

Republique Algérienne Democratique Et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université Larbi Tébessi– Tébessa –

Faculté des Sciences exactes et des sciences de la nature et de la vie

Département de Biologie Appliquée

Spécialité : Biologie Moléculaire et Cellulaire.

## MEMOIRE

Présenté pour l'obtention du diplôme de Master Académique

Présentées par :

- **Azzaz Soumia**

- **Namous Rafika**

thème :

**Étude comparative des algorithmes d'apprentissage  
de la machine dans la prédiction et diagnostic du  
cancer de sein.**

Présenté et soutenu publiquement, le 15/06/2022, devant le jury composé de :

Dr. GASSMI Salim	MCA	Président
Dr. YAHIAOUI Ayoub	MCB	promoteur
Dr. KHEDIRI Abderrazek	MCA	Examineur (invite)
Dr.MENASRIA Taha	MCA	Examineur

Promotion : 2021/2022

## **REMERCIEMENT**

**Tout d'abord, on remercie DIEU qui nous a permis de mener à bien ce modeste recherche scientifique et qui nous a inspirés par la force, la patience, la détermination et surtout la bonne santé sous l'estime de son prophète MOHAMMED(paix et salut sur lui).**

**En second lieu, on tient de remercier nos chers parents pour toute l'aide et tout le soutien qu'ils ont nous apporter durant les années d'étude et nos familles qui nous avoir toujours encouragées.**

**Merci Monsieur YAHIAOUI AYOUB Nous vous sommes très reconnaissantes d'avoir accepté et ce malgré vos innombrables tâches. Cela nous touche énormément e humblement nous vous disons merci. Nous avons eu l'immense honneur de bénéficier de vos enseignements théoriques et l'encadrement. Vous nous avez toujours impressionnés par la clarté et l'aisance avec laquelle vous savez transmettre votre savoir. Votre grande simplicité et votre modestie font de vous un homme admiré et respecté de tous.**

**Être notre encadreur est un grand privilège que vous nous faites et nous vous en sommes reconnaissantes.**

**Trouvez ici, cher encadreur, notre profond respect e toute notre gratitude.**

**Que dieu vous comble de ses bénédictions au-delà de vos attentes.**

**On tient de remercier toutes les personnes qu'ont vieille à notre formation :**

**On remercie également toutes les équipes du service de cancérologie de l'hôpital bacaria.**

**A notre promotion master 2 biologie moléculaire et cellulaire 2017-2022**

**Nos remercient vont également aux honorables membres de jury d'avoir accepté de donner de leur temps pour évaluer notre travail et de leur présence à notre soutenance.**

**Enfin, on adresse nos remerciements les plus sincère aux personnes qui nous ont apporté leur aide de près ou de loin à la réalisation de ce mémoire.**

## **DEDICACES**

**Je dédie ce travail**

**A mes parents**

**A ma famille**

**A mes sœurs**

**A toutes mes amies**

**A tous ceux qui m'ont aidé dans mes études.**

**A tous mes proches, et tous ceux qui m'aiment.**

**soumia**

## **Dedicace**

**Je dédié ce modeste travail à:**

**A l'homme de ma vie ,mon exemple éternel, mon soutien moral et source de joie et de bonheur,celui qu 'est toujours sacrifié pour me voir réussir ,que dieu te garde dans son vaste paradis à toi : mon père.**

**A la lumière de me jours ,la source de mes efforts ,la flamme de mon coeur ,ma vie et mon bonheur:maman que j'adore.**

**C'est un moment de plaisir de dédier cet ouvre ,à més belles soeurs :zienb et Afaf,Warda ,Malika et spécialement souraya pour d'amour,de reconnaissance et de gratitude pour la dévouement et les sacrifes dont vous avez fait toujours preuve à mon égard.**

**A mes petites enfants de mes soeurs ,taiba,Nada,houdaifa,Acherf,Adem,sajida,Ali.**

**il n'ya pas d'ascenseur pour la réussite ,il faut prendre l'escalier.**

**rafika**

**SOMMAIRE**

**REMERCIEMENT** ..... V

**DEDICACES** ..... V

**SOMMAIRE** ..... I

**Liste des figures** ..... IV

**Liste des tableaux** ..... V

**Liste des abréviations** ..... VI

**Résumé** ..... VII

**الملخص** ..... IX

**Introduction Générale** ..... 1

**Problématique** ..... 2

**Organisation du mémoire:** ..... 2

**Chapitre01**

**Généralités sur le cancer de sein** ..... 3

**1.1.Introduction** ..... 4

**1.2.Définition du cancer de sein** ..... 5

**1.2.1Tumeurs bénignes :** ..... 5

**1.2.2.Tumeurs malignes :** ..... 6

**1.3.L'anatomie du sein :** ..... 6

**1.4.La classification TNM** ..... 7

**1.4.1.L'extension locale de la tumeur :** ..... 7

**1.4.2.Adénopathie régionale :** ..... 8

**1.4.3.Métastase à distance :** ..... 8

**1.5.Les types de cancer** ..... 9

**1.6.Les causes d'un cancer du sein :** ..... 10

**1.7.Les facteurs de risque et des symptômes :** ..... 10



## Sommaire

---

<b><u>1.7.1. Les facteurs de risques :</u></b> .....	10
<b><u>1.7.2. Les symptômes</u></b> .....	11
<b><u>1.8.Dépistage :</u></b> .....	12
<b><u>1.9.Diagnostic :</u></b> .....	14
<b><u>1.10.Les traitements médicaux du cancer du sein :</u></b> .....	15
<b><u>1.11.Conclusion :</u></b> .....	17

### Chapitre 2

<b><u>2.1.Introduction</u></b> .....	19
<b><u>2.2.Ou'est-ce que l'apprentissage automatique?</u></b> .....	20
<b><u>2.3.Définition de l'apprentissage automatique (machine Learning):</u></b> .....	20
<b><u>2.4.Techniques d'apprentissage automatique :</u></b> .....	20
<b><u>2.5.Types d'apprentissage automatique:</u></b> .....	21
<b><u>2.5.1.Apprentissage supervisé :</u></b> .....	22
<b><u>2.5.2. Apprentissage Non Supervisé :</u></b> .....	23
<b><u>2.6.Les algorithmes de l'apprentissage automatique utilisés :</u></b> .....	24
<b><u>2.6.1.SVM (Machines à vecteurs de support) :</u></b> .....	24
<b><u>2.6.2. la méthode des k plus proches voisins(KNN) :</u></b> .....	28
<b><u>2.6.3.Les forêts aléatoires ou les forêts de décision aléatoires :</u></b> .....	28
<b><u>2.6.4.L'arbre de décision</u></b> .....	29
<b><u>2.6.5.Naïve Bayes (Apprentissage supervisé-Classification):</u></b> .....	30
<b><u>2.7.Application de l'intelligence artificielle à la prédiction et au diagnostic en oncologie :</u></b> .....	31
<b><u>2.8.Intelligence artificielle VS apprentissage automatique :</u></b> .....	32
<b><u>2.9.Conclusion :</u></b> .....	34

### Chapitre 3 :

<b><u>3.1.Introduction :</u></b> .....	36
<b><u>3.2. Présentation des outils utilisés :</u></b> .....	36



## Sommaire

---

<b><u>3.3 Dataset:</u></b> .....	37
<b><u>1. La description des caractéristiques de dataset</u></b> .....	37
<b><u>2. Attribue de dataset :</u></b> .....	37
<b><u>3. Description des attributs:</u></b> .....	38
<b><u>4. Division de dataset :</u></b> .....	39
<b><u>3.4. Logiciel Utilisé</u></b> .....	39
<b><u>1. Pourquoi weka?</u></b> .....	40
<b><u>2. les avantages de weka:</u></b> .....	41
<b><u>3. Installation su Windows :</u></b> .....	41
<b><u>4. Description générale de l'interface de weka (GUI)</u></b> .....	42
<b><u>3.5. Les étapes de Processus :</u></b> .....	43
<b><u>3.6. Application des algorithmes :</u></b> .....	46
<b><u>3.7. Définition des outils statistiques utilisés :</u></b> .....	46
<b><u>3.8. Application et validation</u></b> .....	48
<b><u>1. Modèle 01 :</u></b> .....	48
<b><u>2. Modèle 02 :</u></b> .....	50
<b><u>3. Modèle 03:</u></b> .....	52
<b><u>4. Modèle 04:</u></b> .....	54
<b><u>5. Modèle 05 :</u></b> .....	56
<b><u>3.9. Évaluation des performances:</u></b> .....	58
<b><u>3.10. Comparaison des algorithmes :</u></b> .....	59
<b><u>3.11. Conclusion :</u></b> .....	60
<b><u>Conclusion générale</u></b> .....	61
<b><u>Références Bibliographies</u></b> .....	63



Liste des figures

<b>N FIGURE</b>	<b>NOM</b>	<b>PAGE</b>
<b>Figure 1.1</b>	<b>Division des cellules cancéreuses</b>	<b>6</b>
<b>Figure 1.2</b>	<b>Anatomie interne du sein</b>	<b>7</b>
<b>Figure 1.3</b>	<b>Stades du cancer du sein utilisation le système TNM</b>	<b>8</b>
<b>Figure 1.4</b>	<b>Types des cancers du sein</b>	<b>9</b>
<b>Figure 1.5</b>	<b>Les facteurs de risques du cancer du sein</b>	<b>11</b>
<b>Figure 1.6</b>	<b>Les symptômes du cancer du sein</b>	<b>12</b>
<b>Figure 1.7</b>	<b>Comment pratiquer L'autopalpation</b>	<b>13</b>
<b>Figure 1.8</b>	<b>Mammographie : dépistage du cancer du sein</b>	<b>14</b>
<b>Figure2.1</b>	<b>Intelligence artificielle vs. Machine Learning</b>	<b>20</b>
<b>Figure2.2</b>	<b>Types d'apprentissage automatique(Machine Learning)</b>	<b>21</b>
<b>Figure2.4</b>	<b>Apprentissage non supervisé</b>	<b>23</b>
<b>Figure2.5</b>	<b>principe de classification en machines à vecteurs de support(SVM)</b>	<b>24</b>
<b>Figure 2.6</b>	<b>Un simple exemple sur le fonctionnement de l'algorithme SVM</b>	<b>25</b>
<b>Figure 2.7</b>	<b>Hyperplan dans les entités 2D et 3D</b>	<b>25</b>
<b>Figure 2.8</b>	<b>Les vecteurs de support</b>	<b>26</b>
<b>Figure2.9</b>	<b>Marge dans l'algorithme SVM</b>	<b>26</b>
<b>Figure2.10</b>	<b>La Méthode de k plus proches voisines</b>	<b>27</b>
<b>Figure2.11</b>	<b>Forets aléatoires dans l'apprentissage automatique</b>	<b>29</b>
<b>Figure2.12</b>	<b>Un simple exemple sur l'algorithme random forest</b>	<b>29</b>
<b>Figure2.13</b>	<b>Arbre de décision</b>	<b>30</b>
<b>Figure3.14</b>	<b>Logo de weka</b>	<b>39</b>
<b>Figure 3.15</b>	<b>Courbe ROC de classificateur de SVM</b>	<b>48</b>
<b>Figure 3.16</b>	<b>Courbe ROC de classificateur de KNN</b>	<b>50</b>
<b>Figure3.17</b>	<b>Courbe Roc de classificateur de les forêts aléatoires</b>	<b>52</b>
<b>Figure3.18</b>	<b>Courbe Roc de classificateur d'arbre de décision</b>	<b>54</b>
<b>Figure3.19</b>	<b>Courbe Roc de classificateur de Naive bayes</b>	<b>56</b>



Liste des tableaux

<b>N TABLE</b>	<b>NOM</b>	<b>PAGE</b>
<b>Tableau 2.1</b>	<b>Différences entre l'apprentissage supervisé et non supervisé</b>	<b>23</b>
<b>Tableau 3.1</b>	<b>Les caractéristique de dataset</b>	<b>37</b>
<b>Tableau 3.2</b>	<b>Les attributs de la base de données « Breast-cancer-Wisconsin »</b>	<b>38</b>
<b>Tableau3.3</b>	<b>Résumé de l'Algorithme svm (machine à vecteurs de support)</b>	<b>49</b>
<b>Tableau3.4</b>	<b>mesures de précision pourSVM (machine à vecteurs de support)</b>	<b>49</b>
<b>Tableau3.5</b>	<b>Matrice de confusion pour l'Algorithme SVM (machine à vecteurs de support)</b>	<b>49</b>
<b>Tableau3.6</b>	<b>Résumé de l'Algorithme KNN (Méthode de K plus proches voisins)</b>	<b>51</b>
<b>Tableau3.7</b>	<b>mesures de précision pour KNN (Méthode de K plus proches voisins)</b>	<b>51</b>
<b>Tableau3.8</b>	<b>Matrice de confusion pour KNN (Méthode de K plus proches voisins)</b>	<b>51</b>
<b>Tableau3.9</b>	<b>Résumé de l'Algorithme les forêts aléatoires ou les forêts de décision aléatoires</b>	<b>52</b>
<b>Tableau3.10</b>	<b>mesures de précision pour les forêts aléatoires ou les forêts de décision aléatoires</b>	<b>53</b>
<b>Tableau3.11</b>	<b>Matrice de confusion pour l'Algorithme les forêts aléatoires ou les forêts de décision aléatoires</b>	<b>53</b>
<b>Tableau3.12</b>	<b>Résumé de l'Algorithme Arbre de décision (REPTree)</b>	<b>55</b>
<b>Tableau3.13</b>	<b>mesures de précision pour Arbre de décision (REPTree)</b>	<b>55</b>
<b>Tableau3.14</b>	<b>Matrice de confusion pour Arbre de décision (REPTree)</b>	<b>55</b>
<b>Tableau3.15</b>	<b>Résumé de l'Algorithme Naïve bayes</b>	<b>57</b>
<b>Tableau3.16</b>	<b>mesures de précision pour Naïve bayes</b>	<b>57</b>
<b>Tableau3.17</b>	<b>Matrice de confusion pour l'Algorithme Naïve bayes</b>	<b>57</b>
<b>Tableau 3.18</b>	<b>les résultats des attributs d'évaluations pour différents modèles</b>	<b>59</b>

## Liste des abréviations

<b>ML</b>	Machine Learning
<b>OMS</b>	Organisation mondiale de la santé
<b>TNM</b>	Tumeur,Nodes,métastase
<b>AS</b>	Apprentissage automatique
<b>ASN</b>	Apprentissage non supervisé
<b>SVM</b>	Machines à vecteurs de support
<b>IA</b>	Intelligence Artificielle

## Résumé

Le cancer de sein est une maladie courante qui touchent principalement les femmes, causée par la perturbation de certaines cellules qui se grandissent souvent et Formet une masse appelée « tumeur »

Récemment, avec l'avènement du machine Learning, de nombreuses études dans toutes les disciplines ont contribué à fournir des méthodes permettant de prédire cette maladie avant se développe, et basées sur l'intelligence artificielle et ses différentes branches,

Dans ce mémoire, nous avons appliqués les algorithmes d'apprentissage automatique supervisé (Random forest ,Decision tree , Naive bayes , Support vector Machine,K nearest neighbors) et le data set extrait du l'université de Californie pour la prédiction précoce de cancer de sein pour prendre de décisions en termes diagnostic et d'analyse afin de lutter contre le cancer de sein ,nous avons comparé les performances entre différentes algorithmes utilisé en fonction la précision et la sensibilité de modèle.

Les résultats expérimentaux montrent que les plus hauts obtenus par l'application de Naïve Bayes et vecteurs support machine sont respectivement 97% et 96%.

**Les mots clés:** L'intelligence artificielle, Machine Learning, prédiction de cancer de sein, Random forest , Decision tree , Naive bayes , Support vector Machine, K Nearset neighbors.

## Abstract

Breast cancer is a common disease that mainly affects women, caused by the disruption of certain cells that often grow and form a mass called a "tumor".

Recently, with the advent of machine learning, many studies in all disciplines have contributed to providing methods to predict this disease before it develops, and based on artificial intelligence and its various branches,

In this thesis, we applied the supervised machine learning algorithms (Randon forest, Decision tree, Naive bayes, Support vector Machine, K nearest neighbors) and the data set extracted from the University of California for the early prediction of cancer of breast to make decisions in terms of diagnosis and analysis in order to fight against breast cancer, we compared the performance of different algorithms used according to the accuracy and sensitivity of the model.

The experimental results show that the highest obtained by the application of Naïve Bayes and machine support vectors are respectively 97% and 96%.

**Keywords:** Artificial intelligence, Machine learning, breast cancerprediction,Randon forest, Decision tree, Naive bayes, Support vector Machine, K Nearset neighbors.

## الملخص

سرطان الثدي هو مرض شائع يصيب النساء بشكل رئيسي ، وينجم عن تمزق بعض الخلايا التي تنمو في كثير من الأحيان وتشكل كتلة تسمى "الورم".  
في الآونة الأخيرة ، مع ظهور التعلم الآلي ، ساهمت العديد من الدراسات في جميع التخصصات في توفير طرق للتنبؤ بهذا المرض قبل أن يتطور ، وبناءً على الذكاء الاصطناعي وفروعه المختلفة ،

في هذه الأطروحة ، طبقنا خوارزميات التعلم الآلي الخاضعة للإشراف ( Random Forest ، Decision Tree ، Naïve Bayes ، Support vector Machine ، K Nearest neighbors ) ومجموعة البيانات المستخرجة من جامعة كاليفورنيا للتنبؤ المبكر بسرطان الثدي لاتخاذ القرارات من حيث التشخيص والتحليل من أجل مكافحة سرطان الثدي قمنا بمقارنة أداء الخوارزميات المختلفة المستخدمة حسب دقة النموذج وحساسيته.  
أظهرت النتائج التجريبية أن أعلى نسبة تم الحصول عليها عن طريق تطبيق Naïve Bayes و Support vector Machine كانت على التوالي 97% و 96%.

الكلمات الرئيسية: الذكاء الاصطناعي ، التعلم الآلي ، التنبؤ بسرطان الثدي ،  
K nearest ، Support vector machine ، Naïve Bayes ، Decision tree ، Random Forest neighbors

### Introduction Générale.

Depuis son existence, l'être humain est exposé à des pathologies diverses, certaines disparaissent avec le temps, pendant que d'autres résistent et prennent de l'ampleur dans l'environnement médico-social et médico-sanitaire.

Parmi ces pathologies, le cancer qui est devenu un problème de santé publique majeur avec une réelle urgence d'intervention et de prise en charge.

Cancer du sein est l'une des principales causes de mortalité chez les femmes dans le monde, entre 1,2 et 2,3 millions de femmes ont reçu un diagnostic de cancer du sein en 2020. Selon l'Organisation mondiale de la santé (OMS), ce taux a déterminé que le cancer du sein est le type le plus courant de cancer chez les femmes dans le monde. Le cinquième taux de mortalité le plus courant (685 000) dans le monde est le cancer du sein. Au cours des 5 dernières années, 7,8 millions de femmes ont été guéries du cancer du sein et ont survécu. Le cancer du sein peut survenir chez les femmes de tout âge. Cependant, il est plus fréquent chez les personnes âgées[1].

Le cancer du sein est l'un des types de cancer les plus dangereux et le plus fréquent chez les femmes. Il affecte généralement les femmes de plus de cinquante ans. Cependant, les femmes de tous âges peuvent développer un cancer du sein et, dans de très rares cas, il peut également affecter Il a coûté la vie à des centaines de milliers de personnes chaque année.

En Algérie, cette maladie représente une source de terreur pour les femmes, car elle représente environ 50% des cancers qui touchent les femmes l'organisation mondiale de la santé (OMS) [2].

Alors que la prédiction précoce du cancer du sein joue un rôle majeur dans le traitement qui en résulte et sauve la vie de milliers de patientes chaque année, les méthodes traditionnelles sont limitées pour fournir cette capacité.

