscient ou incapable d'avaler.
- Passage immédiat dans l'organisme.
- Injection possible au lieu désiré (par exemple injection intracardiaque).

■ inconvénients

- Elle peut entraîner un risque d'infection.
- Il ne faut pas injecter des produits irritants qui pourraient léser la veine.
- La durée d'action médicamenteuse est généralement courte, surtout après administration intraveineuse.[11]

3) **Voies transmuqueuses**

Le médicament est pris à travers les muqueuses riches en vaisseaux sanguins que le sang transporte vers l'organe cible via (nasale, sublinguale, rectale, vaginale ...) . Ses avantages et inconvénients sont :

■ Avantages

- Le médicament n'est pas affecté par l'action des enzymes digestives.
- Convient aux nourrissons et aux enfants, ainsi qu'aux personnes nauséuses, inconscientes ou incapables d'avaler.

■ inconvénients

- Elle peut provoquer une irritation ou même une ulcération de la muqueuse rectale.
- Elle peut être jugée désagréable par le patient.[11]

4) **Voie pulmonaire**

L'arbre pulmonaire, qui a une superficie de 100 m² chez un humain adulte et dilate les gros vaisseaux sanguins, permet une absorption rapide des médicaments. Cette méthode est utilisée pour absorber les gaz (oxygène, chloroforme, éther, etc.).

Il permet la circulation d'un grand nombre de médicaments (sulfamides, antibiotiques, médicaments contre l'asthme, ...), sous la forme d'un spray de fines particules de médicament. [11]

5) **Voies cutanée et percutanée**

Le médicament est appliqué au niveau de la peau, donc seule une peau saine constitue une barrière efficace entre les environnements interne et externe. Dans le cas contraire (lésions, brûlures, eczéma), tout médicament (pommades, gels, lotions ...) appliqué sur la peau sera absorbé de manière significative. [11]

2.3.5 Médicaments dangereux

Nous ne pouvons pas nier l'efficacité de certains médicaments dans le traitement des maladies et donner des résultats positifs, mais il existe des médicaments qui, dans certains cas, constituent une menace pour la santé humaine et ils sont les suivants :

■ Les produits chimiques industriels

Ils sont utilisés dans le développement de la médecine et la fabrication de médicaments prêts à l'emploi qui sont utilisés comme réactifs, catalyseurs ou solvants, ce qui les a rendus dangereux pour leurs utilisateurs, ce qui a conduit l'American Conference of Hygienists à fixer des limites d'exposition professionnelle.

■ Les substances médicamenteuses

Les substances actives dans le domaine de la pharmacie peuvent être classées comme produits naturels d'origine végétale ou animale et drogues synthétiques représentées dans les produits chimiques où nous trouvons plus d'utilisation et d'intérêt pour les matériaux synthétiques avec le développement continu en biologie moléculaire, biochimie, pharmacologie et informatique, où le Le tableau suivant représente certains produits pharmaceutiques.[12]

Un tableau 2.1 représentant certaines des principales catégories de produits pharmaceutiques.

Système nerveux central	Système immunitaire	Système endocrinien
Analgésiques Paracétamol Salicylés	Analgésiques Anti-inflammatoires non stéroïdiens	Inhibiteurs des corticoïdes Prostaglandines
Anticonvulsivants Barbituriques Benzodiazépines	Antitoxines Sérums immunisants	Glucocorticoïdes Hormones
Produits utilisés en psychiatrie Anxiolytiques Antidépresseurs	Vaccins Vaccins	Agents diagnostiques Corticoïdes

TABLE 2.1 – Catégories de produits pharmaceutiques.

2.3.6 Classification des médicaments

Il existe plusieurs types de médicaments qu'un patient peut prendre par voie orale, ou injecter, et les médicaments placés sur la peau, qui sont inhalés par pulvérisation, mais pour suivre ces médicaments et les classer pour faciliter leur extraction, il faut appliquer ce qui est appelé le système de classification des médicaments développé par l'organisation Global Health. qui est le suivant :

- Classification clinique (therapeutic)
- Classification pharmacologique (dépend du mécanisme d'action et du mode d'action)
- Classification chimique
- Classification légale (listes de médicaments et risque de ternissement).

2.3.6.1 Classification clinique (therapeutic)

Cette classification est basée sur le bénéfice des médicaments dans le traitement de maladies spécifiques.

Et ce tableau 2.2 montre quelques classifications clinique de médicaments pour les maladies cardiaques . [13]

Classification clinique	
Country	ALPHA
Influence la coagulation sanguine	Anticoagulant
Abaissier le cholestérol sanguin	Antihyperlipidémique
Abaissier la pression artérielle	Antihypertenseur
Rétablir le rythme cardiaque normal	Antidysrythmique
Traiter l'angine	Antiangineux

TABLE 2.2 – Classification therapeutic

2.3.6.2 Classification pharmacologique

La classification pharmacocinétique est basée sur sa classification grâce à la connaissance du mode d'action du médicament aux niveaux moléculaires, aux tissus et au système corporel. Ou comment le médicament est produit en raison de son effet physiologique sur le corps.

Ce tableau 2.3 présente la classification pharmacologique des antihypertenseurs. [13]

Classification pharmacologique	
Focus sur l'application thérapeutique : pharmacothérapie pour l'hypertension	
Mécanisme d'action	Diurétique
Réduit le volume de plasma	Bloqueur des canaux calciques
Abaissier le cholestérol sanguin	Inhibiteur de l'enzyme de conversion de l'angiotensine
Bloque les canaux calciques du cœur	Antagoniste adrénergique
Bloque l'activité hormonale	Antidysrythmique

TABLE 2.3 – Classification pharmacologique

2.3.6.3 Classification chimique

Sa classification est basée sur trois noms de médicaments, à savoir :

■ Noms chimiques

- Une seule désignation chimique pour un médicament est attribuée par la désignation standard définie par l'Union internationale de chimie pure et appliquée (UICPA). Il nous permet de prédire les propriétés physiques et chimiques d'une substance.

Exemple :

Le nom chimique de l'aspirine est l'acide 2-acétoxybenzoïque.[13]

■ Noms génériques

- Les noms génériques sont plus simples et moins compliqués que les noms chimiques. Les noms de médicaments génériques sont déterminés par le Conseil des noms approuvés aux États-Unis. Alors que la Food and Drug Administration (FDA), la pharmacopée officielle et l'Organisation mondiale de la santé (OMS) décrivent systématiquement le médicament dans son nom générique. [13]

■ Noms commerciales

- La société qui commercialise le médicament est responsable de la définition du nom de marque afin qu'il soit court et facile à retenir. Un nom commercial est parfois appelé nom de marque, nom de produit .[13]

Le tableau 2.4 suivant montre le médicament et certains de ses noms commerciaux :

Exemples de produits de marque	
Substance générique	Les noms de marques
aspirine	Acuprin, Anacin, Aspergum, Bayer, Bufferin, Ecotrin, Empirin, Excedrin, Maprin, Norgesic, Salatin, Salocol, Salsprin, Supac, Talwin, Triaphen-10, Vanquish, Verin, Zorprin
ibuprofène	Advil, Amersol, Apsifen, Brufen, Haltran, Medipren, Midol 200, Motrin, Neuvil, Novoprofen, Nuprin, Pamprin-IB, Rufen. Trendar

TABLE 2.4 – Classification chimique

2.3.6.4 Classification légale (listes de médicaments et risque de ternissement)

■ listes de médicaments

- Certains médicaments sont fréquemment mal utilisés ou présentent un fort potentiel de dépendance. Sur cette base, les drogues à fort potentiel d'abus sont classées en cinq catégories appelées délais, du plus haut potentiel d'abus au plus bas pour l'abus.

Le tableau 2.6 suivant montre que :

Classification légale		
Calendrier des médicaments	Exemples	Usage thérapeutique
I	Heroin, ecstasy	Utilisation thérapeutique limitée ou inexistante
II	Hydromorphone methadone, eperidine, fentanyl	
III	Ketamine, anabolic steroids	Utilisé en thérapeutique sur ordonnance.
IV	Alprazolam, diazépam	
V	Préparations contre la toux ne contenant pas plus de 200 mg de codéine par 100 ml ou par 100 grammes.	Utilisé en thérapeutique sans ordonnance.

TABLE 2.5 – Classification légale

■ risque de ternissement

- Un tératome est une substance qui peut provoquer une anomalie chez un fœtus pendant la grossesse. Peu de médicaments se sont avérés avoir un effet tératogène, tant chez l'homme que chez les animaux de laboratoire. La classification des risques tératogènes place les médicaments dans les classes A, B, C, D et X.[13]

Le tableau 2.6 présente la classification des médicaments en fonction de leurs risques pour la grossesse :

Classification légale	
catégorie	Classification des risques médicamenteux par la FDA pendant la grossesse
A	Les études contrôlées chez les femmes ne montrent pas de risque pour le fœtus et la possibilité de dommages au fœtus semble peu probable .
B	Les études sur la reproduction animale n'ont pas montré de risque fœtal ni d'effet indésirable. Les risques n'ont pas été confirmés dans les études contrôlées chez les femmes .
C	Les études chez l'animal ont révélé des effets indésirables sur le fœtus et il n'y a pas d'études contrôlées chez les femmes ou des études chez les femmes et les animaux ne sont pas disponibles.
D	Il y a confirmation du risque foetal humain, mais les avantages de l'utilisation chez la femme enceinte peuvent être acceptables malgré le risque.
X	Des études animales et humaines ont montré des anomalies fœtales. Le médicament est contre-indiqué chez les femmes enceintes ou susceptibles de l'être.