Ces dernières années, le domaine de l'intelligence artificielle et de l'apprentissage automatique a ouvert la porte aux algorithmes sur les données pour le diagnostic médical et les soins de santé, et la prédiction en tirant parti de ces développements pour prédire le risque de cancer du sein en fonction de l'ensemble de données, en particulier l'utilisation des données ainsi que les symptômes accompagnant les patients, le développement de modèles prédictifs

permettant de classer les patients en deux catégories différentes, bénignes ou malignes, pour optimiser le modèle de prédiction.

### **Problématique**

Plusieurs problématiques sont à mettre en évidence dans le cadre de notre travail.

Premièrement : du côté du cancer du sein, la difficulté de diagnostic d'un cancer .

nous tenterons de résoudre cette problématique par L'IA.

Ensuite, il faudra trouver un modèle ML qui répondra aux contraintes imposées par notre thème : Étude comparative des algorithmes d'apprentissage de la machine dans la prédiction et diagnostic du cancer de sein.

### **Organisation du mémoire:**

Notre mémoire est composé 2 parties en globant trois chapitres en plus l'introduction générale.

**Partie 01** : Partie théorique.

**Chapitre 01:** Généralités sur le cancer de sein.

**Chapitre 02:** Apprentissage automatique (Machine learning).

**Partie 02:** Partie pratique.

**Chapitre 03:** Prédiction du cancer de sein par l'apprentissage automatique.

une conclusion et des perspectives futures concluent ce mémoire.

Bibliographie

# **Chapitre01 :**

## **Généralités sur le cancer de sein**

## 1.1.Introduction

Le cancer est l'un des plus mortels de l'histoire, infectant les femmes par un pourcentage important, et la deuxième cause de décès chez les femmes après un cancer de poumon.

Le cancer de sein correspond à une prolifération anarchique des cellules de la glande mammaire et se traduit dans des nombreux cas par un carcinome ou adénocarcinome.

L'origine du terme de cancer revient au médecin grec "Hippocrate", qui est considéré comme le père de la médecine moderne. Il a utilisé le mot grec "carcinose" qui signifie l'organisme marin "cancer" pour décrire les tumeurs cancéreuses.

IL touchent chaque année entre 11000 algériennes(OMS), un problème de santé publique souvent considéré comme un événement dramatique dans la vie d'une femme. Son pronostic est étroitement lié au stade auquel le diagnostic est posé [3].

Dans ce chapitre, nous abordons le concept de cancer de sein, ainsi que l'anatomie féminine de sein et les types de cancer de sein et les symptômes les plus courants et les facteurs de risque et les traitements de cancer de sein.

## 1.2.Définition du cancer de sein

La notion de « cancer du sein » relève d'une nomenclature générique qui fait référence à tout un ensemble de proliférations néoplasiques de la glande mammaire qui diffèrent tant du point de vue histologique qu'en ce qui concerne leur comportement évolutif.

Le terme de « cancer du sein » ne désigne que les tumeurs malignes, potentiellement agressives, du sein tandis que le terme de « tumeur du sein » désigne à la fois les tumeurs malignes et bénignes.

Le terme « carcinome » ou « épithélioma » est parfois utilisé, comme synonyme de « cancer »

En réalité carcinome est un terme spécifique de morphologie microscopique (histologique) qui désigne les plus fréquents des cancers du sein d'origine épithéliale c'est-à-dire dérivant des unités sécrétoires (lobules) ou des canaux lactifères.

Le cancer du sein est une prolifération maligne des cellules épithéliales au niveau canalaire ou lobulaire. Le cancer du sein humain est une maladie clonale : une seule cellule transformée – le résultat final d'une série de mutations somatiques, acquises ou germinales (héritées) – peut exprimer le plein potentiel malin, dans une série d'événements qui se produisent d'une façon séquentielle et aléatoire. Ainsi, le cancer du sein peut exister pendant une longue période comme maladie localisée non invasive ( in situ) ou bien invasive mais pas métastatique ou invasive et métastatique[4].

Il y a deux grandes catégories de tumeurs sont connues :

- Les tumeurs bénignes.
- Les tumeurs malignes ou cancers.

### 1.2.1 Tumeurs bénignes :

Ils ne sont pas cancéreux, c'est-à-dire qu'ils n'envahissent pas les organes voisins et ne les repoussent pas, ils ont généralement une croissance limitée et ne se propagent pas.

On dit qu'elles sont bien différenciées la tumeur bénigne qui se développe le plus souvent dans le sein est le fibroadénome.les autres affections bénignes du sein sont:

- Les kystes.

- Les changements fibrokystiques.
- Les l'hyperplasie.
- Les l'écoulement du mamelon.
- Gynécomastie.

La plupart des masses du sein ne sont pas synonymes de cancer, mais seul l'examen anatomopathologie réalisé après une biopsie permet de vérifier qu'il ne s'agit pas d'un cancer[5].

### 1.2.2.Tumeurs malignes :

Elles envahissent tout le corps dans la zone, infiltrent les organes voisins, toutes ces tumeurs cancéreuses ne sont pas mortelles, tout dépend de leur degré d'extension, de la rapidité du traitement et du type de cellules qui se forment.

Tumeur (cancer) est une masse qui se développe aux dépens d'un organe et à partir des cellules qui le constituent. Lorsqu'une tumeur est soupçonnée d'être maligne, le médecin procédera à une biopsie afin de déterminer la gravité ou l'agressivité de la tumeur[6].

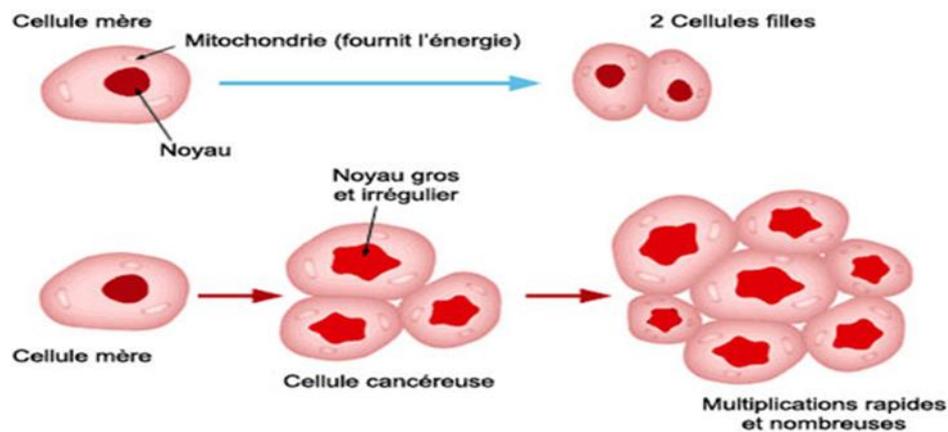
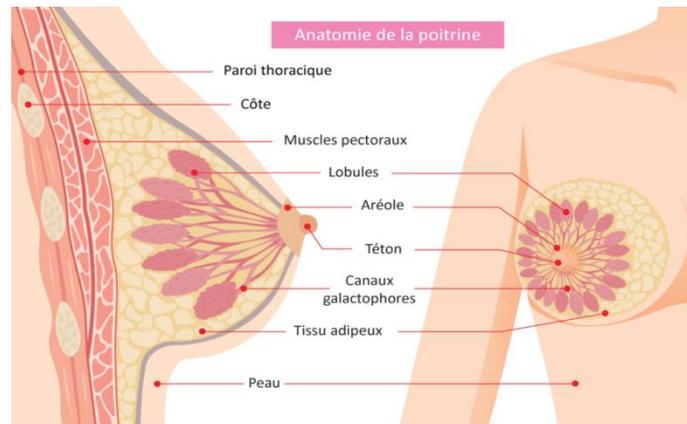


Figure1.1 : Division des cellules cancéreuses[7].

### 1.3.L'anatomie du sein :

Le sein est un organe pair et symétrique de forme hémisphérique, situé en avant du thorax, entre la troisième et la cinquième cote au-dessus du muscle grand pectoral. une

coupe longitudinale passant par le mamelon. la structure du sein est complexe en fait le sein se compose de graisse, de glandes et des canaux servent à transporter le lait jusqu'au mamelon. chaque sein est formé de 15 à 20 compartiments appelés unité lobulaires ou lobes. Chaque lobe est composé de plusieurs lobules qui sont reliés entre eux par des canaux. Les unités lobulaires sont séparées par des ligaments [8].



**Figure1.2 : Anatomie interne du sein[9].**

## 1.4.La classification TNM

L'expression du bilan d'extension d'une tumeur et la plupart du temps, réalisée selon la classification TNM internationalement admis ; ce système est fondé sur l'appréciation de 03 paramètres :

T : Extension locale de la tumeur.

N : Adénopathie régional.

M : Absence ou présence de métastasse à distance.

### 1.4.1.L'extension locale de la tumeur :

T0 : Pas de signe de tumeur primitive.

TI s : Carcinome in situ ; carcinome intra canalaire ou carcinome lobulaire in situ, ou maladie de Paget du mamelon sans tumeur décelable. Une maladie de Paget avec tumeur décelable est classée en fonction de la taille de la tumeur.

T1 : Tumeur <2 cm dans sa grande dimension

T2 : Tumeur >5 cm dans sa plus grande dimension

T4 : Tumeur de toute taille avec extension direct à la paroi thoracique ou à la peau

### 1.4.2. Adénopathie régionale :

N0 : Absence d’envahissement ganglionnaire régional.

N1 : Ganglions axillaire homolatéraux mobiles.

N2 : Ganglions axillaire homolatéraux fixés entre eux ou a d’autre structure.

N3 : Ganglions mammaire internes homolatéraux.

### 1.4.3. Métastase à distance :

M0 : Absence de métastase à distance.

M1 : Présence de métastase à distance (compant des métastases ganglionnaires sus-claviculaire)[10].

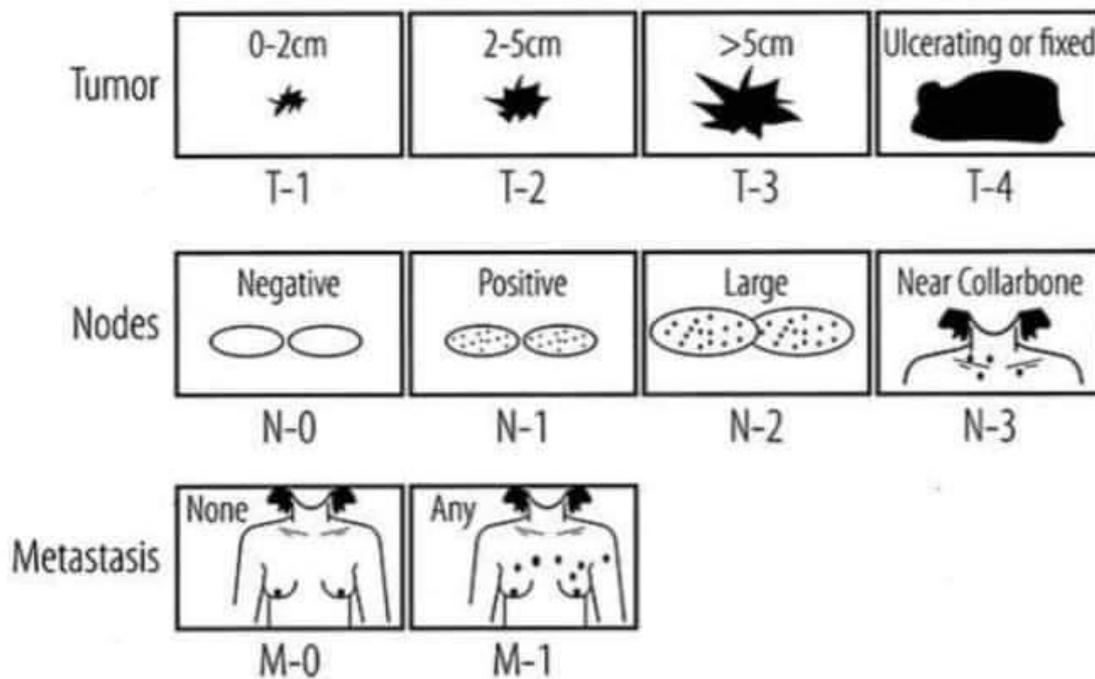


Figure 1.3 : Stades du cancer du sein utilisation le système TNM[11].

## 1.5. Les types de cancer

1-carcinome canalaire in-situ: ce type affecte les conduits qui transportent le lait de la glande mammaire au mamelon.

2-carcinome canalaire infiltrant: ce type atteint les conduits de lait, mais est plus invasif car il propage au tissu mammaire.

3-carcinome lobulaire in-situ: la croissance de cellules anormales dans les glandes mammaires qui sécrètent du lait, mais ce de changement augmente le risque d'infection.

4-le carcinome lobulaire infiltrant: ce type de cancer est plus rare, il débute dans les glandes mammaires puis se propage aux autres tissus du sein.

5-le cancer inflammatoire du sein: il s'agit d'une forme rare de cancer de sein et agressif et semble se développer rapide[12].

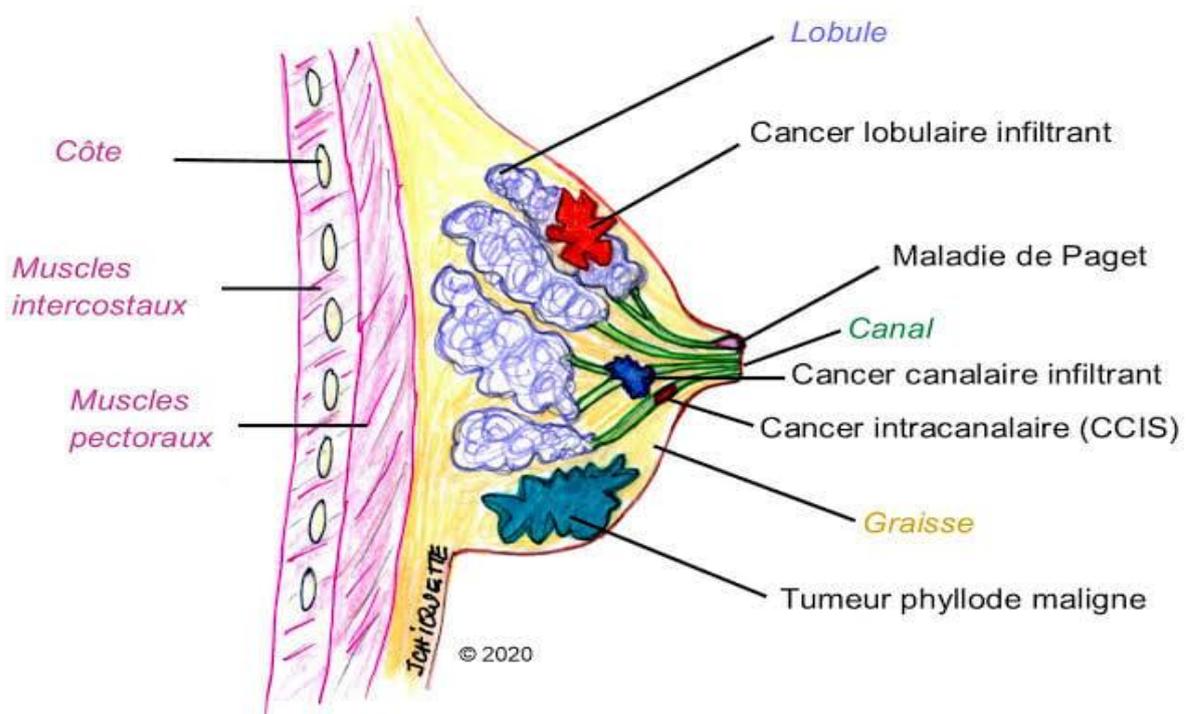


Figure1.4 : Types de cancers du sein[13].

## 1.6. Les causes d'un cancer du sein :

Certaines femmes atteintes d'un cancer du sein cherchent à un moment donné de leur maladie à connaître les causes de leur cancer.

« J'ai toujours voulu savoir pourquoi j'ai développé un cancer du sein »

Les scientifiques cherchent également à connaître les raisons de son apparition. Malgré les progrès qui ont permis de mieux connaître les mécanismes de développement des cancers, les causes du cancer du sein ne sont actuellement pas connues. Néanmoins, les études ont mis en évidence certains facteurs de risque qui favorisent un cancer du sein[14].

## 1.7. Les facteurs de risque et des symptômes :

### 1.7.1. Les facteurs de risques :

- Le sexe (être femme) plus de 99 % des cas de cancer du sein se manifestent chez les femmes. 70% des femmes développant un cancer du sein n'ont d'autre risque que le fait d'être une femme.
- L'âge (vieillir) : le risque augmente avec l'âge ; environ trois quarts des cas de cancer du sein sont décelés chez les femmes âgées de plus de 50 ans.
- Les facteurs liés à la vie reproductive de la femme (système reproducteur) comme la multiparité (jamais avoir eu des enfants), la ménarche précoce (avoir ses premières menstruations avant l'âge de 11-12 ans), une première grossesse tardive (donner naissances à un première enfant après l'âge de 30 ans), la ménopause tardive (achever la période de la ménopause après l'âge de 55 ans).
- Les antécédents familiaux : une augmentation du risque de développer un cancer du sein existe si des proches de la famille au premier degré (la mère et /ou une sœur ) ont contracté un cancer du sein, surtout en période préménopausale (en jeune âge)
- Les antécédents génétique (ou facteur héréditaires) : environ dans 5 à 10 % des cas de cancer du sein sans liées à une mutation génétique. Hériter un des gènes mités liés au cancer du sein augmente considérablement le risque de développer le cancer du sein au cours de sa vie.
- Manque d'activité physique : l'activité physique modérée peut réduire le risque de cancer du sein.



- L'obésité : et la prise de poids corporel pendant l'âge adulte constitue un facteur de risque considérable après la ménopause.
  - La consommation excessive d'alcool : Des résultats indiquant que les femmes qui consomment de l'alcool ont un risque plus élevé de développer un cancer du sein.
  - Le traitement hormonal substitutif (THS) : les traitements à base à la fois d'estrogènes et de progestatifs sont associés à un risque accru de cancer du sein.
- L'exposition aux rayonnements ionisants avant l'âge de 20 ans a été également associée à un risque accru[15].



Figure 1. 5 : Les facteurs de risques du cancer du sein[16].

### 1.7.2. Les symptômes

Les symptômes plus fréquents du cancer du sein sont les suivant :

- Une masse dure et irrégulière ou sensible dans un sein.
- Une masse à l'aisselle.
- Une modification de la peau d'un sein (capitonnage, rougeur, enflure ou démangeaisons).
- Une modification du mamelon (écoulement, formation de croûtes).
- D'autres problèmes de santé peuvent aussi provoquer certaines de ces symptômes ;

- L'apparition de masse dans les seins se produit fréquemment, surtout juste avant les menstruations[17].



Figure1.6 : Les symptômes du cancer du sein[18].