TABLE 2.6 – Classification avec risque de ternissement

Les études sur la reproduction animale n'ont pas montré de risque foetal ni d'effet indésirable. Les risques n'ont pas été confirmés dans les études contrôlées chez les femmes .

2.4 Travaux Connexes

2.4.1 Les objectifs ATC

Le système ATC (anatomique-thérapeutique-chimique) est destiné à servir d'outil de recherche sur l'abus des médicament pour améliorer la qualité de l'abus des médicament. Dans le système de classification ATC, les médicaments sont divisés en différents groupes en fonction de l'organe ou du système sur lequel ils travaillent et de leurs propriétés chimiques, pharmacologiques et thérapeutiques. [14]

La définition de l'unité DDD est : La dose d'entretien moyenne supposée, par jour, pour un médicament, utilisée dans son indication principale chez l'adulte. Il y a donc 3 Élément

principaux :

- Dose journalière moyenne.
- Dose dans l'indication principale.
- Dose utilisée chez l'adulte.

2.4.2 La Structure du système ATC

Dans le système ATC, les médicaments sont divisés en plusieurs groupes selon l'organe ou le système sur lequel ils travaillent et leurs propriétés chimiques, pharmacologiques et thérapeutiques.

Les médicaments sont classés en groupes de 5 niveaux différents . Ils sont d'abord divisés en 14 groupes principaux (le premier niveau de classification selon l'appareil ou le système dans lequel vous travaillez). [14]

A : Appareil digestif et métabolisme

B : Sang et organes hématopoïétiques

C : Système Cardio-vasculaire

D : Dermatologie

G : Système Génito-urinaire et hormones sexuelles .

H : Préparations systémiques hormonales à l'exception des hormones sexuelles et de l'insuline

J : Anti-infectieux à usage systémique

L : Antinéoplasiques et agents immunomodulateurs

M : Système musculo-squelettique

N : Système nerveux

P : Produits antiparasitaires, insecticides et révulsifs

R : Système respiratoire

S : Organes sensoriels

V : Divers

2.4.2.1 Le principes de classification du système ATC

- leur principale utilisation thérapeutique.
- leur principal principe actif.
- sur le principe basique de seulement un seul code ATC pour chaque voie d'administration. ATC utilise des symboles et Structures pécifiques qui représentent les noms de certains médicaments qui ont une forme : (L représente une lettre et C un chiffre)
LCCLLCC

code ATC	Utilisation thérapeutique
A	Appareil digestif et métabolisme (1erniveau, principal groupe anatomique)
A10	traitements utilisés dans le diabète (2èmeniveau, sous-groupe thérapeutique)
A10B	traitements utilisés dans le diabète pharmacologique
A10BA	Biguanides (5ème niveau, principe actif)
A10BA02	Metformine (5ème niveau, principe actif)

TABLE 2.7 – Les 5 niveaux différents des médicaments .

A02BC01= Oméprazole

J01CA04 = Amoxicilline

M01AC01 = Piroxicam

N02BE01 = Paracétamol

Exemple 1 :

Les hormones sexuelles sont utilisées pour des forces spécifiques et sont classées dans le traitement du cancer sous :

L02 – Thérapie endocrinienne.

G03 – Hormones sexuelles et modulateurs du système génital.

Exemple 1 :

Clonidine est disponible en 2 dosages différents pour traiter :

Hypertension (C02 – Antihypertenseurs)

Les migraines (N02C – Préparations antimigraineuses).

Exemple 3 :

La prednisolone a plusieurs codes ATC du fait des différents usages thérapeutiques et des différentes formulations pour applications locales :

A07EA01 : Predisolone utilisé comme agent anti-inflammatoire intestinal.

S01BA04 : Predisolone utilisé comme agent corticoïde ophtalmologique.

S02BA03 : Predisolone utilisé comme agent corticoïdes Otologique.[14]

2.4.3 Classification des antibiotiques

La classification des antibiotiques peut se faire selon :

2.4.4 Origine

- Naturelle : Elaborés par un organisme.

Exemple :

Pénicilline, streptomycine, tétracyclines, céphalosporine.

- Synthétique ou semi synthétique : Produits par synthèse.

Exemple :

Bêta-lactamines, aminosides, sulfamides. [15]

2.4.5 Mode d'action

Les antibiotiques peuvent agir à différents niveaux :

- Agissant sur les enveloppes bactériennes
 - Agissant sur la paroi bactérienne.
 - Agissant sur la membrane plasmique.
- Agissant sur la synthèse des protéines
 - Antibiotique se fixant à la sous-unité 50S du ribosome.
 - Antibiotique se fixant à la sous-unité 30S du ribosome.
 - Antibiotiques interférant avec les facteurs d'élongation.
- Agissant sur la synthèse et la fonction de l'acide nucléique
 - Agissant sur l'ADN.
 - Agissant sur l'ARN.
- Agissant sur le métabolisme bactérien.
 - Inhibiteurs de la synthèse de l'acide folique. [15]

2.4.5.1 Spectre d'activité

Comprend la liste des espèces dans lesquelles l'antibiotique est actif. Il existe deux catégories principales :

(a) **Les antibiotiques à spectre étroit**

Il ne tue que les bactéries à l'origine de l'infection, tout en évitant les bactéries potentiellement bénéfiques, et ces antibiotiques sont généralement prescrits lorsque les bactéries sont connues avec précision. [15]

(b) Les antibiotiques à spectre large

Un spectre efficace contre de nombreuses bactéries, et ce type d'antibiotique est prescrit lorsque les bactéries à l'origine de l'infection sont inconnues ou lorsque la maladie est causée par plusieurs bactéries différentes.

2.4.5.2 Nature chimique

Très variable, elle est basée souvent sur une structure de base (**ex : cycle B lactame**) sur laquelle il y a hémi-synthèse. Elle nous permet de classer les antibiotiques en familles (bêtlactamines, aminosides, tétracyclines, etc.). [15]

2.4.5.3 Expression de l'activité antibactérienne

Le tableau 2.8 montre les antibiotiques .

L'activité anti bactérienne s'exprime par deux modes différents :

Antibiotiques bactériostatiques	Antibiotiques bactéricides
Macrolides	Bêtalactames
Lincosamides	Aminoglycosides
Lincosamides	Polymixines
Sulfamides	Ansamycines

TABLE 2.8 – **Antibiotiques bactériostatiques et bactéricides.**

- Agissant sur la synthèse et la fonction de l'acide nucléique
 - Et c'est par inhibition partielle ou complète du développement des micro-organismes sans les tuer .

- L'effet bactéricide
 - Et cela en stoppant le développement des microorganismes par mort cellulaire, avec ou sans dégradation. [15]

2.5 Conclusion

Compte tenu des obstacles auxquels le pharmacien est confronté, notamment dans la gestion et l'administration des médicaments, nous avons abordé l'application de l'exploration de données, comme nous l'avons évoqué dans ce chapitre sur le domaine des médicaments et sur quelle base ils sont classés et certains de ses composants où nous constatons qu'il existe une relation entre cela et l'exploration de données.

Chapitre 3

Conception

Sommaire

3.1	Introduction	37
3.2	Architecture globale	37
3.3	Architecture détaillée	38
3.3.1	Pré-traitement	39
3.3.2	Apprentissage	40
3.4	Fonctionnement du système	42
3.4.1	Diagramme de classes	42
3.4.2	Diagramme de Séquence	44
3.4.3	diagramme de cas d'utilisation	45
3.5	Conclusion	45

3.1 Introduction

L'augmentation continue et énorme des données par les pharmacies, augmente la difficulté et la complexité de l'extraction et de l'analyse des données, et nous oblige à préparer des techniques et des systèmes capables d'extraire des données en toute crédibilité.

Dans ce chapitre, nous fournirons une description générale de notre système de classification des médicaments, en détaillant toutes les étapes que nous avons traversées et en montrant toutes les techniques et algorithmes qui nous y aident.

3.2 Architecture globale

Notre système classe les médicaments d'une manière spécifique, en prenant le système comme entrée, une base initiale de médicaments et de leurs propriétés et en les transfor-

mant en une base organisée de caractéristiques qui peuvent être utilisées dans la phase d'apprentissage.

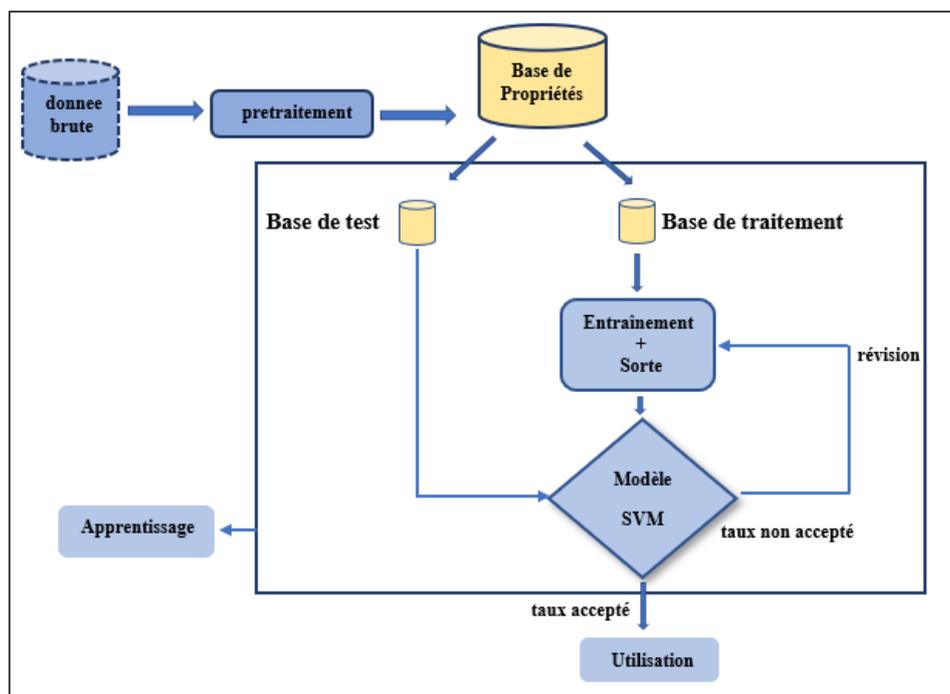


FIGURE 3.1 – Architecture globale

Cette transformation est appelée prétraitement et implique une série d'opérations spécifiques à la base de données où notre base de données n'a pas besoin d'être nettoyée et filtrée car nous venons de la chiffrer. Après avoir obtenu une base de données de traitement, nous la divisons en deux parties, une partie d'apprentissage et une partie de test, l'unité d'apprentissage utilise l'algorithme d'apprentissage pour fournir un modèle de décision qui est appliqué à la base de données de test.