## 1.8.Dépistage :

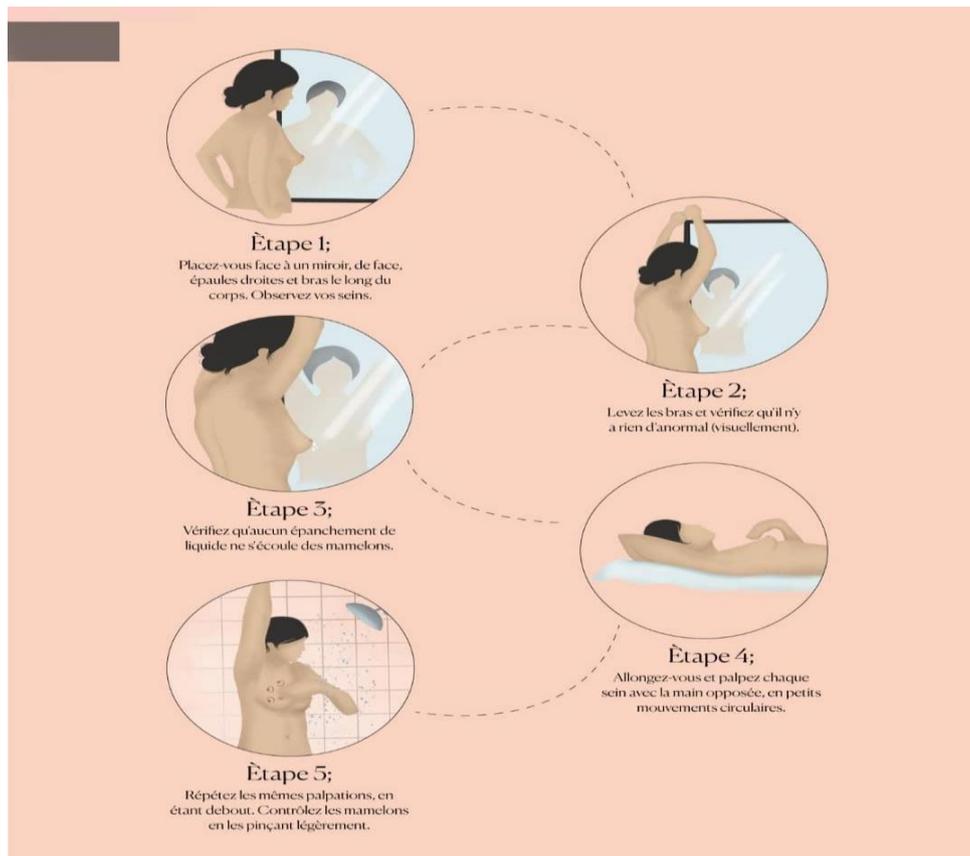
Actuellement, la prévention primaire du cancer du sein n'est pas possible. Compte tenu du taux de mortalité que ce type de cancer engendre, le dépistage reste le seul moyen pratique pour le réduire à terme. Selon l'Organisation Mondiale de la Santé, "le dépistage consiste à détecter à un stade précoce infra clinique (asymptomatique) une maladie pour permettre un diagnostic et une thérapeutique précoces dans le but d'améliorer le pronostic de la maladie et l'état de santé des individus et de la population ".

Le dépistage est une action permettant d'identifier les personnes atteintes d'une maladie ou présentant un risque important d'être atteintes d'une maladie.

Le but: détecter un cancer avant qu'il ne soit palpable ou qu'il ne se traduise par un signe\*anormal comme une modification de la peau ou du mamelon[19].

- **L'autopalpation** : est un geste qu'il est conseillé d'effectuer tous les mois après les règles, afin de repérer une éventuelle grosseur du sein. Mettez-vous d'abord debout devant un miroir, inspectez les deux seins et vérifiez l'absence d'écoulement d'un

mamelon, de crevasses, de plis anormaux ou d'une peau qui pèle. Puis, levez un des deux bras, puis avec les 3 doigts de l'autre main, palpez le sein du côté du bras levé : débutez par la partie externe, les doigts à plat, en effectuant de petits cercles. Il faut rechercher toute sensation de boule, de fossette sur la peau, ou de grosseur. Palpez également le mamelon et la zone entre le sein et l'aisselle, pressez le mamelon et vérifiez qu'aucun écoulement ne se produise[20].



**Figure1.7 :Comment pratiquer L'autopalpation[21].**

➤ **La mammographie:** La mammographie est l'examen radiologique de base. Cet examen est simple, fiable et sera indiqué devant toute anomalie de la glande mammaire. La mammographie est actuellement la seule méthode efficace de dépistage du cancer du sein. Sa sensibilité et sa spécificité sont supérieures à 90 %. Les essais randomisés ont montré que la mortalité par cancer du sein peut être réduite par le dépistage mammographie chez les femmes de plus de 50 ans avec un intervalle entre les vagues de deux à trois ans[22].

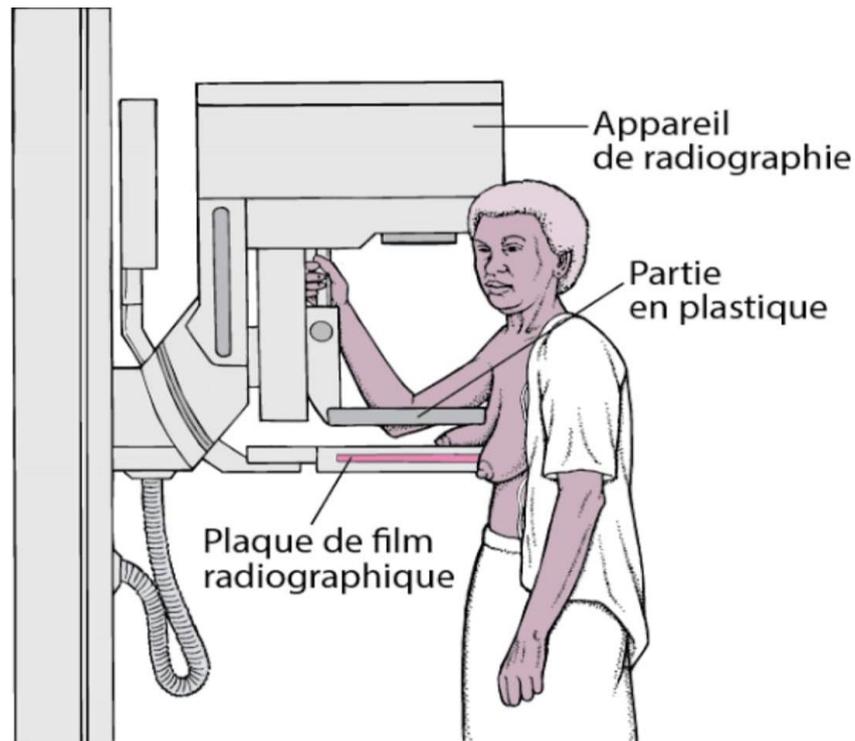


Figure 1.8 : Mammographie : dépistage du cancer du sein[23].

## 1.9.Diagnostic :

Le diagnostic est une étape sensible en général, et en particulier dans le domaine de la médecine, où de nombreuses maladies cancéreuses sont encore à découvrir, cependant, il existe des méthodes de détection précoce où l'on suit les étapes suivantes :

- Bilan sanguin: Lorsque le médecin découvre la présence d'une tumeur ou d'une masse au niveau du sein, ou l'apparition de symptômes et signes inquiétants sur la patiente, le médecin ordonne une prise de sang
- Imagerie médicale : En cas de modification du bilan sanguin, une imagerie médicale est demandée en fonction de la modification découverte parmi les techniques d'imagerie médicale auxquelles nous nous référons. Parmi les techniques d'imagerie on cite la radiographie, échographie, IRM et Pet-scan.
- La radiographie: la radiographie notamment utilisé en oncologie pour examiner le thorax, l'abdomen ou les seins. Un traitement informatique traduit le résultat en image permettant ainsi d'obtenir une représentation indirecte de la région.

Inconvénients : Certains types de tumeurs ne réfléchissent pas ces ultrasons, l'échographie ne permet donc pas de les détecter.

Une échographie est très indirecte et floue, il est souvent ardu pour un médecin de

Distinguer une anomalie.

- L'IRM (imagerie par résonance magnétique) : C'est une méthode d'exploration se fondant sur la détection des réactions de différentes parties du corps exposées à un champ magnétique

Inconvénients : Tous les hôpitaux et centres d'imagerie ne possèdent pas l'équipement spécialisé en IRM du sein.

- Le PET-scan : permet également de visualiser le fonctionnement des organes. Il consiste d'abord à injecter du glucose marqué, c'est-à-dire un sucre faiblement radioactif, puis à analyser l'image obtenue par un scanner. Les cellules cancéreuses sont identifiables car elles ont une activité plus importante que les cellules saines et ont donc besoin de plus de glucose pour fonctionner.

Inconvénients : certaines cellules tumorales ne consomment pas ce glucose, elles ne sont donc pas détectables par pet-scan.

- La biopsie : est un examen incontrôlable lors d'un diagnostic de cancer. Elle seule permet de confirmer le diagnostic et de préciser la nature de la lésion cancéreuse.

Après avoir constaté une tumeur ou quelque anomalie conséquente en imagerie, un prélèvement du tissu suspecté est effectué pour le faire analyser par un anatomopathologiste.

Inconvénients : Tout comme les autres interventions chirurgicales demandant une forme d'anesthésie, la biopsie mammaire présente quelques risques même s'ils sont moindres. En effet, ils dépendent de différents paramètres tels que la méthode d'intervention ou l'état général du patient. A part les réactions allergiques à l'anesthésie qui surviennent rarement, les risques sont limités au saignement, à l'infection et à l'ecchymose[24].

## 1.10. Les traitements médicaux du cancer du sein :

Différents types de traitements peuvent être utilisés pour traiter un cancer du sein : la chirurgie, la radiothérapie, l'hormonothérapie, la chimiothérapie et les thérapies ciblées.

Selon les cas, les traitements peuvent avoir différents objectifs :

- Supprimer la tumeur ou les métastases.
- Réduire le risque de récurrence.
- Ralentir le développement de la tumeur ou des métastases.



- Améliorer le confort et la qualité de vie de la personne malade, en traitant symptômes engendrés par la maladie.
- La chirurgie :

La chirurgie est le meilleur moyen et le traitement de base en cas de cancer du sein, du moins dans ses phases précoces. Elle est fréquemment accompagnée d'autres traitements pour une efficacité accrue et pour minimiser le risque de récurrences. Ces traitements complémentaires sont choisis en fonction de plusieurs éléments, notamment selon le stade d'évolution du cancer, de ses caractéristiques et du type de chirurgie, à savoir la tumorectomie ou la mastectomie. La tumorectomie, appelée aussi chirurgie conservatrice, consiste à enlever la tumeur tout en préservant une partie du sein. Cette pratique est souvent complétée par une chimiothérapie ou une radiothérapie. Quant à la mastectomie, une ablation complète du sein est réalisée, avec une possible reconstruction mammaire ultérieure selon les cas cliniques. Une ablation de plusieurs ganglions axillaires peut aussi être nécessaire, si ces derniers contiennent des cellules cancéreuses [25].

- La chimiothérapie :

La chimiothérapie consiste en un traitement systémique, c'est à dire qu'en pénétrant dans la circulation sanguine, elle atteint toutes les cellules de l'organisme y compris les cellules saines. La chimiothérapie provoque alors l'apparition de nombreux effets secondaires, s'accompagnant de lourdes conséquences chez la femme, qui seront traitées plus loin dans ce travail. Selon la gravité du cancer et ses caractéristiques, la chimiothérapie a plusieurs buts. Elle empêche la prolifération des cellules cancéreuses et vise la diminution du risque de récurrence, en étant administrée après ou entre deux traitements (chirurgie ou radiothérapie) ou encore en renforçant l'effet d'un autre traitement anticancéreux. Lorsqu'elle est administrée après une chirurgie, on parle de chimiothérapie adjuvante [26].

Avant une chirurgie et pour diminuer la taille de la tumeur, on emploie le terme de chimiothérapie néo-adjuvante.

- La radiothérapie :

Afin de compléter le traitement chirurgical, la radiothérapie permet de traiter localement la tumeur et ses environs, ainsi que de potentiels ganglions atteints. Cette intervention sert à prévenir d'éventuelles récurrences avec une fréquence de traitement individualisée selon l'état clinique de la patiente. En général, la radiothérapie se déroule quotidiennement, et varie d'une



à sept semaines selon les cas. Les avantages de la radiothérapie sont la rapidité du traitement et le fait qu'elle ne cause aucune douleur durant la phase aiguë du traitement. Cependant, certains effets secondaires peuvent survenir dans certaines situations mais n'entraînent généralement que peu de complications sérieuses. Ils disparaissent naturellement à la fin du traitement[27].

➤ Hormonothérapie :

On administre souvent une hormonothérapie pour traiter le cancer du sein dont les récepteurs hormonaux sont positifs. Les femmes ménopausées reçoivent des médicaments hormonaux différents de ceux qu'on administre aux femmes pré-ménopausées [28].

➤ Les thérapies ciblées :

Les thérapies ciblées forment une classe innovante de médicaments anticancéreux : à l'inverse de la chimiothérapie, ils agissent spécifiquement sur les cellules cancéreuses en ciblant une

Caractéristique propre à celles-ci, Ce type de traitement permet en principe d'obtenir une efficacité, tout en réduisant le risque d'effets secondaires[29].

### **1.11.Conclusion :**

Le cancer de sein est une maladie très dangereuse pour le corps humain, et il résulte de la destruction de certaines cellules qui se développent souvent et forment une masse appelée "tumeur". Comme la détection précoce de cette maladie augmente le taux de guérison. Récemment, de nombreuses études dans toutes disciplines ont contribué à fournir des méthodes. Prédire cette maladie avant l'infection repose sur l'intelligence artificielle et ses différentes branches.



## **Chapitre 2**

# **Apprentissage Automatique (Machine Learning)**

## 2.1.Introduction

Le développement dans le domaine de la technologie s'est amélioré au fil des ans. Aujourd'hui, nous entendons parler des termes de technologie comme l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.

L'intelligence artificielle est un grand domaine dans lequel les ordinateurs sont formés

Pour montrer un comportement intelligent qui utilise une variété de statistiques et de techniques d'optimisation permettant aux ordinateurs d'apprendre à partir d'exemples précédents et de découvrir des modèles difficiles à distinguer d'un ensemble de données volumineux au complexe. Cette capacité est particulièrement adaptée aux applications médicales, applications où l'apprentissage automatique est utilisé dans le traitement de cancer.

Plus récemment, l'application du diagnostic et de la détection du cancer du sein, où l'apprentissage automatique a été appliqué au diagnostic et à la prédiction, est particulièrement intéressante car elle s'inscrit dans une tendance croissante vers la médecine prédictive.

Les méthodes d'apprentissage peuvent être utilisées pour améliorer la précision de la prédiction du cancer, significativement plus de récurrences (15 % à 25 %) et de mortalité à un niveau plus fondamental, et également aider à améliorer notre compréhension de base du développement du cancer.

Où, dans ce chapitre, nous discuterons des différents types d'algorithmes d'apprentissage automatique et de leurs branches pour prédire et diagnostiquer le cancer du sein.

## 2.2. Qu'est-ce que l'apprentissage automatique?

**L'apprentissage automatique:** l'apprentissage automatique est une technique qui fait partie de l'intelligence artificielle pour former des modèles complexes.

Et se définit comme un domaine d'étude qui consiste à donner la capacité à des algorithmes d'apprendre à partir de données sans être explicitement programmé. Il a été inventé par l'informaticien Américain « Arthur Samuel »

## 2.3. Définition de l'apprentissage automatique (machine Learning):

L'apprentissage automatique (ou artificiel) est l'un des domaines d'étude de l'intelligence artificielle et une science informatique qui fait référence à la capacité d'un système à acquérir et à intégrer des connaissances de manière autonome. L'apprentissage automatique développe et analyse des méthodes qui permettent à une machine d'apprendre et d'effectuer des tâches. Associé à l'intelligence artificielle à travers le processus d'apprentissage, ce dernier permet d'obtenir un système optimisé en fonction de l'environnement. Il met l'accent sur l'utilisation des données (données d'entraînement) et des algorithmes, et parmi ses objectifs est de rendre les systèmes capables d'apprendre et de se développer automatiquement par l'expérience sans effectuer d'opérations logicielles[30].

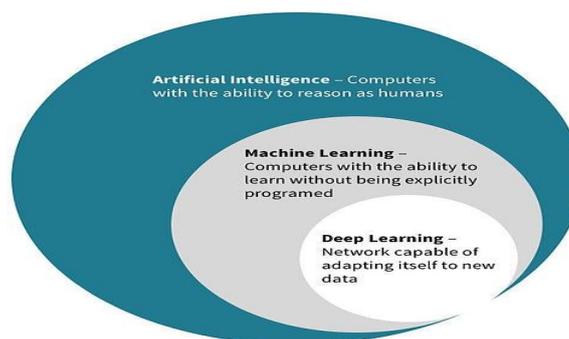


Figure 2.1 : Intelligence artificielle vs. Machine Learning[31].

## 2.4. Techniques d'apprentissage automatique :

- Nombre de points de données: peut utiliser de petites quantités de données pour faire des prédictions.
- Dépendances matérielles: peut fonctionner sur des machines bas de gamme.

Ne nécessite pas beaucoup de puissance de calcul.

- Processus de personnalisation: Nécessite que les caractéristiques soient identifiées et créés avec précision par les utilisateurs.
- Approche d'apprentissage: scinde le processus d'apprentissage en étapes plus petites. Combine ensuite les résultats de chaque étape dans une seule sortie.
- Temps d'exécution: Nécessite relativement peu de temps pour apprendre ,de quelques secondes à quelques heures.

Sortie: la sortie est généralement une valeur numérique telle qu'une note ou une classification[32].

### 2.5.Types d'apprentissage automatique:

L'apprentissage automatique est une méthode employée dans l'intelligence artificielle. C'est une méthode d'apprentissage statistique où chaque instance dans une base de données est décrite par un ensemble de caractéristiques ou d'attributs.

On peut distinguer 2grandes catégories:

- Apprentissage supervisé.
- Apprentissage non supervisé[33].

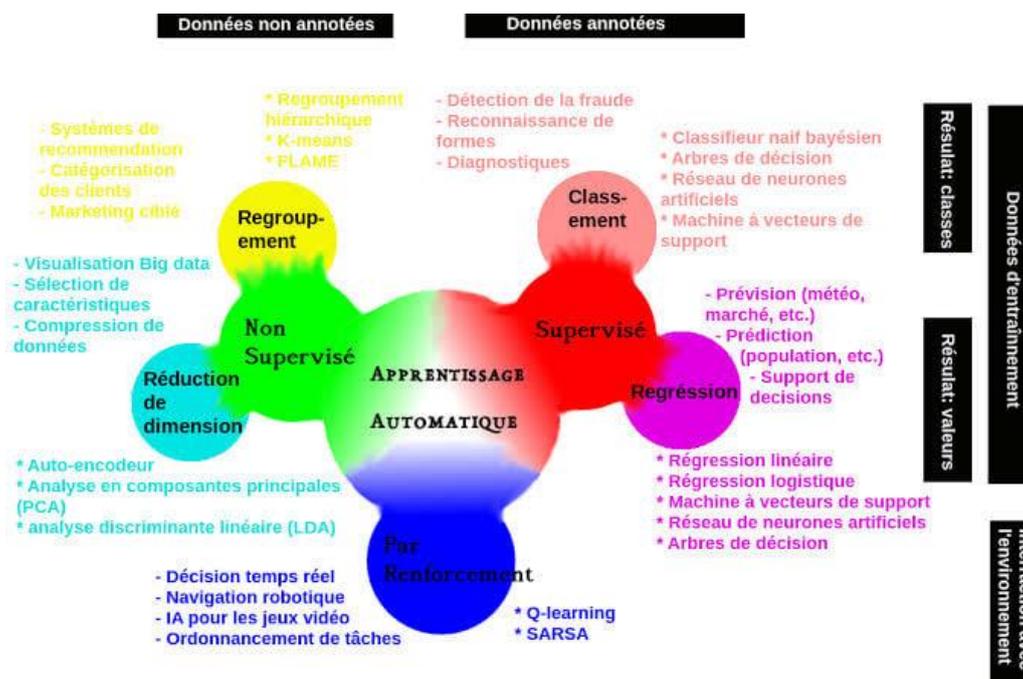


Figure 2.2 : Types d'apprentissage automatique(Machine Learning)[34].