Si le modèle dans lequel nous avons utilisé l'algorithme Machines à vecteur support est accepté, c'est-à-dire qu'il a pu atteindre un taux de reconnaissance acceptable (un certain pourcentage), alors ce modèle peut être utilisé et travaillé.

Sinon, l'algorithme est revu et corrigé afin d'avoir un meilleur taux de reconnaissance que le premier.

3.3 Architecture détaillée

Ci-dessous, nous détaillons chacune des étapes que nous avons traversées dans notre système.

3.3.1 Pré-traitement

Notre objectif est d'organiser les caractéristiques de la base de données brute afin de classer les médicaments, comme nous l'avons dit plus tôt que notre traitement de base de données n'a pas besoin d'être nettoyé et filtré et ce sont des données textuelles représentant le médicament, l'ID du médicament, le médicament actif, la classe de médicament, classe de médicament où il est traité.



FIGURE 3.2 – Pré-traitement de la base

■ Données textuelles

Représente le texte brut de notre base de données auquel nous appliquerons l'opération suivante :

□ Encodage

Nous avons utilisé une numérotation simple de nos sources selon un algorithme auxiliaire dont nous parlerons dans le dernier chapitre, et cela en fait partie, comme dans l'exemple de la figure suivante :

	0	1	2	3	4
0	1.0	0.0	0.0	0.0	0.0
1	1.0	0.0	0.0	0.0	0.0
2	1.0	0.0	0.0	0.0	0.0
3	1.0	0.0	0.0	0.0	0.0
4	1.0	0.0	0.0	0.0	0.0
5	1.0	0.0	0.0	0.0	0.0
6	1.0	0.0	0.0	0.0	0.0
7	1.0	0.0	0.0	0.0	0.0

FIGURE 3.3 – Encodage de la source

3.3.2 Apprentissage

Il comprend deux parties d'apprentissage et de validation de chacune d'entre elles à l'aide de la base de données que nous avons divisée en deux parties, la partie apprentissage sert à présenter un modèle de décision et la partie test sert à mesurer les performances du modèle obtenu.

1) Entraînement

Afin d'entraîner le modèle obtenu, nous avons choisi un algorithme Machine à vecteurs de support dont nous avons parlé dans le premier chapitre, pour deux raisons :

- Il donne les meilleurs résultats en text mining. [16]
- Dans notre modèle, nous avons plus de trois classifications qu'un médicament peut avoir (analgésiques, anesthésiques, antibactériens, ...) et cela correspond à notre algorithme Machine à vecteurs de support (SVM) .

Où nous avons utilisé un type spécifique de classification des médicaments qui est la classification thérapeutique où il se concentre en le classant en fonction de l'usage thérapeutique des médicaments .

La figure 3.4 illustre cette idée plus :

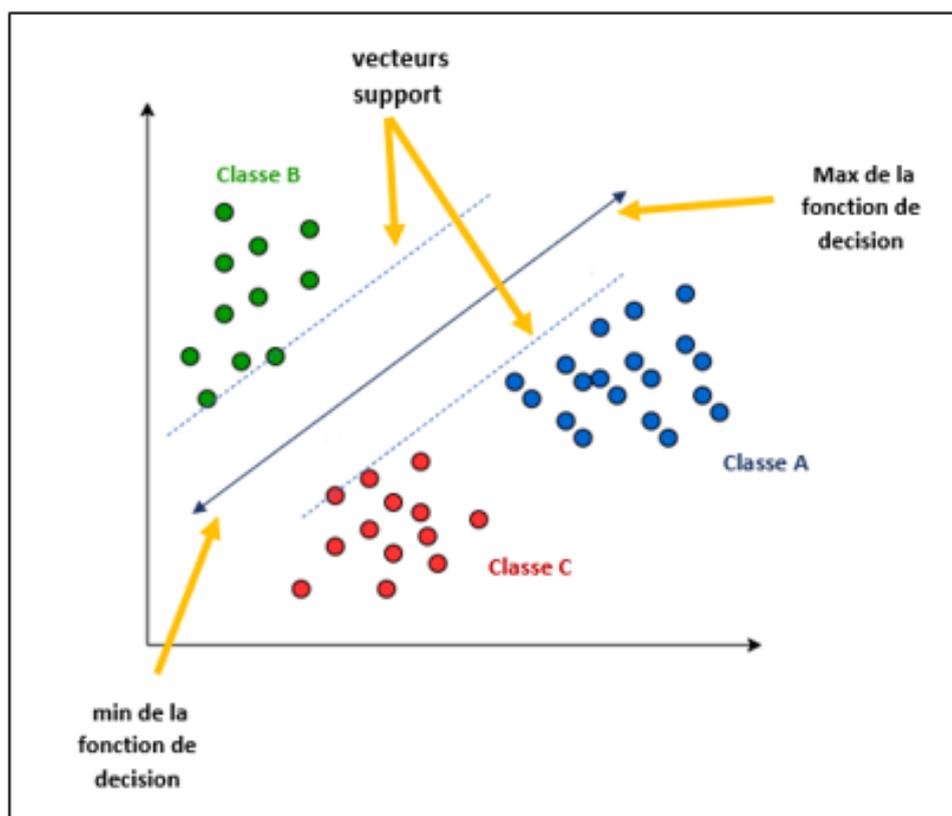


FIGURE 3.4 – Classification des médicaments utilisant SVM

Le résultat de la formation est un modèle qui représente l'analyse des données, leur transformation en informations utiles et l'établissement de relations entre elles. Où nous utilisons certaines métriques et critères (entrées, sorties) pour aider à estimer la qualité du modèle, qui sont les suivants :

■ Entrées

- Médicament : nom du médicament

■ Sorties

- ID du médicament : identificateur de médicament.
- médicament actif : la matière active.
- classe de médicament : La classe de médicaments à laquelle appartient le médicament.
- catégorie de médicament : (classe thérapeutique) fonction de médecine .

2) Validation

Ceci grâce à la capacité du modèle obtenu à identifier de nouveaux exemples, où nous avons divisé la base de données en deux parties, une partie d'apprentissage et une partie de test, afin d'éviter un apprentissage excessif et une dépendance à la même base de données, où nous prenons des valeurs aléatoires. à partir de la base de données pour les deux parties, mais la division ne peut pas se faire. Au hasard, mais selon les termes et conditions :

■ Holdout method

Divisez la base de données en taille X en deux parties, la première partie pour apprentissage 80% ; Le deuxième pour le test 20%.

3) Utilisation

Et il représente la dernière étape et en même temps l'étape la plus importante de notre projet après avoir obtenu le meilleur taux qui nous permet d'utiliser ce modèle, car nous l'utilisons pour prédire la catégorie à laquelle appartient ce médicament grâce à de nouvelles données et informations inconnu à l'avance comme le montre figure suivante :

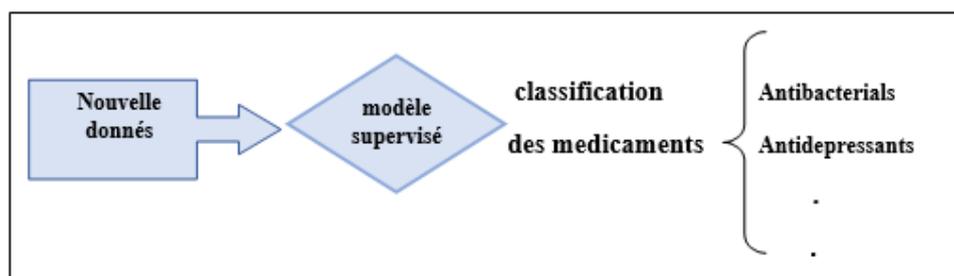


FIGURE 3.5 – Utilisation

3.4 Fonctionnement du système

Afin de rendre notre système plus clair, nous avons dessiné les diagrammes sur le langage UML qui expliquent le fonctionnement de l'application sont les suivants :

Au début, nous avons expliqué que l'utilisateur peut être un professionnel ou un débutant, où s'il est novice et n'a pas d'informations sur une drogue en particulier, il doit suivre une formation dans le domaine de la drogue. professionnel, il lui suffit de saisir le nom du médicament et l'ID du médicament. Il est spécifique à chaque médicament et il identifiera ensuite la catégorie thérapeutique de ce médicament.

3.4.1 Diagramme de classes

Figure 3.6 Schéma de la relation entre les environnements dans notre application.

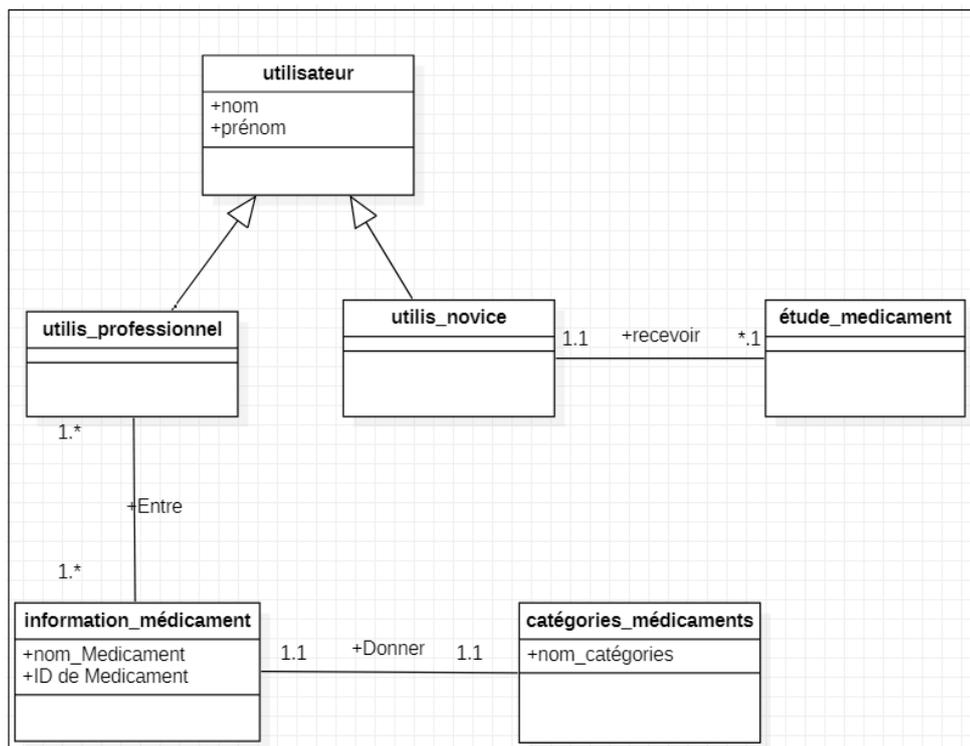


FIGURE 3.6 – Diagramme de classes UML pour l'application

3.4.2 Diagramme de Séquence

Figure 3.7 Explique comment fonctionne notre application .

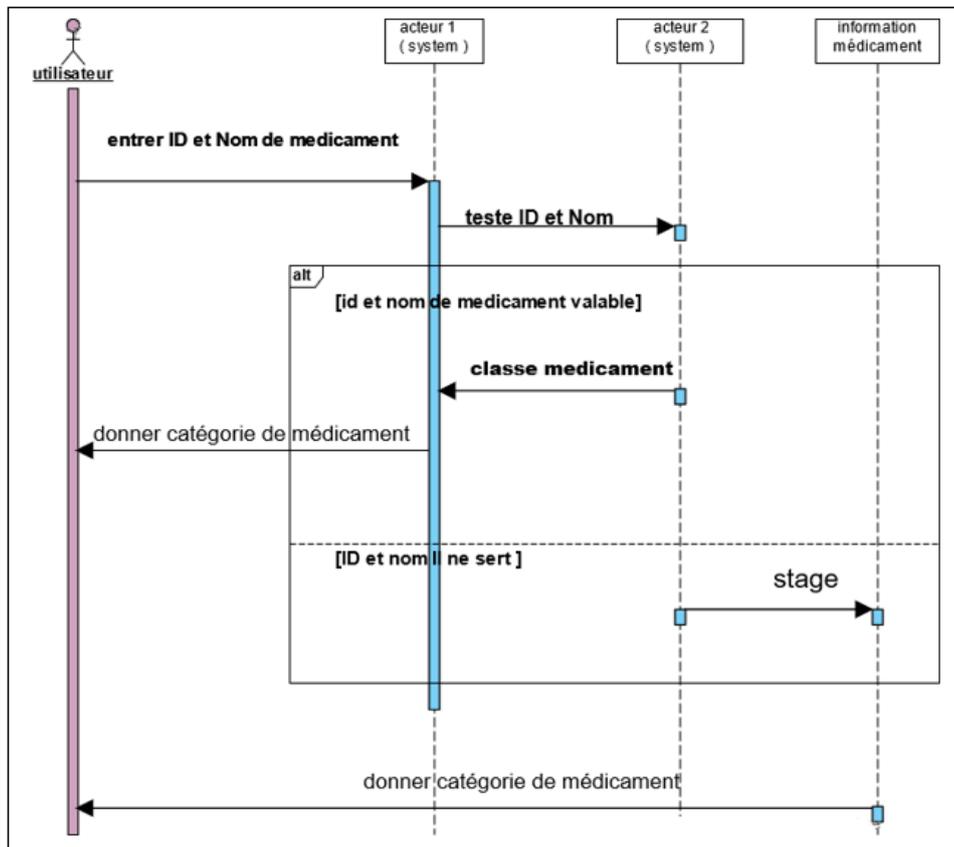


FIGURE 3.7 – diagramme de Séquence UML pour l'application

3.4.3 diagramme de cas d'utilisation

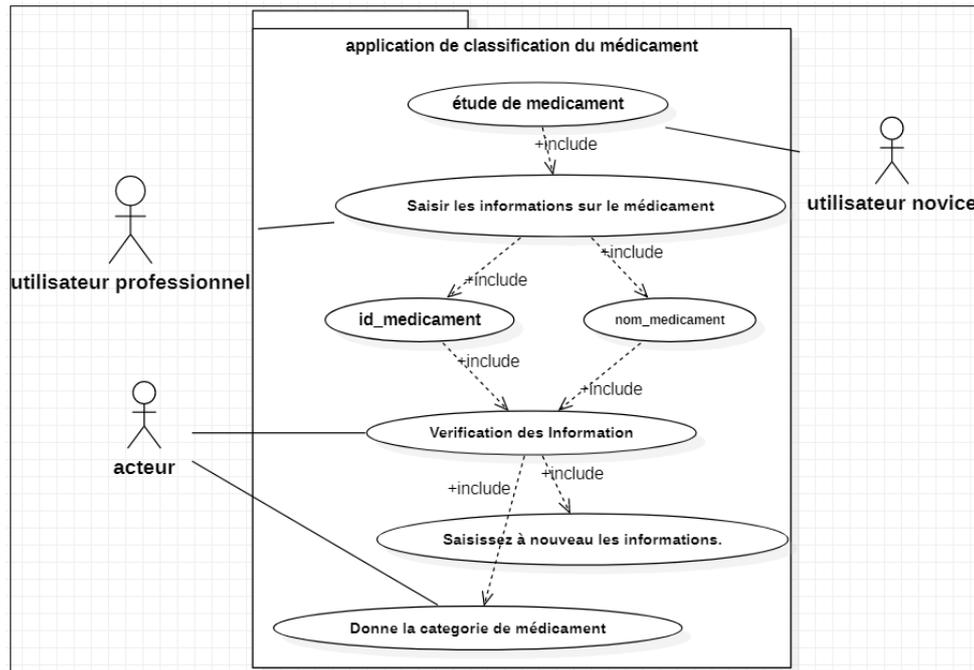


FIGURE 3.8 – diagramme de cas d'utilisation UML pour l'application

3.5 Conclusion

Dans ce chapitre, nous avons essayé de présenter toutes les étapes que nous avons prises pour concevoir un système de classification des médicaments par catégorie, et dans le chapitre suivant, nous expliquerons et décrirons le processus de mise en œuvre de notre système que nous avons obtenu.

Chapitre 4

Implémentation

Sommaire

4.1	Introduction	46
4.2	Environnement de développement	46
4.3	Bibliothèques utilisées	48
4.4	Datase	51
4.5	Application	55
4.6	Conclusion	63

4.1 Introduction

Dans ce chapitre, nous allons mettre en œuvre notre système qui a été obtenu par le fonctionnaire chargé de classer les médicaments selon la catégorie thérapeutique, avec une explication détaillée des moyens, programmes, langages et bibliothèques les plus importants que nous avons utilisés pour obtenir le programme avec une présentation du dossier de candidature et une discussion des résultats.

4.2 Environnement de développement

Dans notre environnement de travail nous avons utilisé anaconda pour lancer Spyder pour le traitement, la formation, la validation des résultats que nous avons obtenus et la création de l'application.