### 2.5.1.Apprentissage supervisé :

C'est une tâche pour l'apprentissage automatique d'apprendre la fonction de prédiction à partir d'exemples annotés, plutôt qu'un apprentissage non supervisé. Un humain aide l'algorithme à apprendre, servant le monde comme guide, et enseigne à l'algorithme les résultats qu'il doit trouver. L'algorithme apprend grâce Les méthodes d'apprentissage supervisé sont bien généralisables, c'est-à-dire l'apprentissage d'une fonction qui fait des prédictions correctes sur des données non présentes

dans l'ensemble d'apprentissage[35].

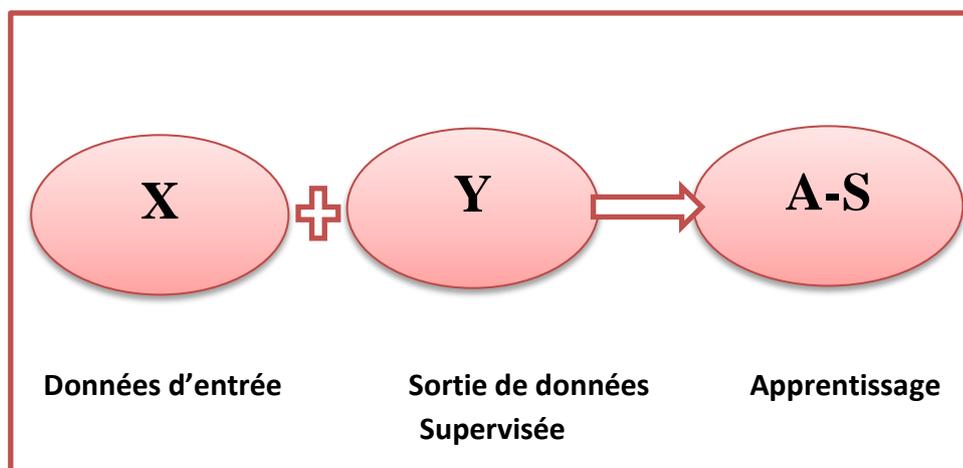


Figure2.3 :Apprentissage supervisé.

Il y a trois types d'apprentissage supervisé :

- **La classification** : Dans les tâches de classification, le programme d'apprentissage automatique doit tirer une conclusion à partir des valeurs observées et déterminer à quelle catégorie appartient les nouvelles observations. Par exemple, lors du filtrage des e-mails comme "spam" ou "non spam", le programme doit examiner les données d'observation existantes et filtrer les e-mails en conséquence.
- **Régression** : dans les tâches de régression, le programme d'apprentissage automatique doit estimer - et comprendre- les relations entre les variables. L'analyse de régression se concentre sur une variable dépendante et une série d'autres variables changeantes, ce qui la rend particulièrement utile pour la prédiction et la prévision.
- **Prévision** : La prévision est le processus consistant à faire des prédictions sur l'avenir sur la base des données passées et présentes, et est couramment utilisée pour analyser les tendances[36].

### 2.5.2. Apprentissage Non Supervisé :

Ici, l'algorithme d'apprentissage automatique étudie les données pour identifier les modèles. Il n'y a pas de clé de réponse ou d'opérateur humain pour fournir des instructions. Au lieu de cela, la machine détermine les corrélations et les relations en analysant les données disponibles. Dans un processus d'apprentissage non supervisé, l'algorithme d'apprentissage automatique doit interpréter de grands ensembles de données et traiter ces données en conséquence. L'algorithme essaie d'organiser ces données d'une manière ou d'une autre pour décrire leur structure. Cela peut signifier regrouper les données en grappes ou les organiser d'une manière qui semble plus organisée.

Au fur et à mesure qu'il évalue davantage de données, sa capacité à prendre des décisions sur ces données s'améliore progressivement et s'affine.

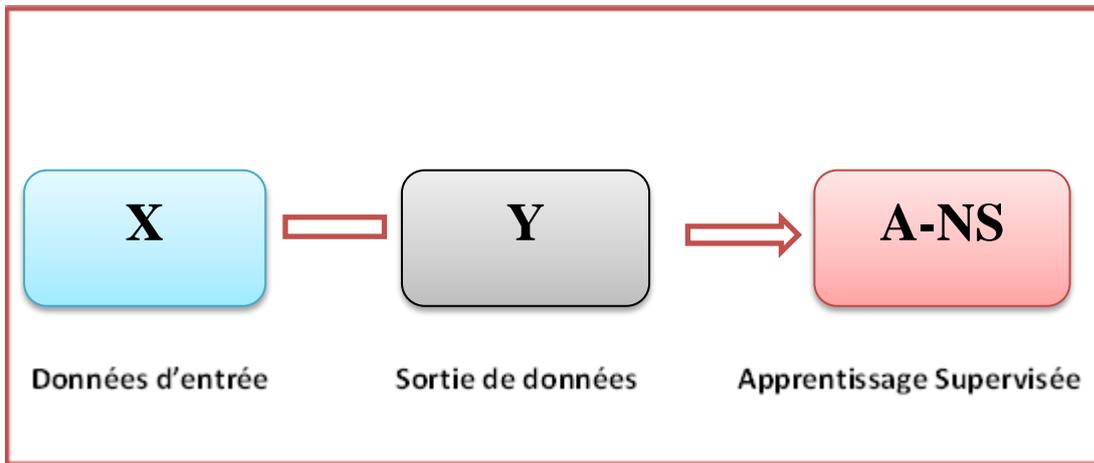


Figure2.4 : Apprentissage non supervisé.

Il y a deux types d'apprentissage non supervisé :

- **Regroupement ( Clustering ) :** Le clustering consiste à regrouper des ensembles de données similaires (sur la base de critères définis). Il est utile pour segmenter les données en plusieurs groupes et effectuer une analyse sur chaque ensemble de données pour trouver des modèles.
- **Réduction de la dimension :** La réduction de dimension réduit le nombre de variables considérées pour trouver l'information exacte requise[37].

*Il y a une Différences entre les types d'apprentissage automatique en résumé dans ce tableau :*

Apprentissage supervisé	Apprentissage non supervisé
Données d'entrée sont étiquetées.	Données d'entrée son non étiquetées.
Utilise le jeu de données d'apprentissage.	Utilise tout le jeu de données en entrée.
Utilisé pour la prédiction.	Utilisé pour l'analyse.
Classification et régression.	Regroupement, estimation de la densité, et réduction de la dimensionnalité.

Tableau 2.1 : Différences entre l'apprentissage supervisé et non supervisé[38].

## 2.6. Les algorithmes de l'apprentissage automatique utilisés :

### 2.6.1. SVM (Machines à vecteurs de support) :

Il s'agit d'un modèle d'apprentissage automatique, puissant et polyvalent, et c'est l'un des modèles d'apprentissage automatique les plus courants.

Il s'agit d'un classificateur qui divise les ensembles de données en catégories pour trouver le niveau hyper marginal maximal sur les points de données les plus proches et minimiser les erreurs de classification.

SVM est une technique de classification ML supervisée couramment utilisée dans le domaine du diagnostic et du pronostic du cancer. SVM sépare les catégories en sélectionnant des échantillons clés de toutes les catégories appelés vecteurs de support et en générant une fonction linéaire qui utilise ces vecteurs de support pour les diviser aussi largement que possible[39].

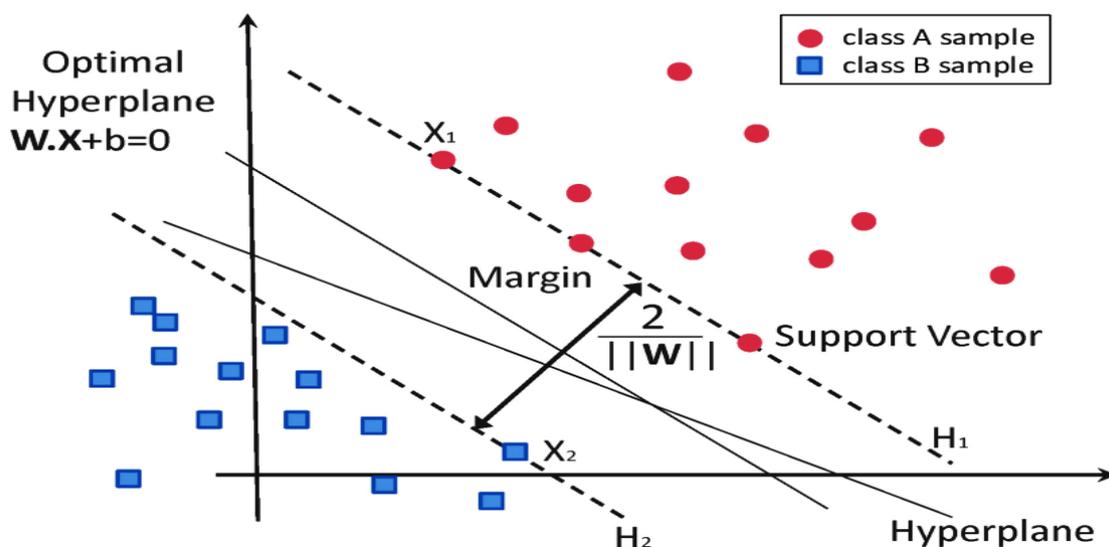


Figure 2.5 : principe de classification en machines à vecteurs de support (SVM)[40].

Exemple

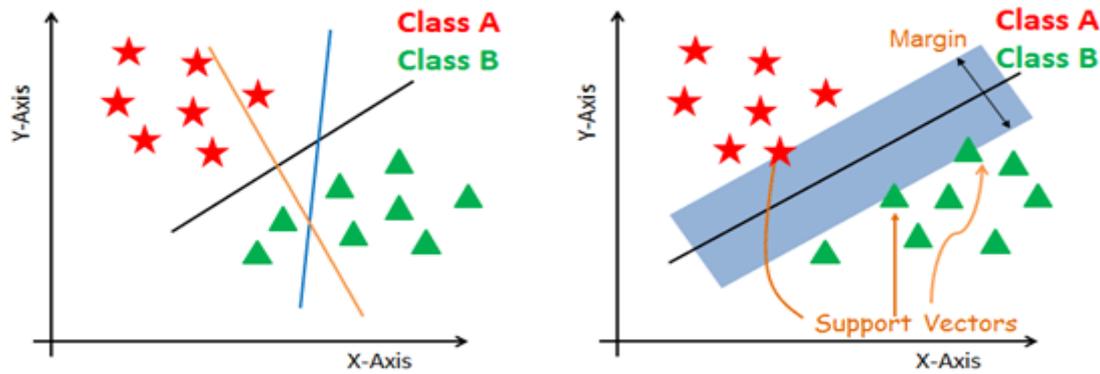


Figure 2.6 : Un simple exemple sur le fonctionnement de l'algorithme SVM[41].

### L'interprétation d'exemple

Dans cet exemple le jeu de données contient des étoiles et des triangles qui sont respectivement classés dans la classe A et B, dans la phase d'apprentissage le classificateur SVM consiste à trouver le meilleur hyperplan qui sépare parfaitement les deux classes, et classe correctement les nouvelles données ainsi comme les vecteurs de support crée une frontière de décision entre les deux classes les nouvelles données sera classées à la base de ces vecteurs.

## Hyperplan et vecteur de support et marge dans l'algorithme SVM

**Hyperplan** : il peut y avoir plusieurs lignes/limites de décision pour séparer les classes dans un espace à  $n$  dimensions, mais nous devons trouver la meilleure limite de décision qui aide à classer les points de données. Cette meilleure frontière est connue sous le nom d'hyperplan de SVM.

Les dimensions de l'hyperplan dépendent des entités présentes dans le jeu de données, ce qui signifie que s'il y a 2 entités (comme indiqué sur l'image), alors l'hyperplan sera une ligne droite. Et s'il y a 3 caractéristiques, alors l'hyperplan sera un plan à 2 dimensions.

Nous créons toujours un hyperplan qui a une marge maximale, c'est-à-dire la distance maximale entre les points de données.

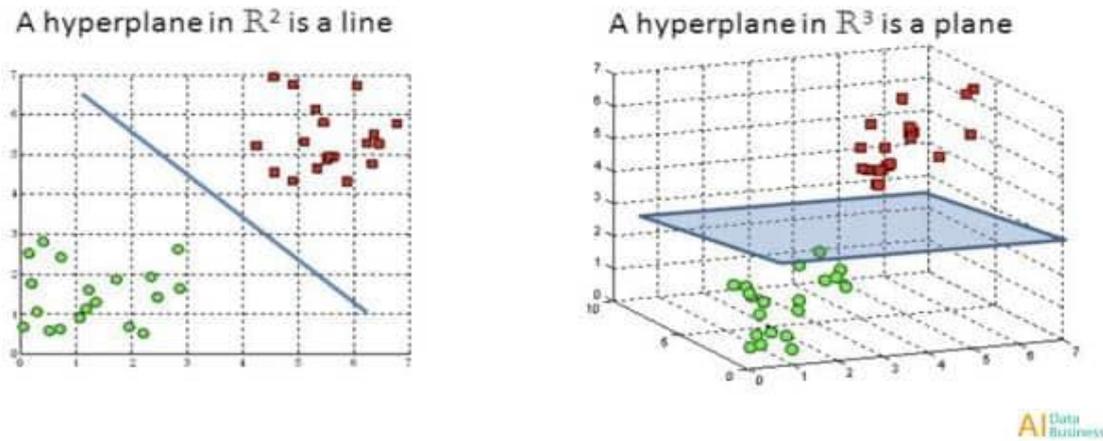


Figure 2.7 :Hyperplan dans les entités 2D et 3D[42].

**Vecteur de support :** Les machines à vecteurs de support, ou support vector machine (SVM), sont des modèles de machine learning supervisés centrés sur la résolution de problèmes de discrimination et de régression mathématiques. Elles ont été conceptualisées dans les années 1990 à partir d'une théorie d'apprentissage statistique développée par les informaticiens russes Vladimir Vapnik et Alexey Chervonenkis : la théorie de Vapnik-Chervonenkis. Ce modèle a été rapidement adopté en raison de sa capacité à travailler avec des données de grandes dimensions, ses garanties théoriques et les bons résultats réalisés en pratique. Requérant un faible nombre de paramètres, les SVM sont appréciées pour leur simplicité d'usage.

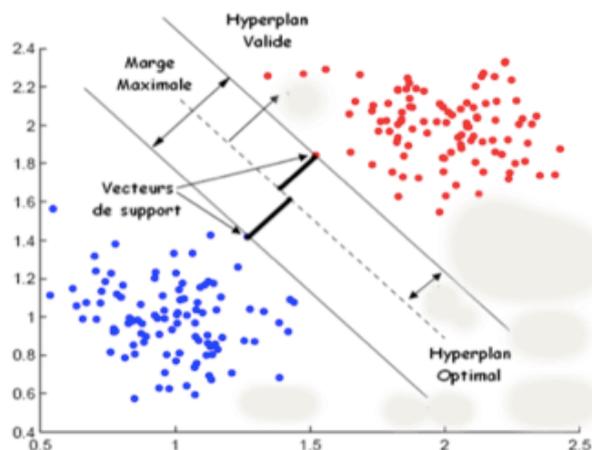


Figure 2.8 : Les vecteurs de support[43].

**la marge :** est la distance entre la frontière de séparation et les échantillons les plus proches. sont appelés vecteurs supports. Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge[44].

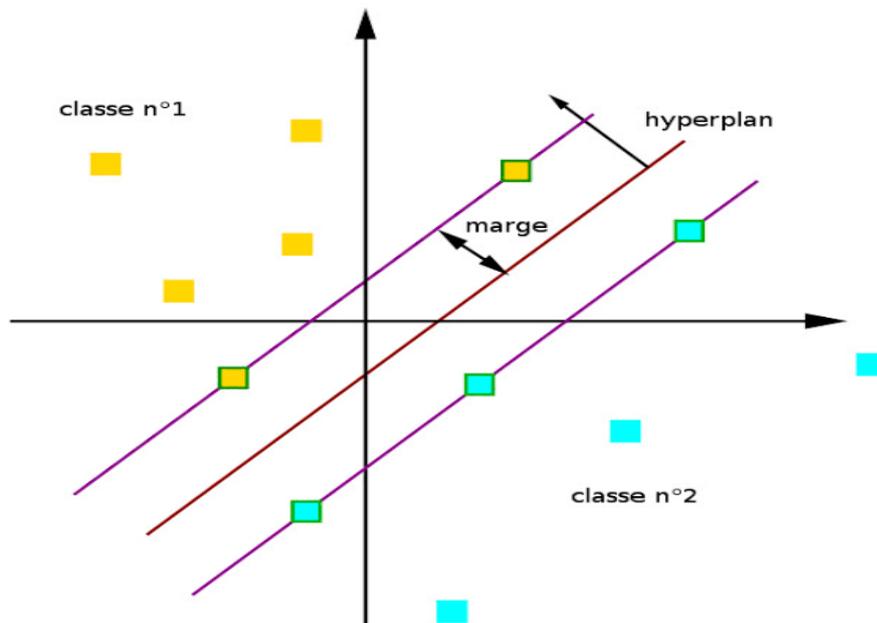


Figure2.9 : Marge dans l'algorithme SVM[45].

#### Avantage de SVM :

1. Leur capacité à manipuler de grandes quantités de données.
2. Le faible nombre d'hyper paramètres utilisés par ces méthodes.
3. Elles sont bien fondées théoriquement.
4. Les résultats pertinents qu'on obtient avec les SVM en pratique.

#### Inconvénients des SVM :

1. Leur utilisation des fonctions mathématiques complexes pour la classification des corpus.
2. Pour trouver les meilleurs paramètres, ce type d'algorithmes demande un temps énorme pendant les phases de test.

**2.6.2. La Méthode de k plus proches voisins:** Le premier algorithme que nous étudierons est l'algorithme du k plus proche voisin, qui est le plus souvent utilisé pour la classification, bien qu'il puisse également être utilisé pour l'estimation et Prédiction. k-Plus proche voisin est un exemple d'apprentissage basé sur les instances, dans lequel L'ensemble de données d'apprentissage est stocké, de sorte qu'une classification pour un nouvel enregistrement non classifié, peut-être trouvé simplement en le comparant aux enregistrements les plus similaires dans l'ensemble d'apprentissage[46].

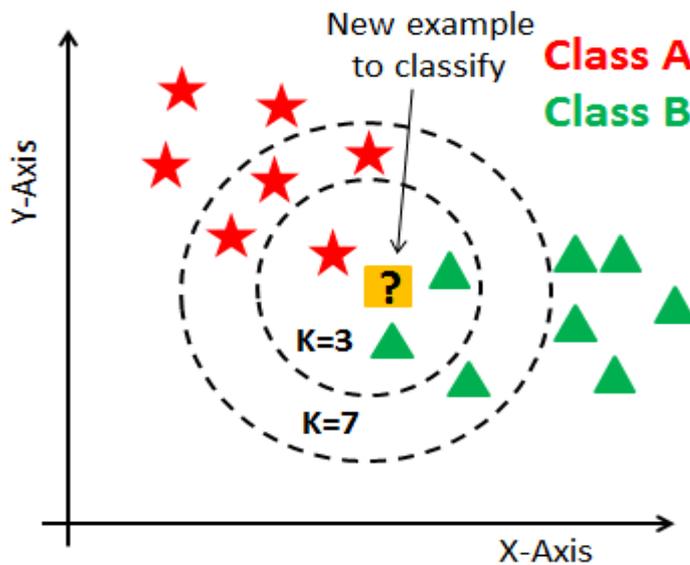


Figure2.10 : La Méthode de k plus proches voisins[47].

### L'interprétation de l'exemple

Dans cet exemple nous avons une donnée non classée et toutes les autres données sont classées (étoile et triangle) chacun avec leur classe (classe A et B).

Si  $k=3$  les données les plus proches de la nouvelle donnée sont qui ont à l'intérieur de premier cercle, et la classe la plus prééminente c'est triangle (Classe B) car 2 triangles et seulement 1 étoile donc la donnée non classée sera classée un triangle (Classe B).