1) Anaconda

Nous avons utilisé Anaconda Navigator comme interface utilisateur graphique (GUI) de bureau intégrée à Anaconda®, nous exécutons des applications et utilisons des packages, des environnements et des canaux conda sans utiliser de commandes

de ligne de commande .[17]

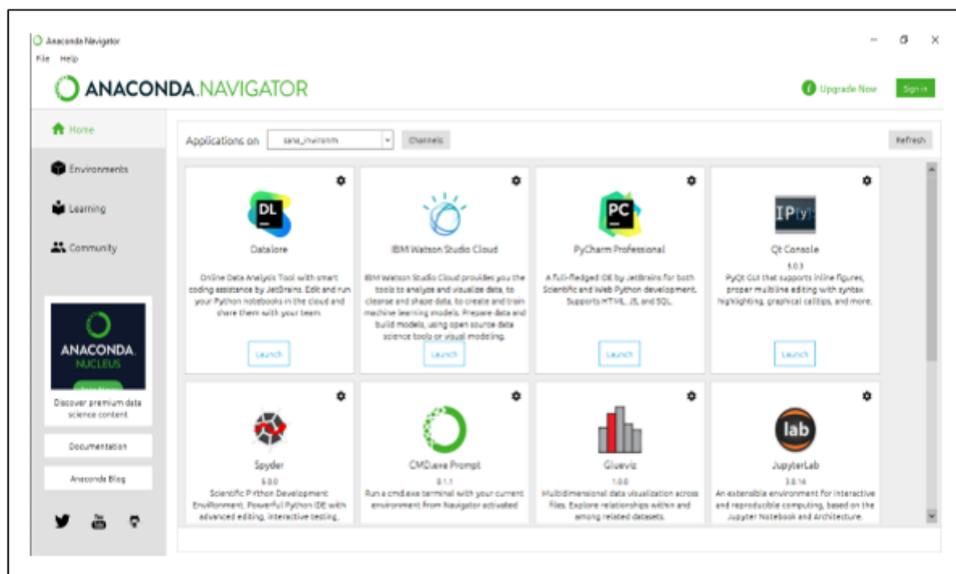


FIGURE 4.1 – Navigateur Anaconda

2) Spyder

Spyder (PYthon Scientific Development Environment) pour une utilisation scientifique du langage Python, avec des fonctionnalités avancées et l'intégration de plusieurs bibliothèques à usage scientifique .[18]

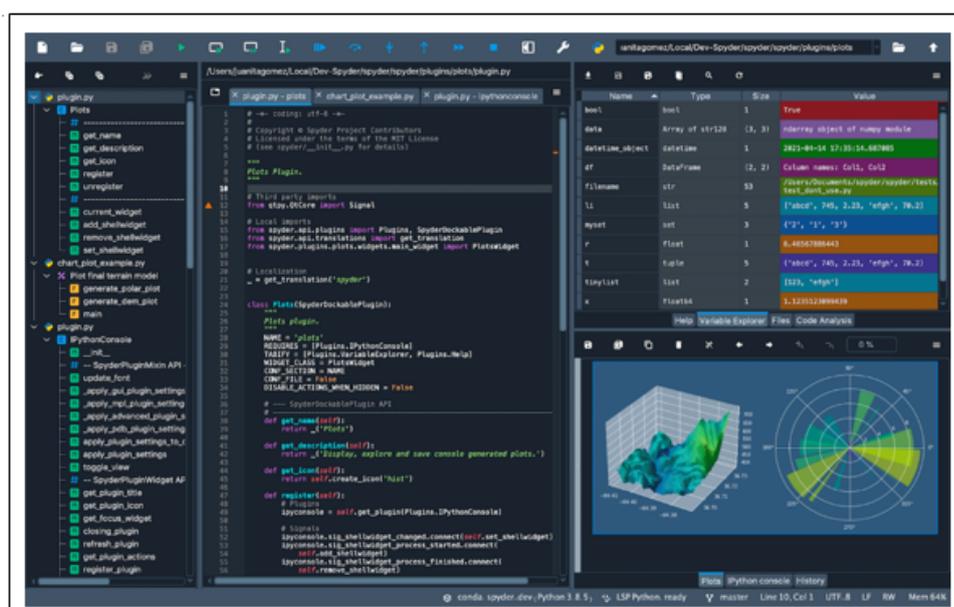


FIGURE 4.2 – Anaconda

Afin de préparer notre programme, nous avons utilisé le langage Python3 pour manipuler la base de données et pour mettre en place l'interface de l'application, nous avons utilisé des bibliothèques spéciales dans le langage Python.

3) Python

Python est un langage de programmation (au même titre que le C, C++, fortran, java..), développé en 1989.

Et ce sont quelques-unes de ses caractéristiques :

- importante quantité des bibliothèques disponibles : pour le calcul scientifique, les Statistiques, les bases de données, la visualisation .
- grande portabilité : indépendant vis à vis du système d'exploitation (linux, windows, Ma-cOS) .
- présente un support pour l'intégration d'autres langages. [19]

4.3 Bibliothèques utilisées

A travers les étapes de réalisation de notre projet, nous avons eu besoin d'intégrer certaines bibliothèques dans la première étape pour traiter et diviser la base de données que nous avons utilisée 'sklearn (scikit-learn)', 'numpy', 'pandas', puis dans la deuxième étape nous avons entraîné nos données en utilisant l'algorithme d'une machine à vecteurs une bibliothèque 'sklearn.svm' et pour dessiner les schémas nous avons utilisé 'matplotlib', 'seaborn', 'mlxtend' et à la dernière étape nous avons utilisé une bibliothèque 'tkinter' pour l'interface de l'application .

1) sklearn (scikit-learn)

Algorithmes pour les tâches d'apprentissage automatique, y compris la classification et la régression. Il fournit également des modules pour l'extraction de caractéristiques, le traitement des données et l'évaluation de modèles. [20]

2) sklearn (Numpy)

Le package de base pour le calcul scientifique en Python, c'est une bibliothèque Python qui fournit un objet tableau multidimensionnel, y compris la manipulation mathématique et logique de la forme, le tri et la sélection. [21]

3) Pandas

Une bibliothèque open source sous licence BSD de structures de données hautes performances et faciles à utiliser et d'outils d'analyse de données conçus pour le langage de programmation Python. [22]

```

import numpy as np
import pandas as pd
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import LabelEncoder
from sklearn.datasets import make_blobs
from sklearn.svm import SVC

import matplotlib.pyplot as plt
from matplotlib.pyplot import plt
import seaborn as sns
from mlxtend.plotting import plot_decision_regions
import tkinter as tk
import requests

HEIGHT = 800
WIDTH = 800

def get_cat(predict_medic):
    try:
        categorie='categorie de médicaments: \n %s' % (predict_medic)
    except:
        categorie='IL y avait un problème '
    return categorie

def gete_category(drug, drug_id):
    x_predict= [[drug, drug_id]]
    x_predict = pd.DataFrame(x_predict).values

    dataset = pd.read_csv('drug.csv')
    x = dataset.iloc[:, :-1].values
    y = dataset.iloc[:, -1].values
    x = np.concatenate([x,x_predict])

    ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(sparse=False), [0]), remainder='passthrough'])
    x = np.array(ct.fit_transform(x))
    ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [1497]), remainder='passthrough'])

```

FIGURE 4.3 – sklearn avec algorithme

4) sklearn.svm

Il s'agit d'une bibliothèque d'apprentissage automatique à usage général écrite en Python. Fournit des implémentations efficaces des derniers algorithmes. [23]

```

#Classifier training using Support Vector Machine
from sklearn.svm import SVC
model = SVC()
model.fit(x_train,y_train)
print("pourcentage : ",accuracy_score(y_test, predict_medic))
predict_medic = model.predict([x[-1,:]])
predict_medic = le.inverse_transform([0])

```

FIGURE 4.4 – bibliothèque SVM

5) Matplotlib

L'une des bibliothèques Python les plus utilisées pour représenter des graphiques 2D. Il peut produire une variété de graphiques et est de haute qualité. [24]

6) seaborn

fournit une interface de haut niveau pour dessiner des graphiques statistiques attrayants .[25]

7) **Mlxtend**

module du langage de programmation Python qui implémente divers algorithmes, ainsi que des utilitaires pour l'apprentissage automatique et l'exploration de données. L'interface de programmation d'applications (API). [26]

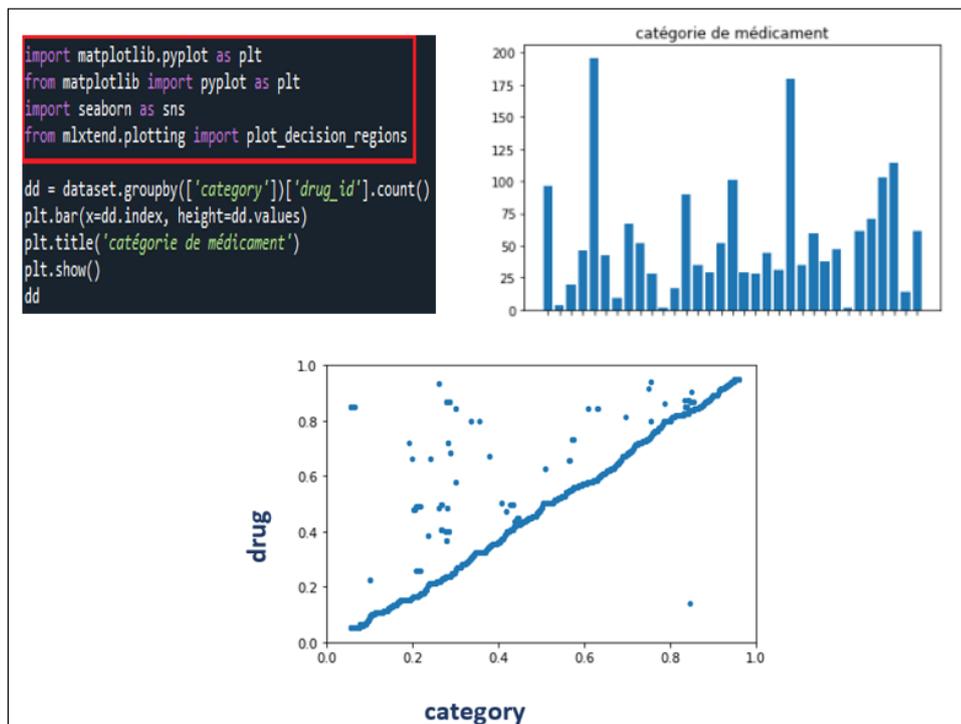


FIGURE 4.5 – bibliothèque graphique

8) **Tkinter**

Le module Tkinter ("interface Tk") est l'interface Python standard de la boîte à outils Tk GUI de Scriptics (anciennement développée par Sun Labs). [27]

```

label['text'] = get_cat(predict_med)

import tkinter as tk
root = tk.Tk()
canvas = tk.Canvas(root, height=HEIGHT, width=WIDTH)
canvas.pack()

root.title('application classification des médicaments')

background_image = tk.PhotoImage(file='C:\\Users\\sana\\Desktop\\1122\\opo.png')
background_label = tk.Label(root, image=background_image)
background_label.place(relwidth=1, relheight=1)

root.resizable(False, False)
root.configure(background='green')

#####
frame = tk.Frame(root, bg='#B3DAFF', bd=10)
frame.place(relx=0.5, rely=0.06, relwidth=0.80, relheight=0.08, anchor='n')

label=tk.Label (frame, text="Drug" , bg='#B3DAFF' , font=("Helvetica" , 10))
label.place(relwidth=0.20, relheight=1)

entry1 = tk.Entry(frame, font=40)
entry1.place(relx=0.3,relwidth=0.65,relheight=1)

#####
frame2 = tk.Frame(root, bg='#B3DAFF', bd=10)
frame2.place(relx=0.5, rely=0.12, relwidth=0.80, relheight=0.08, anchor='n')

label2=tk.Label (frame2, bg='#B3DAFF', text="ID de Médicaments" ,font=("Helvetica" , 10))
label2.place(relwidth=0.2, relheight=1)

```

FIGURE 4.6 – bibliothèque Tkinter

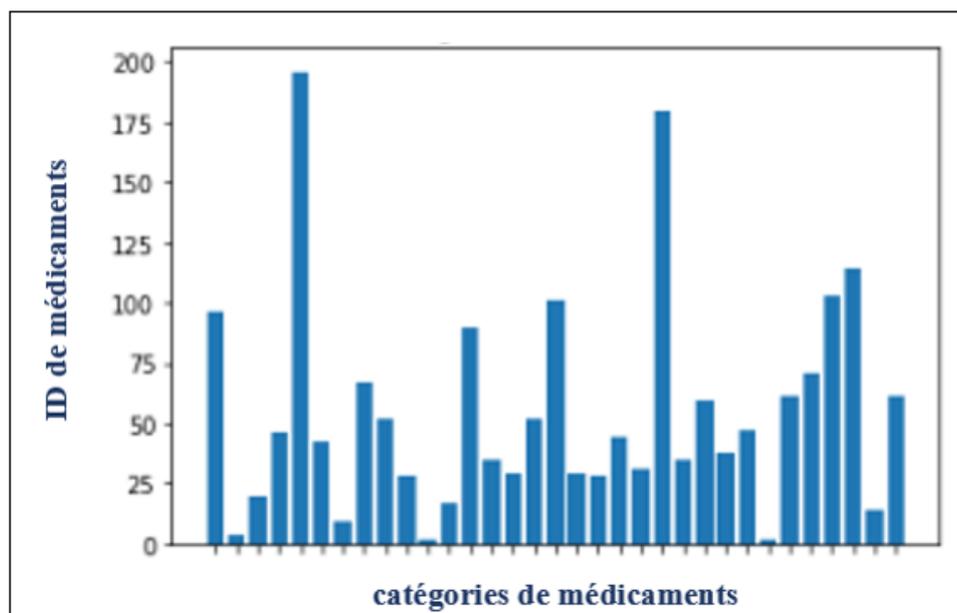
4.4 Dataset

Puisque nous sommes dans un système qui utilise l'apprentissage automatique, nous avons besoin d'une base de données figure 4.7 pour l'entraîner sur les données en utilisant des algorithmes pour prédire et tester une classe de données.

	A	B	C	D	E
1	drug	drug_id	drug_actif	class_drug	category_drug
2	celecoxib	D00567	celecoxib	Nonsteroidal Anti-inflammatory Drugs	Analgesics
3	diclofenac	D07816	diclofenac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
4	diclofenac potassium	D00903	diclofenac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
5	diclofenac sodium	D00904	diclofenac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
6	diclofenac diethylamine	D07817	diclofenac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
7	diclofenac hydroxyethylpyrrolidine	D07818	diclofenac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
8	diclofenac calcium	D07819	diclofenac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
9	diflunisal	D00130	diflunisal	Nonsteroidal Anti-inflammatory Drugs	Analgesics
10	etodolac	D00315	etodolac	Nonsteroidal Anti-inflammatory Drugs	Analgesics
11	fenoprofen	D02350	fenoprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
12	fenoprofen calcium	D00968	fenoprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
13	flurbiprofen	D00330	flurbiprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
14	flurbiprofen axetil	D01475	flurbiprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
15	flurbiprofen sodium	D02290	flurbiprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
16	esflurbiprofen	D10254	flurbiprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
17	ibuprofen	D00126	ibuprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
18	ibuprofen piconol	D01122	ibuprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
19	ibuprofen aluminum	D04490	ibuprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
20	ibuprofen lysine	D06606	ibuprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics
21	ibuprofen arginine salt	D08058	ibuprofen	Nonsteroidal Anti-inflammatory Drugs	Analgesics

FIGURE 4.7 – Base de Données.csv

Le graphique de la Figure 4.8 montre l'évolution des classes de médicaments en termes de ID de médicaments.

FIGURE 4.8 – **diagramme de catégories**

Notre base de données que nous avons utilisée et préparée est constituée de données relatives aux médicaments, qui ont été enregistrées dans un fichier .CSV puis le fichier est divisé en deux parties, la première partie est pour la formation et la deuxième partie est pour la validation, puis nous avons déménagé au stade d'utilisation, qui est la prédiction pour connaître la classe à laquelle appartient le médicament, et ils sont les suivants :

```
[ 'Analgesics' 'Anesthetics'  
'Anti-Addiction/Substance Abuse Treatment Agents' 'Antibacterials'  
'Anticonvulsants' 'Antidementia Agents' 'Antidepressants' 'Antiemetics'  
'Anti-inflammatory Agents' 'Antimigraine Agents' 'Antimyasthenic Agents'  
'Antimycobacterials' 'Antineoplastics' 'Antiparasitics'  
'Antiparkinson Agents' 'Antipsychotics' 'Antivirals' 'Anxiolytics'  
'Bipolar Agents' 'Blood Glucose Regulators'  
'Blood Products/Modifiers/Volume Expanders' 'Cardiovascular Agents'  
'Central Nervous System Agents' 'Gastrointestinal Agents'  
'Genitourinary Agents'  
'Hormonal Agents, Stimulant/Replacement/Modifying (Sex Hormones/Modifiers)'  
'Hormonal Agents, Suppressant (Thyroid)' 'Immunological Agents'  
'Inflammatory Bowel Disease Agents' 'Ophthalmic Agents'  
'Respiratory Tract/Pulmonary Agents' 'Sleep Disorder Agents'  
'Therapeutic Nutrients/Minerals/Electrolytes']
```

FIGURE 4.9 – **Catégorie de médicament**

D'après ce qui est à la hauteur de nos travaux et sur la base de nos recherches, il n'existe pas de base de données adaptée pour classer les médicaments selon la classe thérapeutique dont nous avons besoin, et c'est ce qui nous a poussé à nous débarrasser de certains

éléments indésirables pour disposer de notre propre base de données.

Ce que l'on tire de Kaggle qui est un site américain créé en 2010 qui permet aux amateurs comme aux professionnels de s'exercer seul ou en équipe sur des challenges de machine learning proposés par les entreprises ou le site Kaggle lui-même. [28]

Identification de la classe à laquelle appartient le médicament : Contient des données descriptives pour un médicament particulier ('Analgésiques' 'Anesthésiques' 'Antibactériens' 'Antiémétiques'...).

Comme le montre l'image Figure 4.7, nos données sont des données textuelle C'est pourquoi nous avons utilisé un algorithme (Figure 4.10) spécial pour les encoder (entraînement et validation) afin que Python puisse comprendre notre entrée .

```
def gete_category(drug,drug_id):
    x_predict= [[drug,drug_id]]
    x_predict = pd.DataFrame(x_predict).values

    dataset = pd.read_csv('drug.csv')

    x = dataset.iloc[:, :-1].values
    y = dataset.iloc[:, -1].values
    x = np.concatenate([x,x_predict])

    ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(sparse=False), [0]), remainder='passthrough'])
    x = np.array(ct.fit_transform(x))
    ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [1497]), remainder='passthrough'])
    x = np.array(ct.fit_transform(x))

    ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(), [3]), remainder='passthrough'])
    x = np.array(ct.fit_transform(x))

    le = LabelEncoder()
    y = le.fit_transform(y)
```

FIGURE 4.10 – algorithme de codage de base

Et le résultat de codage était comme le montre Figure 4.12 et 4.7

The screenshot shows a Jupyter Notebook window titled 'x_train - NumPy object array (read only)'. It displays a 19x13 matrix of binary data. The columns are indexed 0 to 12, and the rows are indexed 0 to 18. The data is as follows:

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
12	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FIGURE 4.11 – resultat du codage entraînement

The screenshot shows a Jupyter Notebook window titled 'x_train - NumPy object array (read only)'. It displays a 19x13 matrix of binary data, identical to the one in Figure 4.11. The data is as follows:

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
12	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
18	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

FIGURE 4.12 – resultat du codage et validation

Après avoir pré-traité nos données et testé le modèle, nous avons obtenu le meilleur taux de précision qui était de 80% .

```
#Classifier training using Support Vector Machine
model = SVC()
model.fit(x_train,y_train)

print("Model accuracy : ",accuracy_score(y_test, predict_med))

In [3]: runfile('C:/Users/sana/Desktop/nan/drugs_model.py', wdir='C:/Users/sana/Desktop/nan')
Model accuracy : 0.8060941828254847
```

FIGURE 4.13 – Précision du modèle

4.5 Application

Comme nous l'avons évoqué dans les chapitres précédents, notre objectif est d'exploiter les données afin qu'il soit plus facile pour le pharmacien de classer les médicaments et d'identifier leur catégorie thérapeutique. Par conséquent, après avoir préparé les données nécessaires, nous avons créé une application qui correspond à ces données et répond au désir du pharmacien.