Si  $k=7$  les données les plus proches de la nouvelle donnée sont qui ont à l'intérieur de deuxième cercle, et la classe la plus prédominante c'est l'étoile (Classe A) car on a 4 étoiles et 3 triangles donc la donnée non classée sera classée une étoile (Classe A).

### 2.6.2. Avantages de la méthode des k plus proches voisins (KNN) :

1. La facilité de mise en œuvre de cet algorithme.
2. Son efficacité pour des classes réparties de manière irrégulière.
3. Son efficacité pour des données incomplètes.
4. La méthode des k plus proches voisins n'utilise pas de modèle pour classer les documents.

### Inconvénients de la méthode des k plus proches voisins (KNN) :

1. le choix de la valeur de k (le nombre de voisins le plus proche)
2. Le coût de calcul est élevé (pour chaque instance de l'ensemble de données on a besoin de calculer la distance)
3. Stockage de données

4. Sensible aux fonctionnalités non pertinentes

**2.6.3. Les forêts aléatoires ou les forêts de décision aléatoires :**

sont une méthode d'ensemble pour la classification, la régression et d'autres tâches, qui fonctionnent en construisant une multitude d'arbres de décision au moment de la formation et en produisant la classe qui est le mode des classes (classification) ou la prédiction moyenne (régression) des arbres individuels. Aléatoire les forêts de décision corrigent l'habitude des arbres de décision de trop s'adapter à leur ensemble d'apprentissage[48].

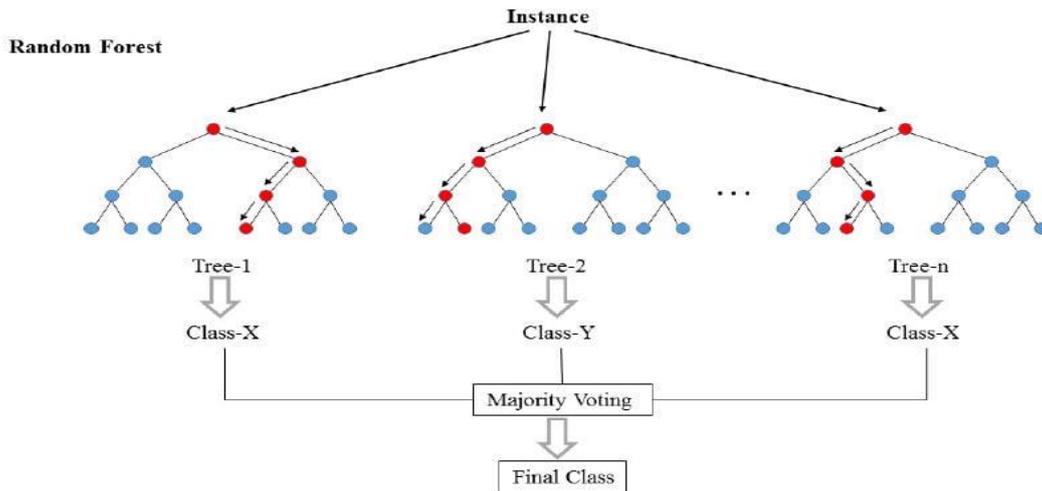


Figure2.11 : Forêts aléatoires dans l'apprentissage automatique[49].

Exemple :



Figure2.12 : Un simple exemple sur l'algorithme random forest[50].

L'interprétation d'exemple :

Comme son nom l'indique, un arbre de décision aide le Data Scientist à prendre une décision grâce à une série de questions (aussi appelées tests) dont la réponse (oui/non) mènera à la décision finale.

Prenons un exemple de classification binaire : on cherche à savoir si un champignon est comestible en fonction des critères -ou features en anglais- suivants : couleur, taille du champignon, forme du chapeau, odeur, taille de la tige, présence de tâches, etc.

### 2.6.4.L'arbre de décision

est un outil de modélisation prédictive qui peut être appliqué dans de nombreux domaines. Il peut être construit par une approche algorithmique qui peut diviser l'ensemble de données de différentes manières en fonction de différentes conditions[51].

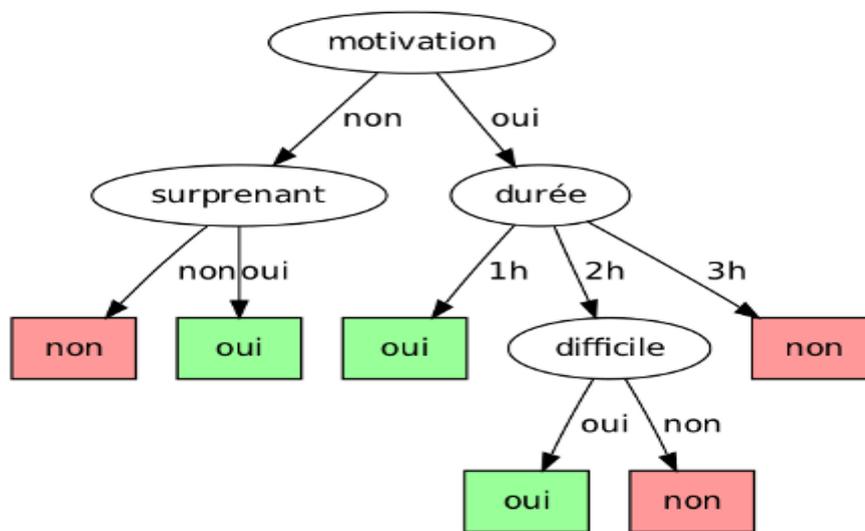


Figure2.13 : Arbre de décision[52].

#### Avantages des arbres de décision

1. Leur capacité à travailler sur des données symboliques.
2. Leur grande capacité et efficacité à faire de la classification.
3. Leur facilité d'apprentissage et d'utilisation.

#### Inconvénients des arbres de décision

1. Ce type d'algorithmes est très sensible aux points aberrants et au bruit.
2. Leur sensibilité au changement des données.
3. Une détection difficile des interactions entre les variables a retiré un message

### 2.6.5.Naïve Bayes (Apprentissage supervisé-Classification):

La méthode de classification naïve de Bayes est un algorithme d'apprentissage automatique supervisé qui classe un ensemble d'observations selon les règles déterminées par l'algorithme

lui-même. L'outil de classification doit d'abord être entraîné sur un ensemble de données d'entraînement, qui affiche la catégorie attendue en fonction de l'entrée. Dans la phase d'apprentissage, l'algorithme développe ses règles de classification sur cet ensemble de données et les applique à la classification de l'ensemble de données prédit dans la deuxième étape. Le classificateur NB signifie que la classe des données d'apprentissage est connue et fournie. La classification NB a obtenu des résultats remarquables dans de nombreuses applications quotidiennes, ce qui en fait l'algorithme préféré des outils d'apprentissage automatique. La base de la classification NB est le théorème de Bayes, qui simplifie l'hypothèse d'indépendance entre toutes les paires de variables, qui est appelée naïve. Le théorème de Bayes énonce la relation suivante, étant donné la variable de classe  $y$  et le vecteur de fonction dépendante  $x_1$  à  $x_n$  (Godé, 2020)  $P(A/B) = P(A)P(B/A)P(B)$

$P(A/B)$  : Est la probabilité postérieure de classe ( $y$ , cible) donnée prédicteur ( $x$ , attributs).

$(A)$  : est la probabilité a priori de classe.

$(B/A)$  : est la vraisemblance qui est la probabilité du prédicteur de la classe donnée.

$(B)$  : est la probabilité a priori du prédicteur[53].

### Avantages des méthodes Naïve Bayes

1. La facilité et la simplicité de leur implémentation.
2. Leur rapidité.
3. Les méthodes Naïve Bayes donnent de bons résultats.

### Inconvénient de Naïve Bayes

limitation de Naïve Bayésienne est l'hypothèse de fonctionnalités indépendantes.

Dans la vraie vie, il est presque impossible d'obtenir un ensemble de fonctionnalités complètement indépendants.

### Remarque

Un bon modèle d'apprentissage choisit la fonction de prédiction qui réalise la plus faible erreur de prédiction.

## 2.7. Application de l'intelligence artificielle à la prédiction et au diagnostic en cancérologie :

L'intelligence artificielle est devenue un outil précieux dans la lutte contre le cancer. Cette technologie est mise au service de l'imagerie mammaire et de la santé de la femme. En effet, les outils d'imagerie mammaire jouent un rôle important dans le diagnostic et dans le dépistage du cancer du sein. Néanmoins l'interprétation directe d'un très grand nombre des

clichés est peut-être fastidieuse. Pendant les débuts des années 1980, des systèmes de diagnostic assisté par ordinateur (CAD) ont été développés. Dans le début des années 1970 aux années 1990, l'analyse des images est

Réalisée à l'aide d'un traitement de bas niveaux ; le débruitage, le rehaussement du contraste, la détection de bords et de lignes, la segmentation des régions en se basant sur les différents outils mathématiques. A l'heure actuelle de nouveaux systèmes d'intelligences artificielles inspirées du cerveau humain, sont à la disposition des médecins. L'intelligence artificielle a fait preuve d'une meilleure précision, pour identifier un cancer du sein à un stade précoce et, réduire le taux de faux positifs et de faux négatifs.

Les outils d'intelligence artificielle peuvent analyser un nombre important d'images médicales, en un temps restreint. Avec l'adoption des techniques adéquates, l'intelligence artificielle est aussi capable de prédire un cancer du sein, avant qu'il ne soit présent[54].

## 2.8.Intelligence artificielle VS apprentissage automatique :

Paramètre	Intelligence artificielle	Apprentissage automatique
<b>Concept</b>	C'est un concept plus large pour créer des machines intelligentes pour simuler la pensée et le comportement humains.	Il s'agit d'un sous-ensemble de l'intelligence artificielle pour aider les machines à apprendre en analysant les données sans programmation explicite.
<b>Objectif</b>	Il vise à créer des systèmes plus intelligents dotés de capacités de réflexion humaines pour résoudre des questions complexes. Il est préoccupé par l'augmentation des taux de réussite.	Il vise à permettre aux machines d'analyser les données afin de fournir une sortie précise. Il est préoccupé par les modèles et la précision
<b>Ce qu'ils font</b>	L'IA permet à un système d'être capable d'effectuer des tâches comme le ferait un	Les machines apprennent en permanence à améliorer et à exécuter une tâche afin

	humain, mais sans erreurs et à une vitesse plus rapide.	qu'elle puisse fournir plus de précision.
<b>Sous-ensembles</b>	Ses sous-ensembles sont l'apprentissage en profondeur et l'apprentissage automatique.	Son sous-ensemble est l'apprentissage en profondeur
<b>Types de séchoir</b>	Il est de trois types - IA générale, IA forte et IA faible	Ses types sont l'apprentissage par renforcement, supervisé et non supervisé
<b>Procédé</b>	Cela comprend le raisonnement, l'apprentissage et l'autocorrection	Il comprend l'apprentissage ainsi que l'autocorrection pour les nouvelles données
<b>Types de données</b>	Il traite des données non structurées, semi-structurées et structurées	Il traite des données semi-structurées et structurées
<b>Portée</b>	Son champ d'action est plus large. Les systèmes d'IA peuvent effectuer plusieurs tâches au lieu du ML qui est formé pour des tâches spécifiques.	Sa portée est limitée par rapport à l'IA. Les machines ML effectuent des tâches spécifiques pour lesquelles elles sont formées
<b>Usages</b>	Ses applications sont les chatbots, les robots, les systèmes de recommandation, les jeux, les médias sociaux et bien d'autres.	Les principales applications sont les recommandations en ligne, les suggestions d'amis Facebook, la recherche Google, etc.

[55]

## 2.9.Conclusion :

L'apprentissage automatique est un paradigme important et largement utilisé pour nombreux problèmes . Dans ce chapitre ,nous avons présenté l'apprentissage automatique (ML)et leur technique. Nous avons d'abord commencé par la définition de l'apprentissage automatique et ses différents types ,ensuite détaillé les algorithmes utilisés pour étude la prédiction et diagnostic de cancer de sein ,qui peuvent nous aider la détection de l'apparition précoce de cancer de sein et réduire les risques de cette maladie sur la santé se patient.

Le prochain chapitre, sera consacré d'étude comparative pour application des différentes algorithmes, Decision Trees, Random Forest, Support Vector Machine, Naïves Bayes, K-nearst neighbors ; Dans la prédiction et diagnostic de cancer de sein.

## **Chapitre 3 :**

**Prédiction de cancer de sein  
par l'apprentissage  
automatique.**

### 3.1.Introduction :

Tout au long des chapitres précédents ,nous avons détaillé les algorithmes de l'apprentissage automatique dans la prédiction et diagnostic du cancer de sein.

Dans ce chapitre ,nous présentons la définition de dataset et la description des caractéristiques et les attributs de dataset du notre modèle des maladies du cancer de sein, puis nous comprend le cadre mis en œuvre et présenter les résultats avec logiciel utilisé depuis de jusqu'à la phase d'apprentissage et la validation des modèles de prédiction.

### 3.2. Présentation des outils utilisés :

Un PC personnel HP avec un processeur Intel (R) pentium (R) CPU B960 @ 2.20 GHz et une capacité mémoire de 4 Go avec Windows 10, 64 bit type système .

Ce présent document a pour objet le Data Mining avec l'outil Weka. Sa réalisation a nécessité l'utilisation de l'outil de Data Mining Weka, l'application du processus de Data Mining KDD et de l'exploitation de datasets au format ARFF.

L'outil Weka permet de faire du preprocessing, d'effectuer une classification et une régression, de réaliser un clustering, de déduire les règles d'association, de rechercher les corrélations entre attributs et de visualiser les données, sur des cas réels.

Le logiciel Weka satisfait les propriétés principales du Data Mining. En perspective, il est possible d'optimiser encore plus nos prédictions face à un cas précis en implémentant nos propres algorithmes dans cet outil.

### 3.3Dataset:

l'ensemble de données (dataset) présentés dans l'ensemble de données de cancer de sein du Wisconsin breast cancer de machine learning le référentiel est au format structuré de l'Université de Californie.

Cet ensemble de données contient 444 instances ,10 variables d'entrée(caractéristique)et 1 variable de sortie(Réponse) dans l'attribut de réponse « classe », 2 est remplacé par «bénigne» et 4 est remplacé par « malin ».

### 1.La description des caractéristiques de dataset en voir le tableau suivant :

Les caractéristiques de dataset	Multivarié
Caractéristique des attributs	Réel
Taches associés	Classification
Nombre d'instances	444
Nombre d'attributs	10
Valeurs manquantes	Non
Surface	Vie
Date du don	1995

Tableau 3.1 : Les caractéristique de dataset.

### 2.Attribue de dataset :

No	attribue En Anglais	En Français
01	Clump Thickness	Epaisseur de l'aggloméré.
02	Uniformity of Cell Size	Uniformité de la taille de la cellule.
03	Uniformity of Cell Shape	Uniformité de la forme de la cellule.
04	Marginal Adhesion	Adhésion marginale.
05	Single Epithelial Cell Size	Taille des cellules épithéliales simples.
06	Bare Nuclei	Noyau nu.
07	Bland Chromatin	Chromatine fade .
08	Normal Nucleoli	Nucléoles normales.
09	Mitoses	mitose.
10	Class	classe.

Tableau3.2 : Les attributs de la base de données « Brest-cancer-Wisconsin »

### 3.Description des attributs:



## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

---

- **Épaisseur de touffe** Les cellules regroupées en monocouches sont identifiées comme des cellules bénignes, tandis que les cellules regroupées en multicouches sont identifiées comme des cellules malignes, c'est-à-dire cancéreuses.
- **Uniformité de la taille des cellules** La taille des cellules diffère des cellules malignes aux cellules bénignes.
- **Uniformité de la taille des cellules** La taille des cellules diffère des cellules malignes aux cellules bénignes.
- **Uniformité de la forme des cellules** La forme des cellules diffère des cellules malignes aux cellules bénignes.
- **Adhérence marginale** : Les cellules normales sont de nature compacte tandis que les cellules cancéreuses sont de nature lâche.
- **Taille de cellule épithéliale unique** Les cellules épithéliales distendues et assez grandes peuvent être une cellule cancéreuse.
- **Noyaux nus** Les noyaux des cellules qui ne sont pas entourés de cytoplasme sont appelés noyaux nus ou nus. Cela se trouve dans les tumeurs bénignes.
- **Chromatine fade** Dans les cellules bénignes, on retrouve une texture fine et

uniforme du noyau tandis que la chromatine est plus grossière dans les cellules cancéreuses.

- **Nucléoles normaux** Généralement, les cellules contiennent des nucléoles à peine

visibles mais dans les cellules cancéreuses, les nucléoles sont proéminents.

- **Mitoses** Le processus de division des cellules normales varie de celui des cellules cancéreuses.
- **Classe** Attribut de réponse.

### 4.Division de dataset :

Nous avons d'abord identifié la base de données « Brest-cancer-Wisconsin », puis nous avons divisé la base de données en deux parties différentes ,partie pour l'entrainement"training"66% et une partie pour les tests 34% par l'utilisation de la fonction "train -test percentage split".

### 3.4.Logiciel Utilisé

Weka (Waikato Environment for Knowledge Analysis) est un ensemble d'outils permettant de manipuler et d'analyser des fichiers de données, implémentant la plupart des



## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

algorithmes d'intelligence artificielle, entre autres, les arbres de décision et les réseaux de neurones. Les algorithmes peuvent être appliqués directement à un ensemble de données ou appelés à partir d'un code Java. Il est également bien adapté au développement de nouveaux modèles de Machine Learning. Weka est utilisé dans divers domaines comme dans la recherche, l'éducation et même les entreprises.

Weka est un logiciel open source publié sous licence publique générale GNU. Il a été développé à l'Université de Waikato en Nouvelle-Zélande.

Weka est extensible et est devenu une collection d'algorithmes d'apprentissage automatique pour résoudre les problèmes d'exploration de données dans le monde réel. Il est écrit en Java et fonctionne sur presque toutes les plateformes. Il est facile à utiliser et à appliquer sur plusieurs niveaux différents. Il est possible d'accéder à la bibliothèque de classes Weka à partir d'un programme Java et implémenter de nouveaux algorithmes d'apprentissage automatique.

### Il se compose principalement :

- De classes Java permettant de charger et de manipuler les données.
- De classes pour les principaux algorithmes de classification supervisée ou non supervisée.
- D'outils de sélection d'attributs, de statistiques sur ces attributs.
- De classes permettant de visualiser les résultats.

### On peut l'utiliser à trois niveaux :

- Via l'interface graphique, pour charger un fichier de données, lui appliquer un algorithme, vérifier son efficacité.
- Invoquer un algorithme sur la ligne de commande.
- Utiliser les classes définies dans ses propres programmes pour créer d'autres méthodes, implémenter d'autres algorithmes, comparer ou combiner plusieurs méthodes.

La figure 1 : montre le logo du logiciel Weka.



Figure3.14. Logo de Weka

### Format d'entrée

Le format d'entrée par défaut de Weka est le format ARFF (Attribute Relation File Format). Mais il prend en charge aussi d'autres formats : CSV, JSON, fichiers Matlab, BSI, les bases de données utilisant ODBC, etc.

### 1.Pourquoi weka?

Weka est un langage L'open source intègre un grand nombre de technologies, intégrées et est peut-être le meilleur parmi les logiciels open source,

Mais cela nécessite une formation et une lecture attentive des documents d'accompagnement. Suggère 4 façons de traiter les données : Interface ligne par ligne

commandes, un explorateur pour le traitement des données et une application de plusieurs techniques d'exploration de données sur des bases de données

Constructions graphiques multiples de procédures de traitement de données. Fournit une construction de variables, un apprentissage supervisé et sans surveillance, Règles d'engagement.

### 2.les avantages de weka:

- Gratuité: libre et gratuit, distribué selon les termes de la licence publique générale GNU.

Le programme contient un ensemble d'outils de traitement de données et de techniques de modélisation.