La Figure 4.14 montre la fenêtre qui nous apparaît au démarrage de l'application.

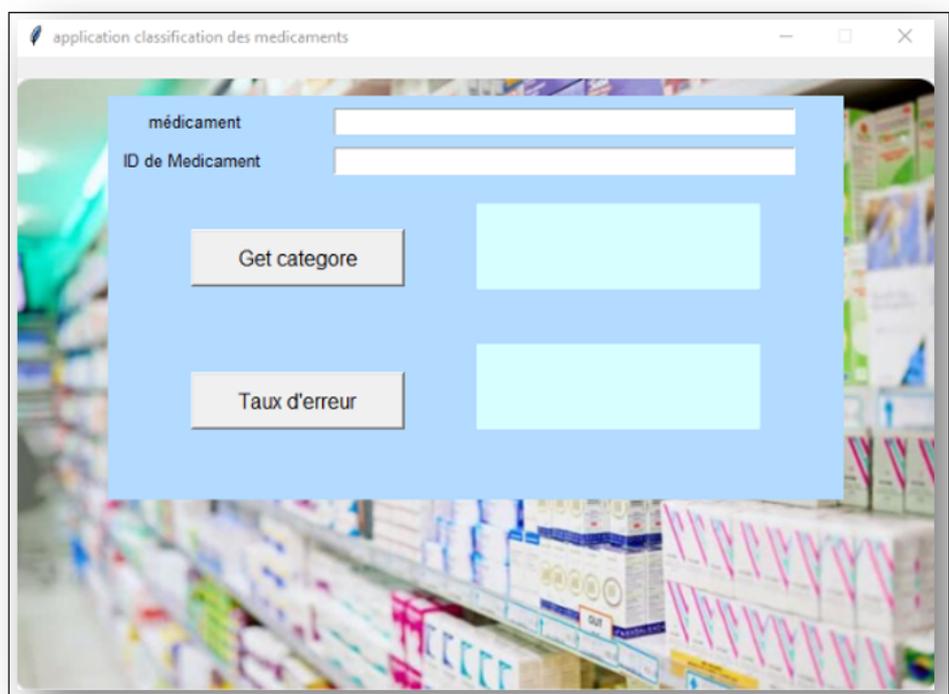


FIGURE 4.14 – L'interface principale de l'application

4.5. Application

Après cela, l'utilisateur est tenu de saisir le nom du médicament qu'il souhaite déterminer sa catégorie thérapeutique en précisant l'identifiant de ce médicament, sachant que chaque médicament a son propre identifiant qui le distingue du reste des médicaments comme indiqué dans Figure 4.15 et 4.16 .

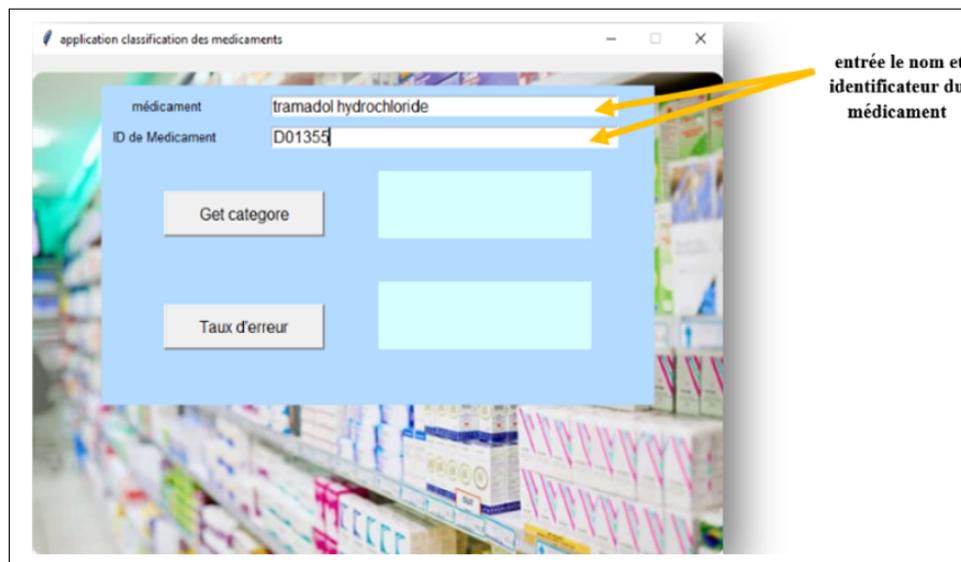


FIGURE 4.15 – Première entrée utilisateur

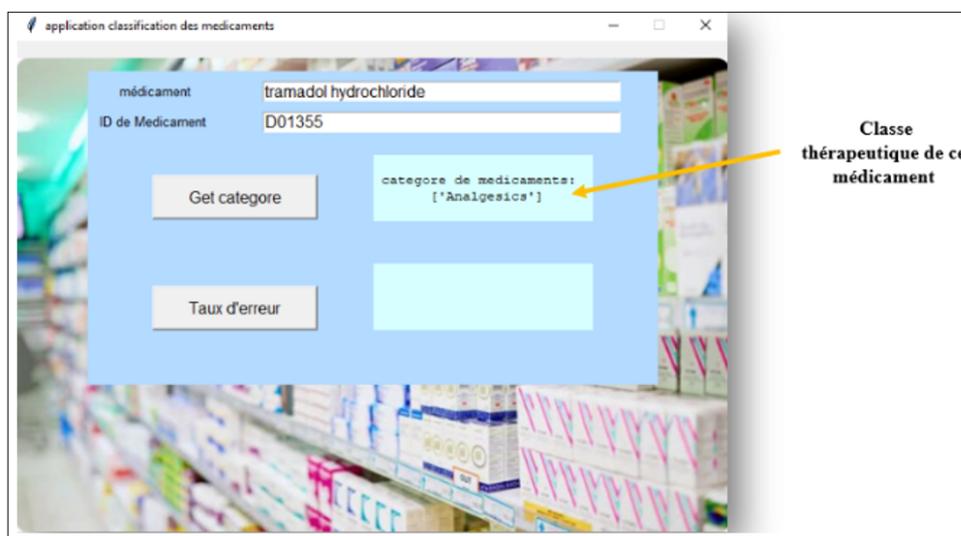


FIGURE 4.16 – Première entrée utilisateur

4.5. Application

Afin de déterminer à quel point nous pourrions nous tromper dans notre résultat, nous avons utilisé le taux d'erreur comme indiqué dans Figure 4.17 .

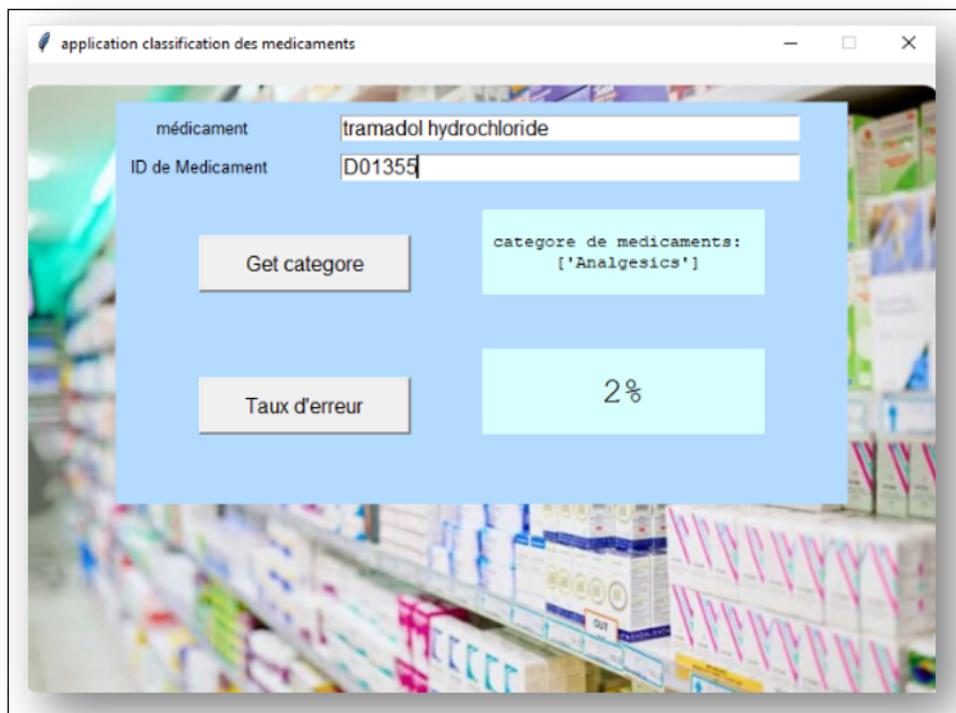


FIGURE 4.17 – Explication du taux d'erreur

Code Source

```
import numpy as np
import pandas as pd
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import LabelEncoder
from sklearn.datasets import make_blobs
from sklearn.svm import SVC
import matplotlib.pyplot as plt
from matplotlib import pyplot as plt
import seaborn as sns
from mlxtend.plotting import plot_decision_regions
import tkinter as tk
import requests
import math

HEIGHT = 500
WIDTH = 700

def get_cat():
    taux=format(math.ceil(number))
```

label2=taux

def get_cat(predict_medic):

try:

categori='categori de medicaments: \n %s' % (predict_medic)

except:

categori='Il y avait un problème '

return categori

def gete_category(drug,drug_id):

x_predict= [[drug,drug_id]]

x_predict = pd.DataFrame(x_predict).values

dataset = pd.read_csv('drug.csv')

x = dataset.iloc[:, :-1].values

y = dataset.iloc[:, -1].values

x = np.concatenate([x,x_predict])

**ct = ColumnTransformer(transformers=[('encoder',
OneHotEncoder(sparse=False), [0]), remainder='passthrough')**

x = np.array(ct.fit_transform(x))

**ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(),
[1497]), remainder='passthrough')**

x = np.array(ct.fit_transform(x))

**ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(),
[3]), remainder='passthrough')**

```
x = np.array(ct.fit_transform(x))
le = LabelEncoder()
y = le.fit_transform(y)
#split data
x_d = x[:-1,:]
x_train, x_test, y_train, y_test = train_test_split(x_d, y, test_size = 0.2,
random_state = 1)

#Classifier training using Support Vector Machine
model = SVC()
model.fit(x_train,y_train)
#print("Model accuracy : ",accuracy_score(y_test, predict_medic))
predict_medic = model.predict([x[-1,:]])
predict_medic = le.inverse_transform([0])
#print(predict_medic)
#return predict_medic
label['text']= get_cat(predict_medic)

import tkinter as tk
root = tk.Tk()
canvas = tk.Canvas(root, height=HEIGHT, width=WIDTH)
canvas.pack()
root.title('application classification des medicaments')
```

```
background_image =
tk.PhotoImage(file='C:\\Users\\sana\\Desktop\\122\\opo.png')
background_label = tk.Label(root, image=background_image)
background_label.place(relwidth=1, relheight=1)
root.resizable(False,False)
root.configure(background='green')
#####
frame = tk.Frame(root, bg='#B3DAFF', bd=10 )
frame.place(relx=0.5, rely=0.06, relwidth=0.80, relheight=0.08, anchor='n')
label=tk.Label (frame, text=" médicament" , bg='#B3DAFF' ,
font=("Helvetica", 10))
label.place(relwidth=0.20, relheight=1)
entry1 = tk.Entry(frame, font=40)
entry1.place(relx=0.3,relwidth=0.65,relheight=1)
#####
frame2 = tk.Frame(root, bg='#B3DAFF', bd=10 )
frame2.place(relx=0.5, rely=0.12, relwidth=0.80, relheight=0.08, anchor='n')
label2=tk.Label (frame2, bg='#B3DAFF', text='ID de Medicament'
,font=("Helvetica", 10))
label2.place(relwidth=0.2, relheight=1)
entry2 = tk.Entry(frame2, font=40)
entry2.place(relx=0.3,relwidth=0.65,relheight=1)
#####
frame5 = tk.Frame(root, bg='#B3DAFF', bd=10 )
frame5.place(relx=0.5, rely=0.20, relwidth=0.80, relheight=0.08, anchor='n' ,
height=200)
```

```
button = tk.Button(frame5, text="Get categorie", font=40 , command=lambda:  
gete_category(entry1.get(),entry2.get() )
```

```
button.place(relx=0.1, relheight=0.2, relwidth=0.3 , rely=0.1)
```

```
label=tk.Label (frame5,bg='#D8FFFF', font=('courier',10))
```

```
label.place(relx=0.5,relwidth=0.4, relheight=0.3, rely=0.01)
```

```
#####
```

```
button = tk.Button(frame5, text="Taux d'erreur", font=40,  
command=lambda:taux_e())
```

```
button.place(relx=0.1, relheight=0.2, relwidth=0.3, rely=0.6 )
```

```
label2=tk.Label (frame5,bg='#D8FFFF', font=('courier',20))
```

```
label2.place(relx=0.5,relwidth=0.4, relheight=0.3, rely=0.5 )
```

```
root.mainloop()
```

4.6 Conclusion

Dans ce chapitre, nous avons expliqué en détail les différents outils utilisés pour mettre en œuvre notre projet et les résultats que nous avons obtenus sous forme de graphiques illustratifs, et nous avons fourni un exemple de classification des médicaments et comment l'application est mise en œuvre, et nous avons parlé de l'erreur taux.

Conclusion Générale

Ce mémoire présente une méthode d'exploration de données en classant les médicaments à l'aide d'une Support Vector Machine, et afin de déterminer plus précisément les meilleures propriétés auxiliaires pour la classification des médicaments.

Nous avons commencé à étudier l'exploration de données qui représente un ensemble de techniques pour extraire des données d'une très grande base de données.

Ensuite, nous avons conçu et mis en œuvre une solution basée sur l'utilisation de techniques de nettoyage, de codage et de techniques de codage pour pré-traiter la base de données que nous avons utilisée, puis extraire les propriétés, et enfin nous avons appliqué l'algorithme Support Vector Machine à notre base de données de médicaments pour construire un modèle pour la classification des médicaments et l'identification des classes thérapeutiques.

Grâce aux recherches qui ont été menées au cours de cette étude, nous avons obtenu les résultats suivants :

- Les meilleures propriétés pour détecter la classe de médicaments sont dans l'ordre :ID médicamente , médicament actif,classe de médicamente.
- Malgré la grande quantité de données, l'algorithme Support Vector Machine donne le résultat en moins de temps.
- Chaque fois que nous utilisons des données différentes, le taux d'erreur reste entre 2% et 3.5%.

Le travail que nous avons effectué peut être étendu en étudiant avec un ensemble de données plus important, en utilisant le data mining en tant que domaine riche en sujets de recherche prospective et en développant son apprentissage à l'aide d'autres algorithmes d'apprentissage automatique.

Bibliographie

- [1] Abdelhamid DJEFFAL. *Fouille de données avancée*. 2014.
- [2] P CABENA et al. “Discovering data mining from concept to implementation–Prentice Hall PTR”. In : *Inc. USA* (1998).
- [3] EG TALBI. “Fouille de données (Data Mining) : Un tour d’horizon”. In : *Laboratoire d’informatique de Lille* (2008).
- [4] Djazia CHAMI. *Une plate forme orientée agent pour le data mining thèse de magister*. 2010.
- [5] J BOULICAUT et al. “Knowledge Discovery in Databases : PKDD 2004”. In : *8th European Conference on Principles and Practice of Knowledge Discovery in Databases, Pisa, Italy*. Springer. 2004.
- [6] Dr.Salim BITAM. “Entrepôts et préparation de données”. In : (2020).
- [7] Dr DJEROLEILA. “Algorithmes génétiques”. In : (2018).
- [8] Marc BENINCA. “Machine à Vecteurs de Support”. In : (2018).
- [9] Pr B BELDJILALI. “Système d’Aide à la Distribution des Produits Pharmaceutiques en Algérie en utilisant le Data Mining”. Thèse de doct. Université d’Oran, 2010.
- [10] James SWARBRICK. “Encyclopedia of Pharmaceutical Technology, pdf”. In : (2015/2016).
- [11] Mr FATMI SOFIANE. *Procèdes pharmaceutiques*. 2015/2016.
- [12] OR ZAIANE. “Principles of Knowledge Discovery in Databases, CMPUT690”. In : *University of Alberta, Department of Computing Science* (1999).
- [13] Parasuram BALASUBRAMANIAN. “Optimal resource utilization in communicable disease control with special reference to syphilis control.” Thèse de doct. Purdue University, 1977.
- [14] A BELAICHE. “Le système ATC-DDD : intérêt en pharmacovigilance [Internet]. 2010 [cité le 06 dec 2015]”. In : *Disponible sur : http://www.who.int/medicines/areas/quality_safety/safety_efficacy/trainingcourses/1_1atc_ddd.pdf* ().

- [15] ouissam kichou amina ouissam, kamelia manceur kamelia et menza ouyed menza. “Profil de la consommation des antibiotiques à usage systémique en milieu hospitalier : Expérience du CHU de Tizi-Ouzou”. In : (2020).
- [16] Hadeer AHMED, Issa TRAORE et Sherif SAAD. “Detection of online fake news using n-gram analysis and machine learning techniques”. In : *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*. Springer. 2017, p. 127-138.
- [17] Vasudevan LAKSHMINARAYANAN et al. *Understanding Optics with Python*. CRC Press, 2018.
- [18] Asma REMACHE et Adenane HIDOUCI. “Conception et proposition de modele de risque de crédit”. In : (2019).
- [19] *Introduction à la programmation en langage Python, U. Paris-Sud*. 2015-2016.
- [20] Gavin HACKELING. *Mastering Machine Learning with scikit-learn*. Packt Publishing Ltd, 2017.
- [21] *NumPy User Guide, Release 1.20.0*. 31/01/2021.
- [22] S.RAJAMOHAN. *Introduction to Python Pandas for DataAnalytics*. 19/07/2016.
- [23] Alexandre ABRAHAM et al. “Machine learning for neuroimaging with scikit-learn”. In : *Frontiers in neuroinformatics* 8 (2014), p. 14.
- [24] *Introduction aux graphiques en Python avec matplotlib.pyplot*. 2020.
- [25] Katia OLEINIK. “Python for Data Analysis”. In : ()
- [26] Bagus Prabowo AJI et Muh Arif WIBISONO. “Strategi Pengambilan Keputusan Penjualan Dalam Rangka Optimasi Profit Industri Ritel Berbasis Unsupervised Machine Learning Algorithm (Studi Kasus Modern Minimarket-X)”. In : (2018).
- [27] Fredrik LUNDH. “An introduction to tkinter”. In : *URL : [www. pythonware. com/library/tkinter/introduction/index. htm](http://www.pythonware.com/library/tkinter/introduction/index.htm)* (1999).
- [28] *Rapport de projet de spécialité Challenge Kaggle 4*. 2016.