- Le programme dispose d'une interface utilisateur graphique facile à utiliser.
- Le programme prend en charge plusieurs types de formats de fichiers différents tels que (.ARFF, .CSV, .C4).
- Il peut traiter d'autres programmes (tels que EXCEL2016.MS)

### 3.Installation su Windows :

Étape 1 : Téléchargez le logiciel à partir d'ici .

Vérifiez la configuration du système informatique et téléchargez la version stable de WEKA (actuellement 3. à partir de cette page.

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

Étape 2 : Une fois le téléchargement réussi, ouvrez l'emplacement du fichier et double-cliquez sur le fichier téléchargé. L'assistant Step Up apparaîtra. Cliquez sur Suivant.

Étape 3 : Les termes du contrat de licence s'ouvriront. Lisez-le attentivement et cliquez sur "J'accepte".

Étape 4 : Selon vos besoins, sélectionnez les composants à installer. L'installation complète des composants est recommandée. Cliquez sur Suivant.

Étape 5 : Sélectionnez le dossier de destination et cliquez sur Suivant.

Étape 6 : Ensuite, l'installation commencera

Étape 7 : Si Java n'est pas installé dans le système, il installera d'abord Java.

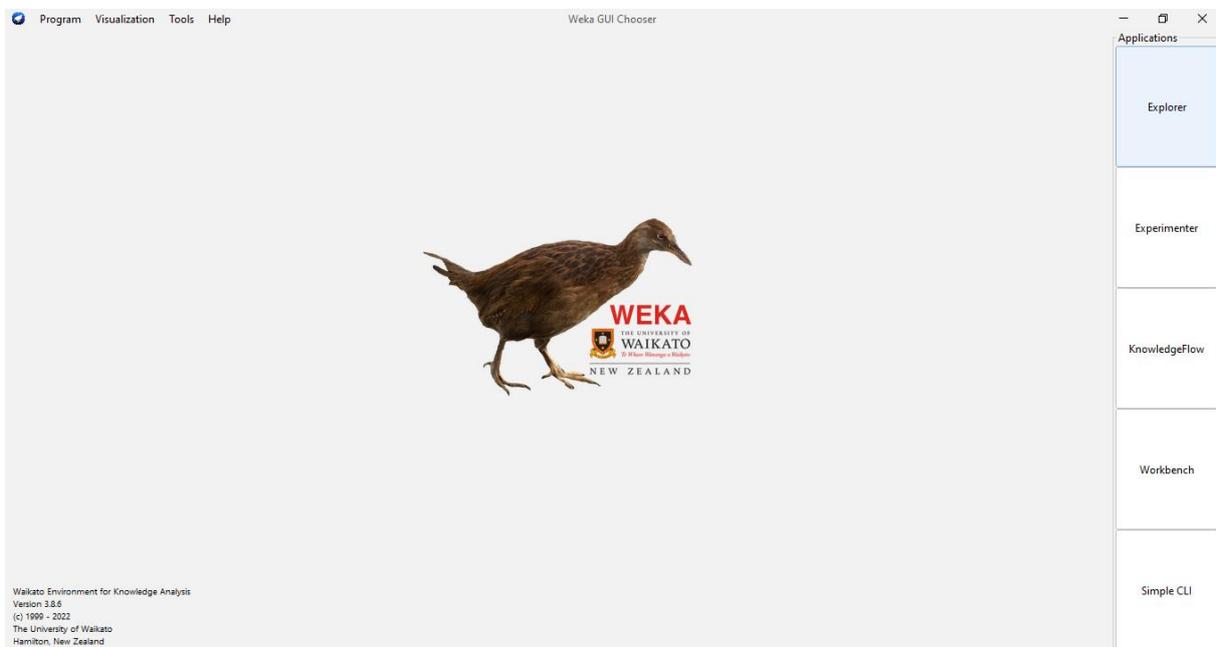
Étape 8 : Une fois l'installation terminée, la fenêtre suivante apparaîtra. Cliquez sur Suivant.

Étape 9 : Cochez la case Démarrer Weka. Cliquez sur Terminer.

Étape 10 : La fenêtre de l'outil et de l'explorateur WEKA s'ouvre.

Étape 11 : Le manuel WEKA peut être téléchargé ici.

## 4.Description générale de l'interface de weka (GUI)



Interface utilisateur graphique de WEKA

## **Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique**

---

L'interface graphique de WEKA propose cinq options : Explorer, Expérimenter, Knowledge flow, Workbench et Simple CLI. Comprenons chacun d'eux individuellement.

### 1) CLI simple ( Simple CLI) :

La CLI simple est Weka Shell avec ligne de commande et sortie. Avec "aide", la vue d'ensemble de toutes les commandes peut être vue. Simple CLI offre un accès à toutes les classes telles que les classificateurs, les clusters et les filtres, etc.

### 2) Explorateur (Explorer) :

Les fenêtres de l'explorateur WEKA affichent différents onglets en commençant par le prétraitement. Initialement, l'onglet de prétraitement est actif, car l'ensemble de données est d'abord prétraité avant de lui appliquer des algorithmes et d'explorer l'ensemble de données.

### 3) Expérimentateur (Experimenter) :

Le bouton Expérimentateur WEKA permet aux utilisateurs de créer, d'exécuter et de modifier différents schémas en une seule expérience sur un ensemble de données. L'expérimentateur dispose de 2 types de configuration : Simple et Avancé. Les deux configurations permettent aux utilisateurs d'exécuter des expériences localement et sur des ordinateurs distants.

### 4) Flux de connaissances (Knowledge Flow) :

Le flux de connaissances montre une représentation graphique des algorithmes WEKA. L'utilisateur peut sélectionner les composants et créer un flux de travail pour analyser les ensembles de données. Les données peuvent être traitées par lots ou de manière incrémentielle. Des flux de travail parallèles peuvent être conçus et chacun s'exécutera dans un thread séparé.

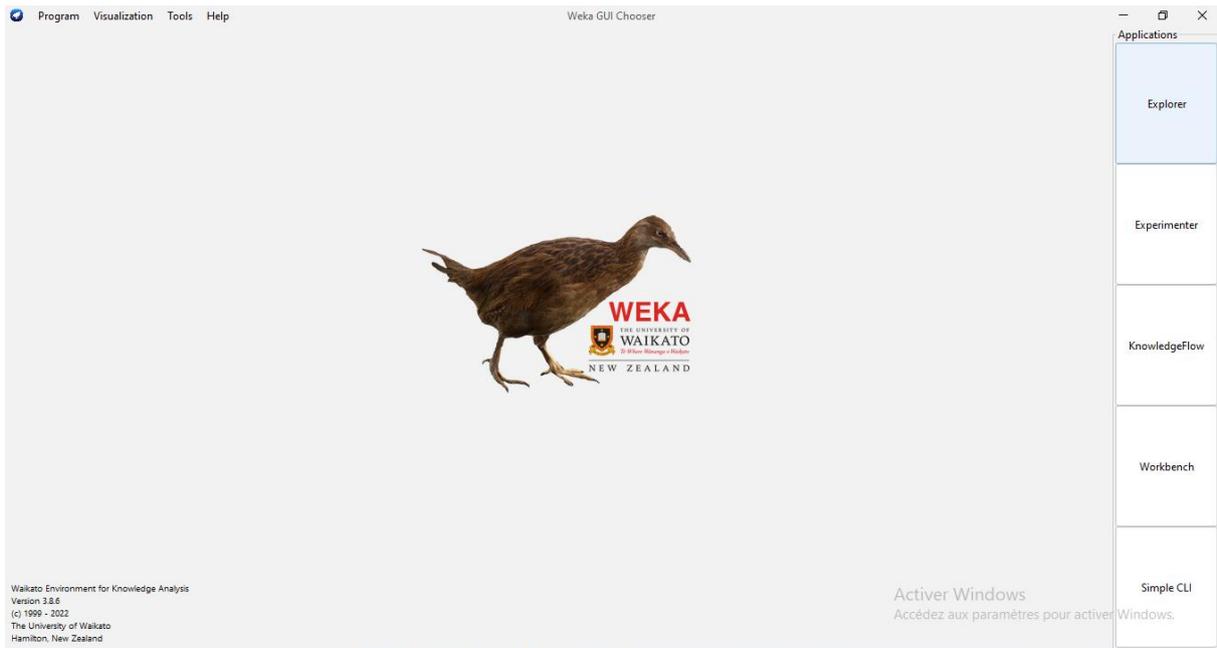
### 5) Établi (Workbench) :

WEKA a un module workbench qui contient toutes les interfaces graphiques dans une seule fenêtre.

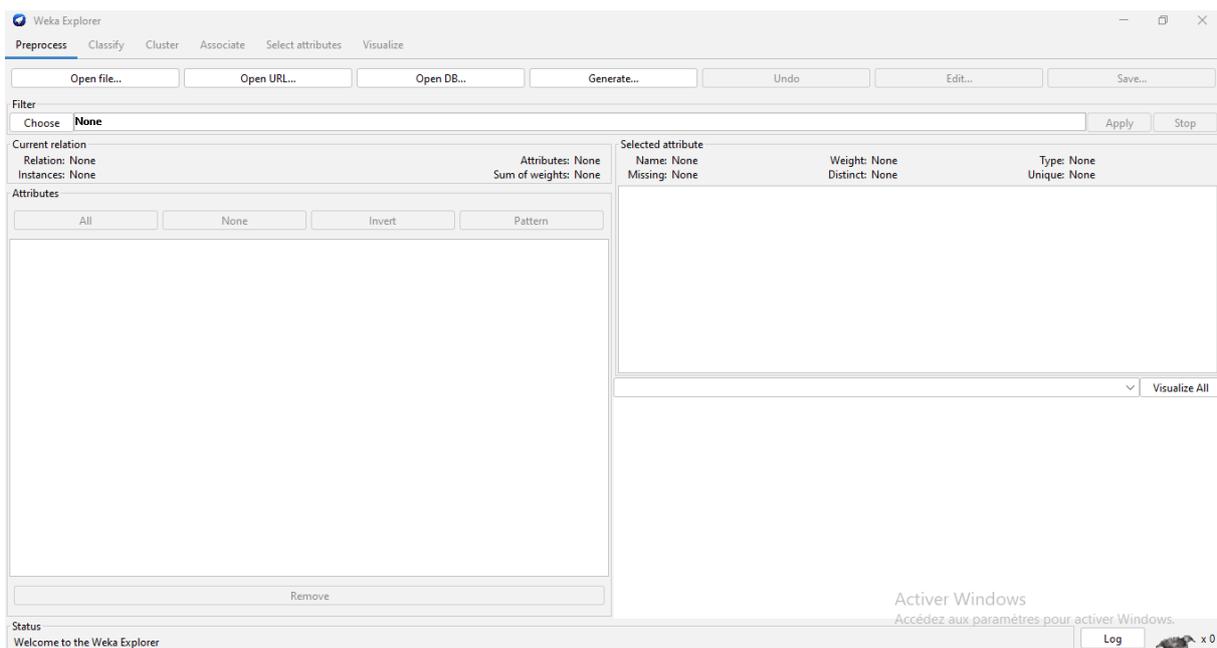
## **3.5.Les étapes de Processus :**

- **Première étape: ouvrez le waka GUI**

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique



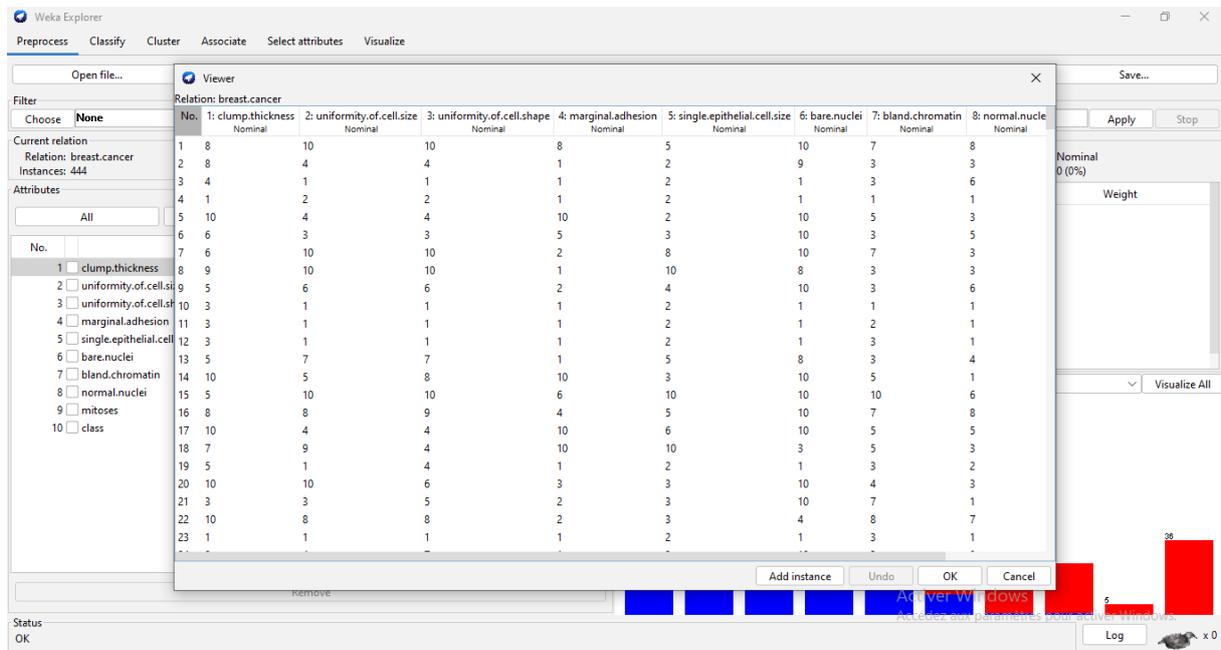
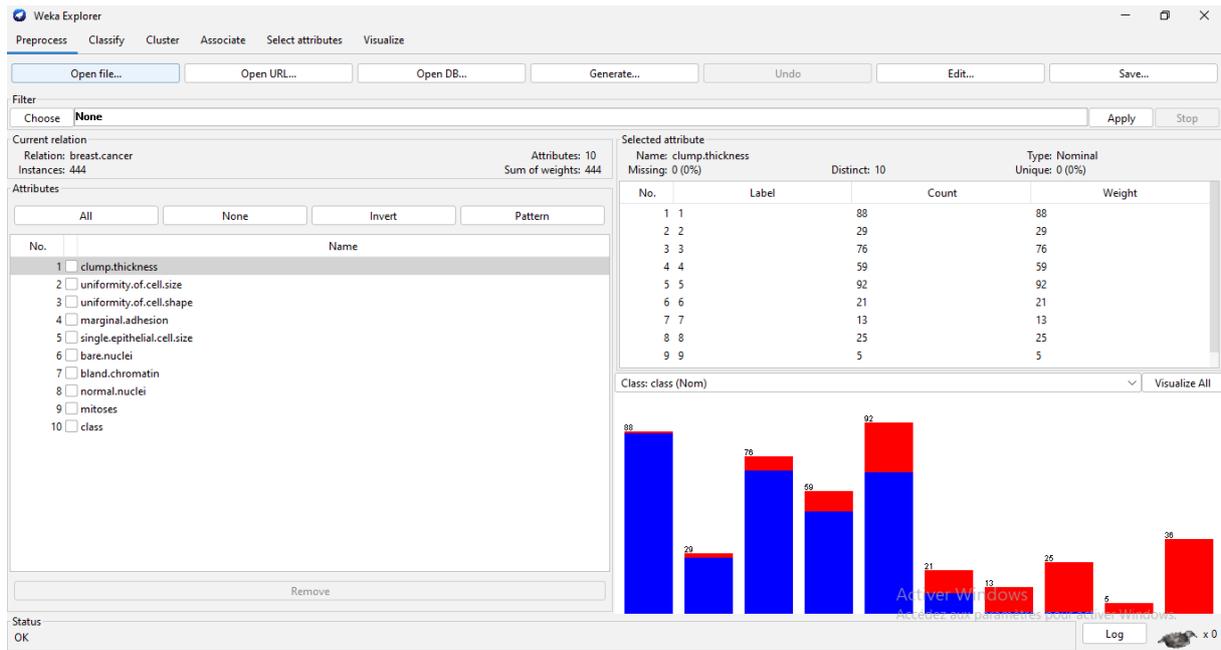
➤ **deuxième étape: choisir et cliquez sur le bouton "Explorer" pour ouvrir l'exploration weka.**



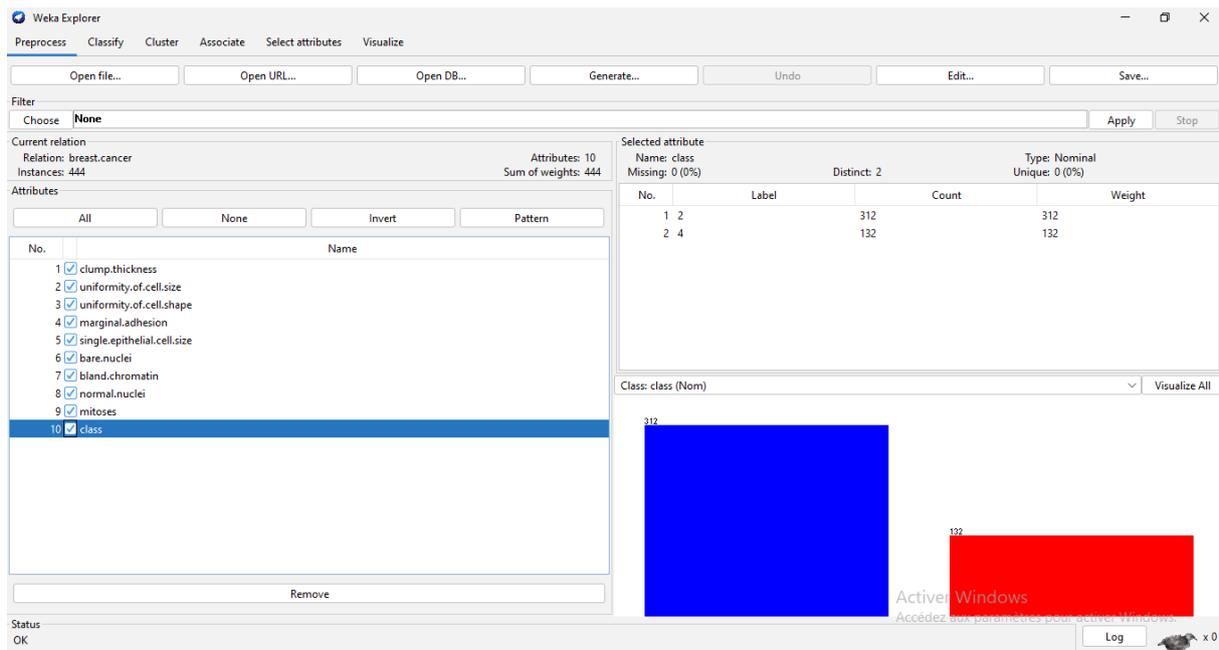
➤ **troisième étape :** charger L'ensemble de données à partir de fichier data.breast cancer

➤ **quatrième étape :**chargement de la base des données.

# Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique



Cet Histogramme représente le nombre de personnes atteintes d'une tumeur sont 132, et le nombre de personnes qui ne sont pas cancéreux sont 312.



The screenshot shows the Weka Explorer interface. The 'Selected attribute' table is as follows:

No.	Label	Count	Weight
1	2	312	312
2	4	132	132

Below the table is a bar chart with two bars: a blue bar for label '2' with a count of 312, and a red bar for label '4' with a count of 132. The interface also shows a list of attributes on the left, with 'class' selected at the bottom.

### 3.6. Application des algorithmes :

Les modèles utilisés pour la prédiction de cancer de sein, Comme leur nom sur Logiciel Weka est le suivant :

- 1-kNN (K plus proches voisines) (IBK).
2. L'arbre de décision (Decision tree) (REPTree).
3. SVM (support vector Machin)(SMO).
4. Random Forest (les forêts aléatoires ou les forêts de décision Aléatoires)(Random Forest).
5. Naïve bayes (Gaussian Naive Bayes)(Naive Bayes).

### 3.7. Définition des paramètres statistiques utilisés :

➤ **Le coefficient kappa de Cohen ( $\kappa$ ) :** est une statistique utilisée pour mesurer la fiabilité inter-évaluateur (et également la fiabilité intra-évaluateur ) pour les éléments qualitatifs (catégoriels). On pense généralement qu'il s'agit d'une mesure plus robuste que le simple calcul du pourcentage d'accord, car  $\kappa$  prend en compte la possibilité que l'accord se produise par hasard. Il existe une controverse autour du kappa de Cohen en raison de la

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

difficulté à interpréter les indices d'accord. Certains chercheurs ont suggéré qu'il est conceptuellement plus simple d'évaluer le désaccord entre les éléments.

➤ **Erreur absolue moyenne (MAE) :** Le MAE mesure l'ampleur moyenne des erreurs dans un ensemble de prévisions, sans tenir compte de leur direction. Il mesure la précision des variables continues. L'équation est donnée dans les références de la bibliothèque. Exprimée en lettres, la MAE est la moyenne sur l'échantillon de vérification des valeurs absolues des écarts entre la prévision et l'observation correspondante. Le MAE est un score linéaire ce qui signifie que toutes les différences individuelles sont pondérées de manière égale dans la moyenne.

➤ **Erreur quadratique moyenne (RMSE) :** La RMSE est une règle de notation quadratique qui mesure l'ampleur moyenne de l'erreur. L'équation de la RMSE est donnée dans les deux références. Exprimant la formule en mots, la différence entre les prévisions et les valeurs observées correspondantes est chacune mise au carré puis moyennée sur l'échantillon. Enfin, la racine carrée de la moyenne est prise. Étant donné que les erreurs sont mises au carré avant d'être moyennées, la RMSE donne un poids relativement élevé aux erreurs importantes. Cela signifie que la RMSE est plus utile lorsque de grandes erreurs sont particulièrement indésirables.

- Le MAE et le RMSE peuvent être utilisés ensemble pour diagnostiquer la variation des erreurs dans un ensemble de prévisions. Le RMSE sera toujours supérieur ou égal au MAE ; plus la différence entre eux est grande, plus la variance des erreurs individuelles dans l'échantillon est grande. Si le  $RMSE=MAE$ , alors toutes les erreurs sont de même grandeur
- Le MAE et le RMSE peuvent aller de 0 à  $\infty$ . Ce sont des scores orientés négativement : les valeurs inférieures sont meilleures.

➤ **Erreur absolue :** L'erreur absolue est une mesure de la distance entre une mesure et une valeur vraie ou une indication de l'incertitude d'une mesure. Par exemple, si vous mesurez la largeur d'un livre à l'aide d'une règle avec des marques millimétriques, le mieux que vous puissiez faire est de mesurer la largeur du livre au millimètre près. Vous mesurez le livre et trouvez qu'il fait 75 mm. Vous signalez l'erreur absolue dans la mesure comme 75 mm +/- 1 mm. L'erreur absolue est de 1 mm. Notez que l'erreur absolue est rapportée dans les mêmes unités que la mesure.

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

Alternativement, vous pouvez avoir une valeur connue ou calculée et vous souhaitez utiliser l'erreur absolue pour exprimer à quel point votre mesure est proche de la valeur idéale. Ici, l'erreur absolue est exprimée comme la différence entre les valeurs attendues et réelles.

**Erreur absolue = valeur réelle - valeur mesurée**

Par exemple, si vous savez qu'une procédure est censée produire 1,0 litre de solution et que vous obtenez 0,9 litre de solution, votre erreur absolue est de  $1,0 - 0,9 = 0,1$  litre.

**Erreur relative :** Vous devez d'abord déterminer l'erreur absolue pour calculer l'erreur relative. L'erreur relative exprime la taille de l'erreur absolue par rapport à la taille totale de l'objet que vous mesurez. L'erreur relative est exprimée sous forme de fraction ou est multipliée par 100 et exprimée en pourcentage .

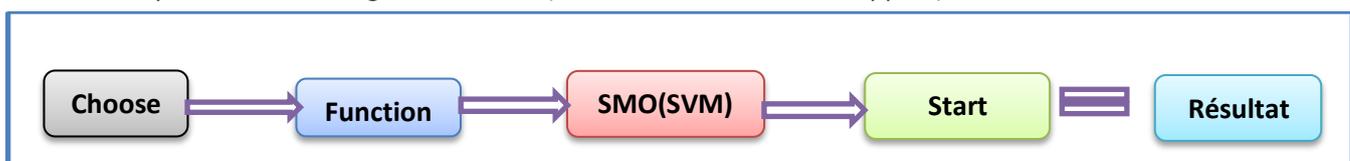
**Erreur relative = erreur absolue / valeur connue**

Par exemple, le compteur de vitesse d'un conducteur indique que sa voiture roule à 60 miles par heure (mph) alors qu'elle roule en réalité à 62 mph. L'erreur absolue de son compteur de vitesse est de  $62 \text{ mph} - 60 \text{ mph} = 2 \text{ mph}$ . L'erreur relative de la mesure est de  $2 \text{ mph} / 60 \text{ mph} = 0,033$  ou 3,3%

### 3.8.Application et validation

#### 1 .Modèle 01 :

Implémentation d'algorithmme SVM (Machines à vecteurs de support):



#### Résultats :

On ce qui suit les résultats de modèle « SVM » de l'analyse du cancer de sein a été effectué sur 10 Attributs pour prédire de cancer de sein .

Instances correctement classées	429	96.6216 %
Instances mal classées	15	3.3784 %
Statistique Kappa		0.919

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

Erreur absolue moyenne	0.0338
Erreur quadratique moyenne	0.1338
Erreur absolue relative	8.0775 %
Erreur quadratique relative racine	40.2122 %
Nombre total d'instances	444

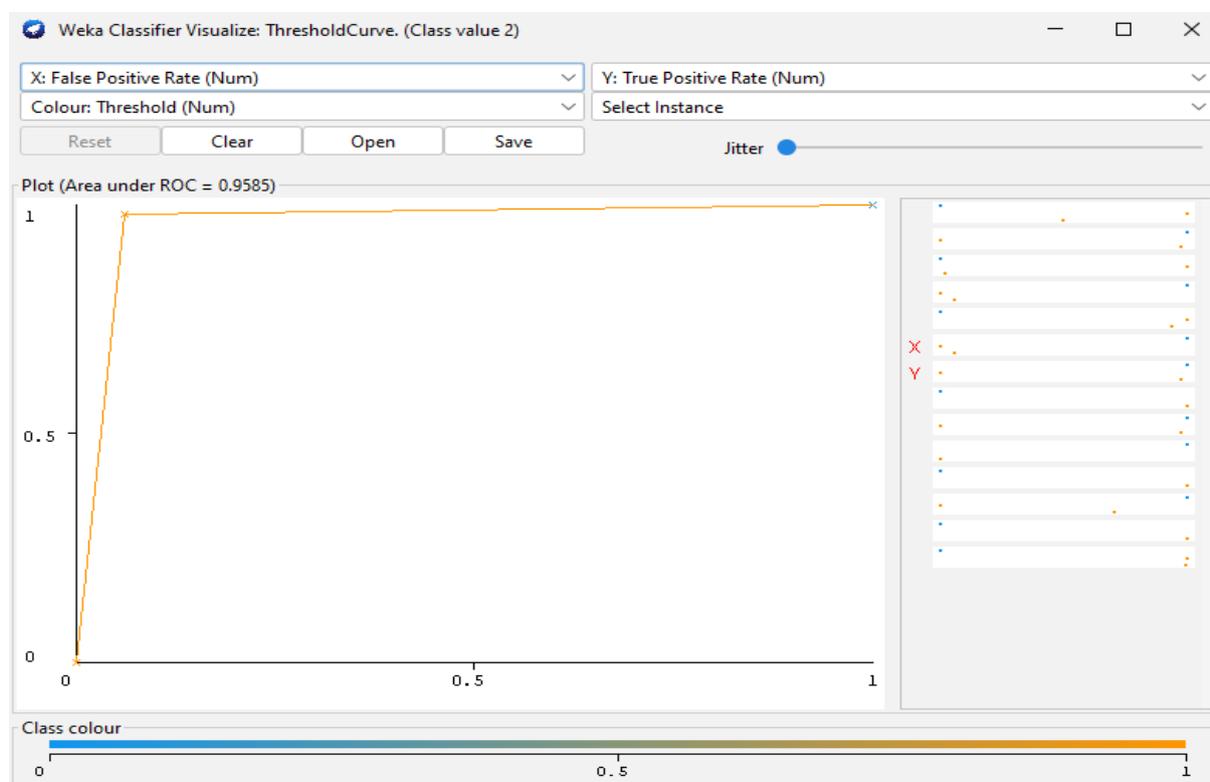
**Tableau3.3: Résumé de l'Algorithme svm (machine à vecteurs de support).**

TP Rate	FP Rate	Precision	Recall	Class
0,978	0,978	0,974	0,978	Bening
0,939	0,022	0,947	0,939	Malignant

**Tableau3.4:mesures de précision pourSVM (machine à vecteurs de support).**

Classifieur	Benign	Malignant
Session 1	305	7
Session 2	8	124

**Tableau3.5:Matrice de confusion pour l'Algorithme SVM (machine à vecteurs de support).**



**Figure 3.15 : Courbe de classificateur SVM (machine à vecteurs de support).**

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

### ❖ Analyses et interprétations :

D'après le tableau ci-dessus, nous avons une valeur de Kappa égale à 91.9% ;ainsi, en utilisant cette échelle de Kappa, nous savons que notre prédiction appartient à (Accord presque parfait), avec une instance distribuée légèrement étalée en fonction de la valeur de l'erreur absolue moyenne (0,0338), la distribution de la moyenne est autour de l'unité ( 1).

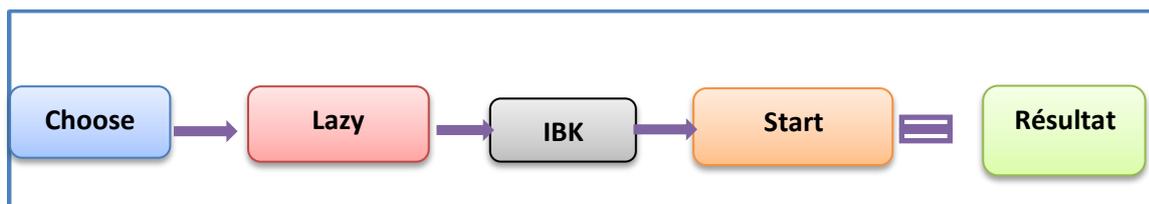
De plus, dans le deuxième tableau des sorties SVM, nous avons une valeur (Rappel) égale à 97,8 % pour Bening et 93,9 % pour Malignant, cela signifie que la sensibilité de notre modèle est forte et qu'il est capable de sélectionner une certaine classe dans notre ensemble de données. (La valeur de rappel est plus élevée, donc la valeur de TP -True positive- est également supérieure à la valeur FN -False Negative-). instances d'une certaine classe à partir d'un ensemble de données est excellente.

Nous pouvons confirmer qu'en examinant la valeur de précision, et elle est de 97,4 %, cela signifie que notre classe a été si bien prédite dans l'ensemble de données. on peut dire que les classes ont été prédites. avec excellence, une classe bénigne et une classe maligne dans l'ensemble de données.

Les résultats nous montrent que le graphe ROC pour SVM est excellent et que notre modèle est fiable.

### 2.Modèle 02 :

Implémentation d'algorithme KNN (Méthode de K plus proches voisins)



Instances correctement classées	428	( 96.3964 %)
Instances mal classées	16	(3.6036 %)
Statistique Kappa		0.9122
Erreur absolue moyenne		0.0393
Erreur quadratique moyenne		0.1762
Erreur absolue relative		9.4005 %
Erreur quadratique relative racine		38.5569 %

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

Nombre total d'instances	444
--------------------------	-----

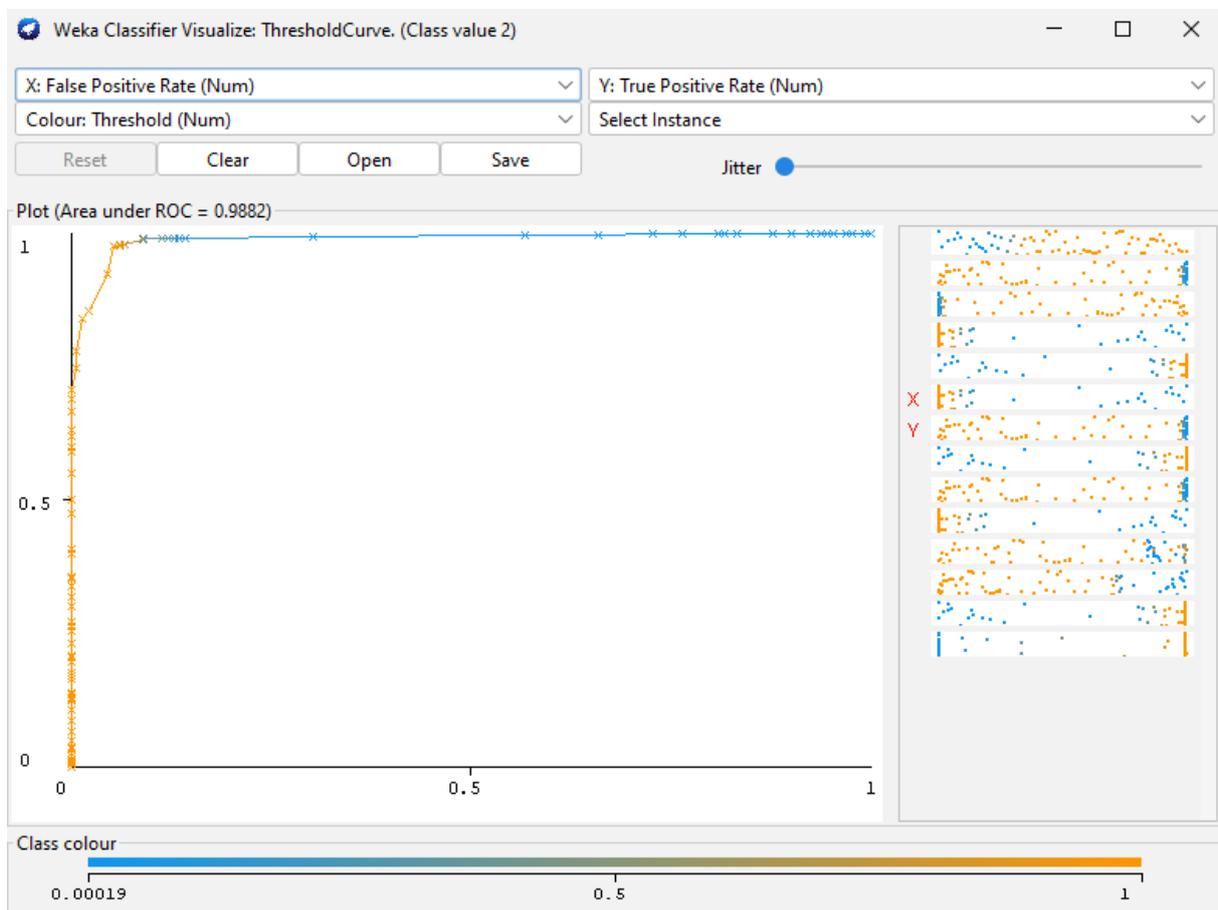
**Tableau3.6: Résumé de l'Algorithme KNN (Méthode de K plus proches voisins).**

TP Rate	FP Rate	Precision	Recall	Class
0,987	0,091	0,963	0,987	Bening
0,909	0,013	0,968	0,909	Malignant

**Tableau3.7:mesures de précision pour KNN (Méthode de K plus proches voisins).**

Classifieur	Benign	Malignant
Session 1	308	4
Session 2	12	120

**Tableau3.8:Matrice de confusion pour KNN (Méthode de K plus proches voisins).**



**Figure 3.16 : Courbe de classificateur KNN (Méthode de K plus proches voisins).**

### ❖ Analyses et interprétations :

Nous savons que notre prédiction appartient à l'échelon (presque parfait), avec une instance assez largement distribuée basée sur la valeur de l'erreur absolue moyenne, car la

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

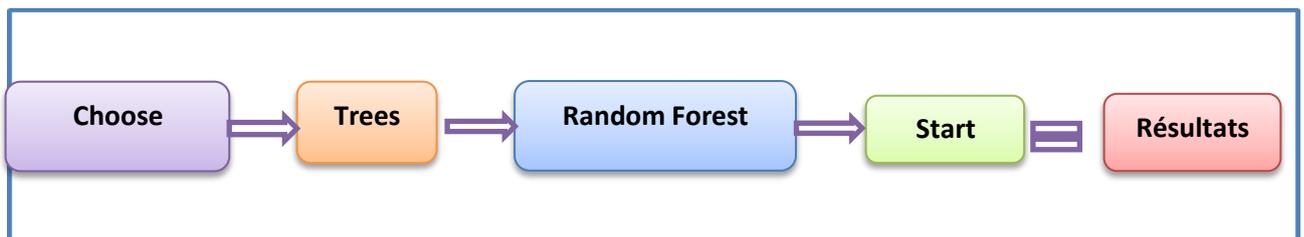
valeur de kappa est égale à 91,22 % dans le tableau ci-dessus, l'erreur Absolut égale (0,033) il est indiqué que notre instance n'est pas distribuée d'une manière étendue.

Dans le tableau séparé des sorties KNN, nous obtenons une valeur (Recall) de 98,7 % pour Bening et de 90,9 % pour Malignant, ce qui indique que la sensibilité de notre modèle est élevée et qu'il peut identifier correctement une classe à partir de notre ensemble de données. Et les valeurs de précision pour Bening et Malignant sont toutes deux de 96 %, ce qui indique que la capacité d'un modèle de prédiction à choisir des instances d'une classe spécifique d'un ensemble de données est excellente. Nous pouvons vérifier cela en examinant la valeur de précision, qui est de 97,4 %, démontrant que notre classe a été prédite avec succès depuis l'intérieur de l'ensemble de données. On peut dire que les cours ont été prévus. Dans l'ensemble de données, il existe une classe bénigne et une classe maligne.

Si on Plot les résultats en ROC Graph, nous assurons que le graph indiquera KNN comme un excellent, et que notre modèle est fiable.

### 3.Modèle 03:

implémentation d'algorithme les forêts aléatoires ou les forêts de décision aléatoires.



Instances correctement classées	429	(96.6216 %)
Instances mal classées	15	( 3.3784 %)
Statistique Kappa		0.919
Erreur absolue moyenne		0.0659
Erreur quadratique moyenne		0.1634
Erreur absolue relative		15.7457 %
Erreur quadratique relative racine		35.7498 %
Nombre total d'instances		444

**Tableau3.15: Résumé de l'Algorithme les forêts aléatoires ou les forêts de décision Aléatoires.**

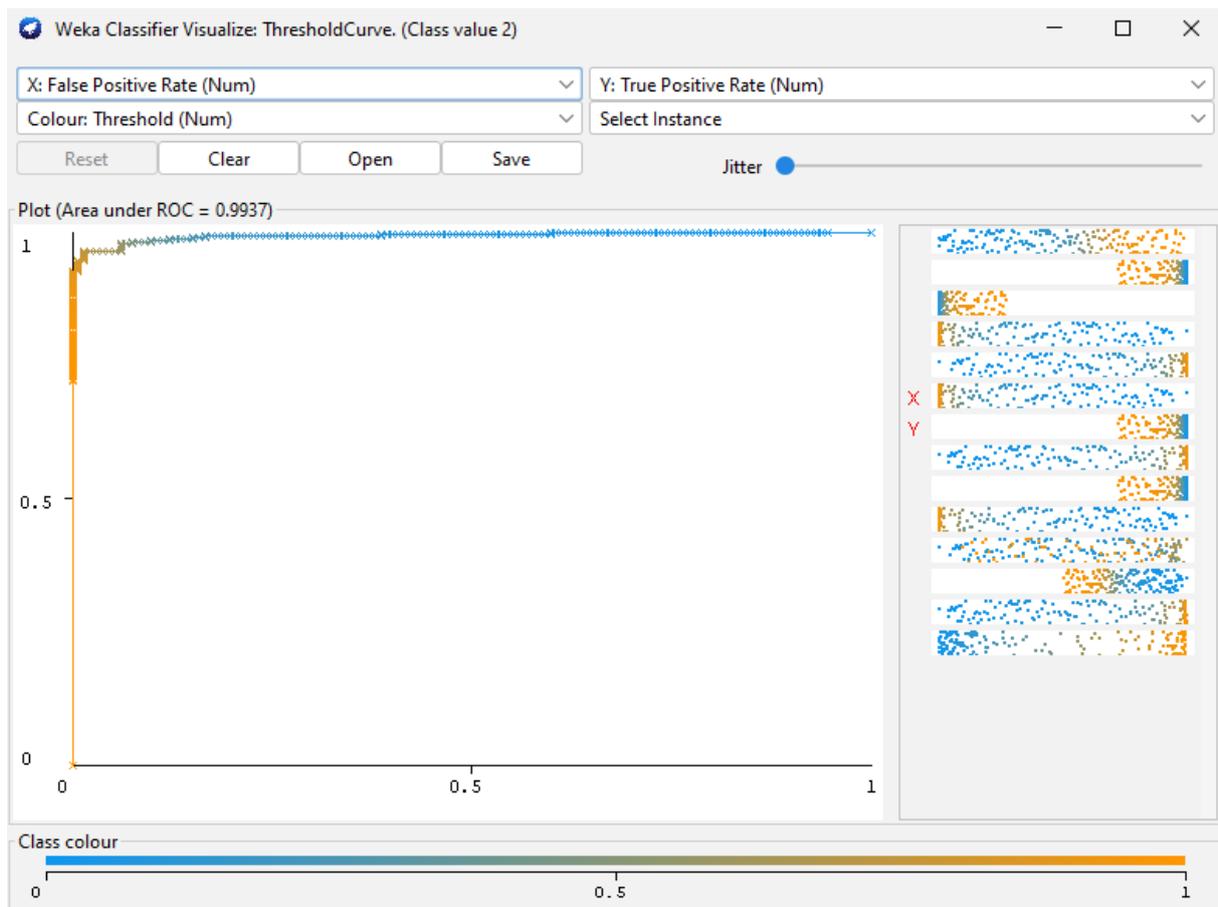
## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

TP Rate	FP Rate	Precision	Recall	Class
0,978	0,061	0,974	0,978	Bening
0,939	0,022	0,947	0,939	Malignant

**Tableau3.10:mesures de précision pour les forêts aléatoires ou les forêts de décision Aléatoires.**

Classifier	Benign	Malignant
Session 1	305	7
Session 2	8	124

**Tableau3.11:Matrice de confusion pour l'Algorithme les forêts aléatoires ou les forêts de décision aléatoires.**



**Figure 3.17 : Courbe de classificateur les forêts aléatoires ou les forêts de décision aléatoires.**

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

### ❖ Analyses et interprétations :

Étant donné que notre ensemble de données contient deux classes distinctes qui sont différentes l'une de l'autre et que les valeurs ne s'éloignent pas de la valeur moyenne, nous avons constaté que kappa est égal à 91,9 % dans le tableau RF Algorithm. Cela indique que notre modèle de prédiction est très bon, avec une instance distribuée légèrement étalée en fonction de la valeur de l'erreur absolue moyenne (0,0338).

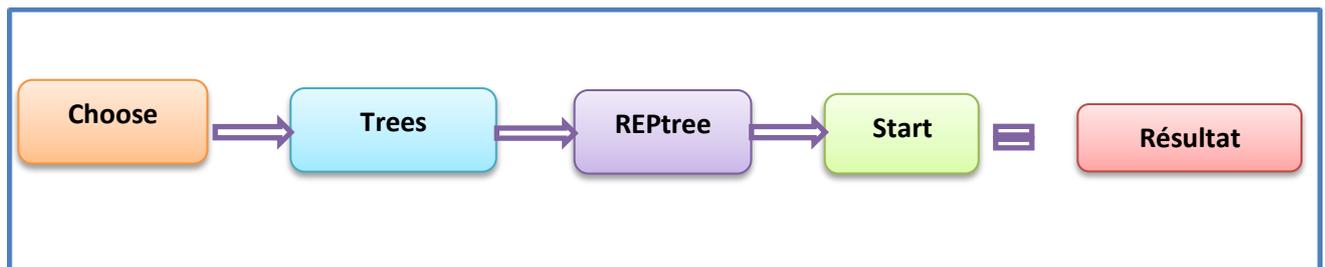
Le deuxième tableau, qui appartient également aux sorties RF, a une valeur (Rappel) de 97,4 % pour Bening et de 93,9 % pour Malignant, ce qui indique que la sensibilité de notre modèle est élevée et qu'il peut choisir une certaine classe dans notre ensemble de données. De plus, les scores de précision pour Bening sont de 97,4 % et pour Malignant de 97,4 %, ce qui indique que la capacité d'un modèle de prédiction à choisir des instances d'une classe spécifique à partir d'un ensemble de données est excellente.

Cependant, le nombre de précision, qui est de 97,4 %, indique que notre classe a été prédite avec précision dans l'ensemble de données. On peut dire que les deux classes ont été prévues. Dans l'ensemble de données, il existe une classe bénigne et une classe maligne.

Le graphique ROC pour RF est exceptionnel, et notre modèle est fiable, selon les résultats.

### 4.Modèle 04:

Implémentation d'Algorithme d'arbre de décision (REPTree)



Instances correctement classées	421	( 94.8198 %)
Instances mal classées	23	( 5.1802 %)
Statistique Kappa		0.8747
Erreur absolue moyenne		0.082
Erreur quadratique moyenne		0.2202
Erreur absolue relative		19.6011 %

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

Erreur quadratique relative racine	48.17 %
Nombre total d'instances	444

Tableau3.12: Résumé de l'Algorithme Arbre de décision (REPTree).

TP Rate	FP Rate	Precision	Recall	Class
0,971	0,106	0,956	0,971	Bening
0,894	0,029	0,929	0,894	Malignant

Tableau3.13:mesures de précision pour Arbre de décision (REPTree).

Classifieur	Benign	Malignant
Session 1	303	9
Session 2	14	118

Tableau3.14:Matrice de confusion pour Arbre de décision (REPTree).

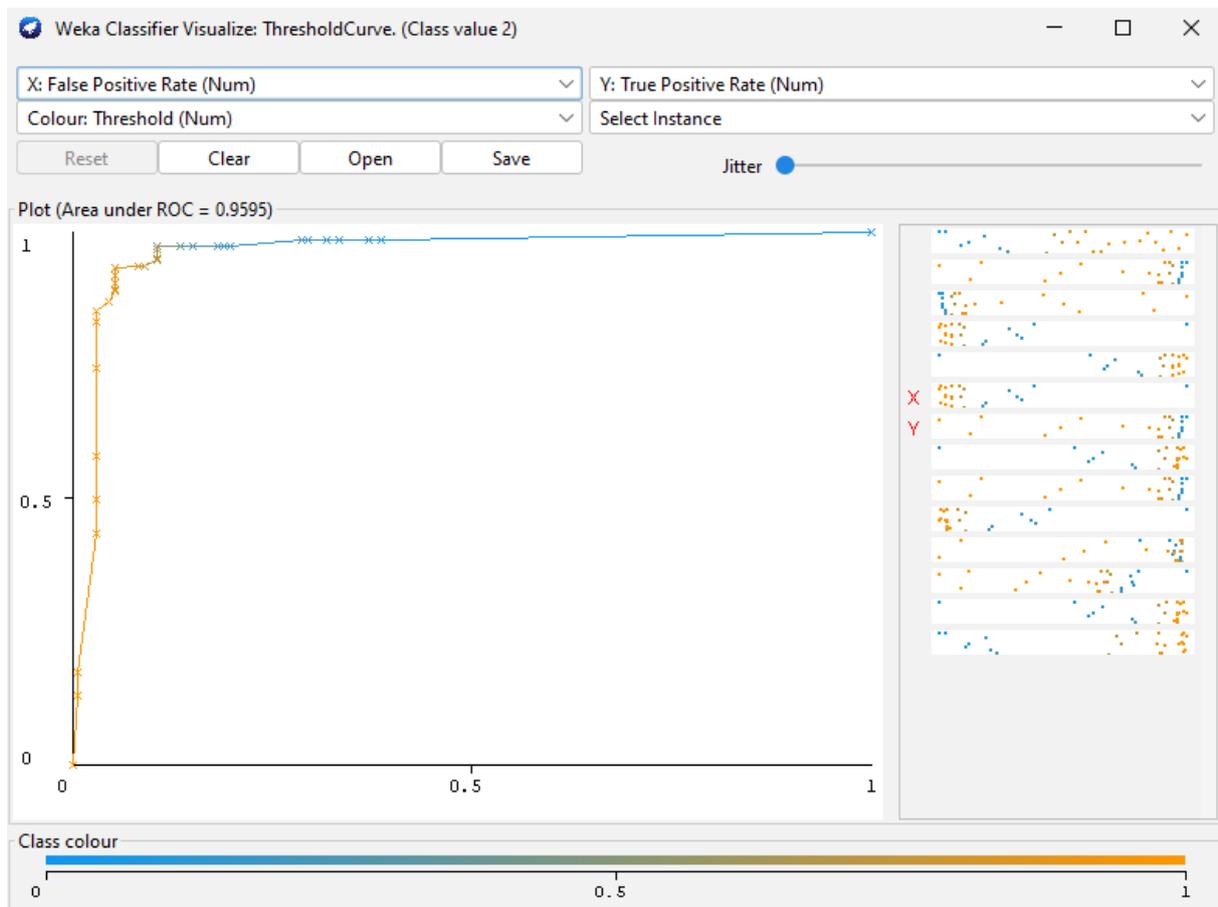


Figure3.18 : Courbe de classificateur Arbre de décision (REPTree).

❖ Analyses et interprétations :

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

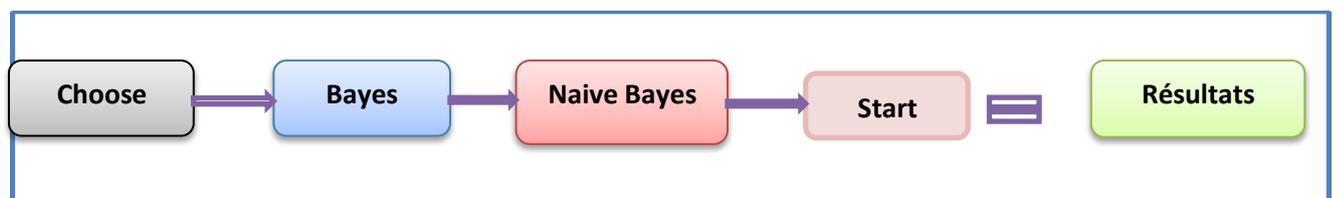
Nous allons maintenant examiner les résultats dans les tables DT (REPTree). Il ressort clairement des tableaux de l'algorithme que la valeur kappa est de 87,47 %, ce qui est un peu inférieur aux deux autres algorithmes indiqués précédemment, mais toujours dans la plage d'une bonne prédiction. Avec une erreur absolue moyenne de 0,082, il existe une instance dispersée raisonnablement petite autour de la valeur moyenne. Nous voyons également que l'algorithme avait un nombre d'erreur quadratique moyen enraciné assez élevé ; cela suggère que l'erreur quadratique moyenne est élevée et qu'il existe un écart entre les instances mesurées dans chaque classe.

De plus, la valeur de précision, qui est de 95,6 %, indique que notre classe a été prédite avec précision dans l'ensemble de données. On peut dire que les deux classes ont été prévues. Dans l'ensemble de données, les classes bénignes et malignes ont une précision bonne et raisonnable. De plus, le système a une grande possibilité de prédire correctement Bening, avec une précision de 97% contre 89% dans la classe maligne.

Bien que les résultats de DT (REPTree) montrent qu'il est moins efficace que les autres, il a un graphique ROC décent et notre modèle est fiable.

### 5.Modèle 05 :

Implémentation d'algorithme Naïve bayes:



Les tableaux suivants présentent les résultats d'application de " Naïve Bayes" sur la base des données

Instances correctement classées	432	( 97.2973 %)
Instances mal classées	12	( 2.7027 %)
Statistique Kappa		0.9362
Erreur absolue moyenne		0.0258
Erreur quadratique moyenne		0.1539
Erreur absolue relative		6.1594 %

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

Erreur quadratique relative racine	33.6744 %
Nombre total d'instances	444

Tableau3.15: Résumé de l'Algorithme Naïve bayes.

TP Rate	FP Rate	Precision	Recall	Class
0,971	0,971	0,990	0,971	Benign
0,977	0,029	0,935	0,977	malignant

Tableau3.16:mesures de précision pour Naïve bayes.

Classifieur	Benign	Malignant
Session 1	303	9
Session 2	3	129

Tableau3.17:Matrice de confusion pour l'Algorithme Naïve bayes.

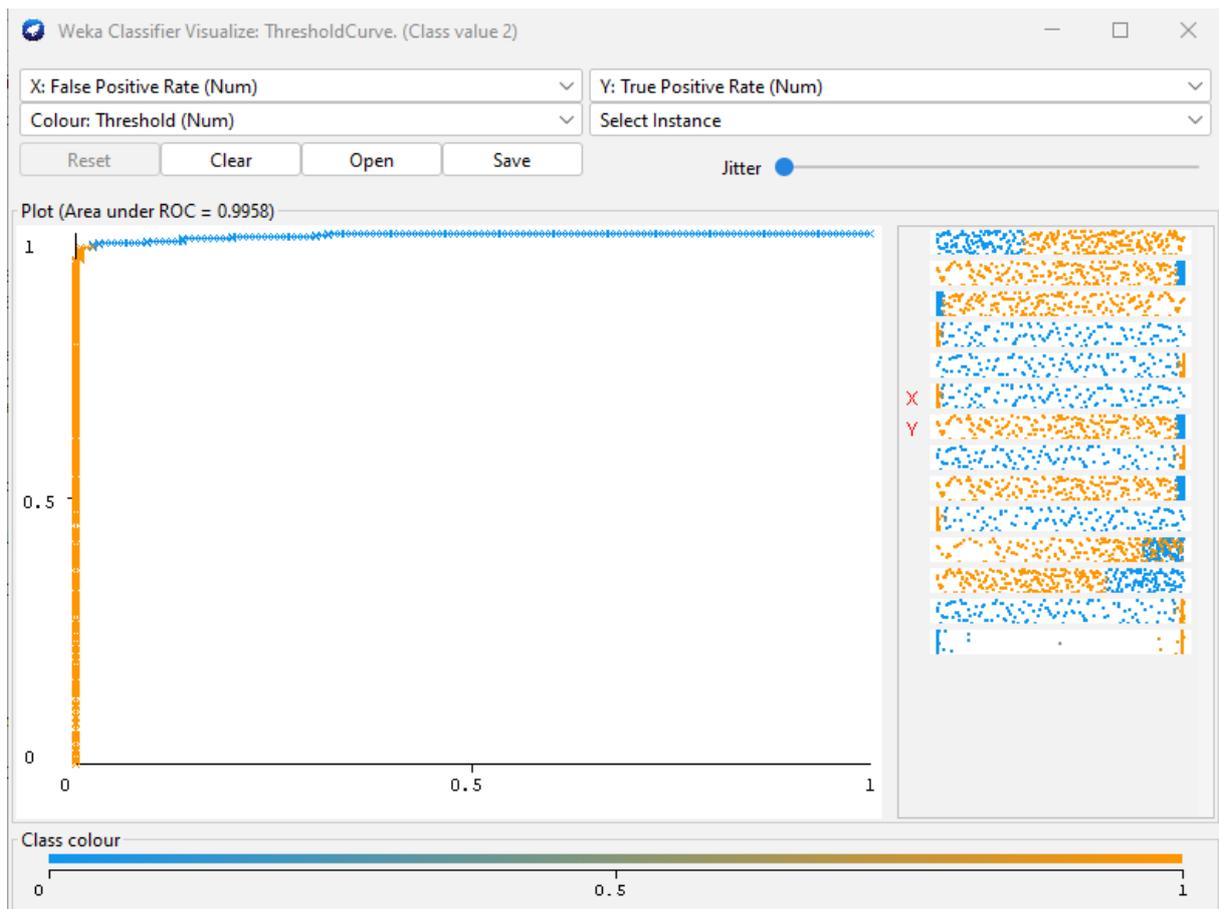


Figure 3.19 : Courbe ROC de classificateur Naïve bayes.

### ❖ Analyses et interprétations :

Les tableaux de l'algorithme Naïve Bites montrent une valeur kappa égale à 93,62 % ; son indication nous indique que notre modèle de prédiction est très bon, avec une instance distribuée légèrement étalée en fonction de la valeur de l'erreur absolue moyenne (0,0258), car notre jeu de données contient deux classes distinctes différentes l'une de l'autre, et les valeurs ne vont pas loin de la valeur moyenne.

les sorties du deuxième tableau, nous avons une valeur (Rappel) égale à 97,1 % pour Bening et 97,7 % pour Malignant, cela signifie que la sensibilité de notre modèle est forte et qu'il est capable de sélectionner une certaine classe dans notre ensemble de données avec une chance égale. De plus, les valeurs de précision pour Bening sont de 99,0 % et pour Malignant de 93,5 %, de sorte que la capacité d'un modèle de prédiction à sélectionner des instances d'une certaine classe à partir d'un ensemble de données est excellente et, en même temps, les chances de bénin sont plus élevées.

Les résultats nous montrent que le graphique ROC pour NB est excellent, néanmoins des valeurs légèrement inférieures mais notre modèle est toujours fiable.

### 3.9.Évaluation des performances:

Nous avons utilisé les mesures du modèle pour évaluer et montrer les méthodes d'apprentissage automatique à partir de la matrice de confusion elle disgne la proportion des prédictions correctes effectué par le modèle et parmi lesquelles la précision est définie

#### Accuracy :

La précision (Accuracy ) est la mesure de performance la plus intuitive et il s'agit d'un rapport entre les observations correctement prédites et les observations totales.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

- TP : est le nombre de vrais positifs
- FP : le nombre de faux positifs.
- TN : est le nombre de vrais négatifs
- FN : le nombre de faux négatifs.

Précision :Cela définit l'exactitude du modèle en termes de prédiction



## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

$$\text{Précision} = \frac{VP}{VP+FP} \text{ pour cas bénigne}$$

$$\text{Précision} = \frac{VN}{VN+FN} \text{ pour cas malin}$$

Recall (Sensitivity) : Cette mesure de performance implique comment différentes valeurs et variables indépendantes affectent une variable dépendante.

$$\text{Recall} = \frac{VP}{VP+FN} \text{ pour cas bénigne}$$

La Sensibilité : Elle mesure le pourcentage de valeurs positives correctement identifiées

(*TruePositive*) par le modèle (e.g., le pourcentage de patients correctement identifiés comme atteints par la maladie). Elle représente également le rappel (the recall) du modèle ou la sensibilité du modèle. Elle est calculée en utilisant l'équation

$$\text{Sensibilité} = \text{Recall} = \frac{\text{True\_Positive}}{\text{True\_Positive} + \text{False\_Negative}}$$

F1-Score : Cela traduit l'équilibre entre la précision et le rappel, c'est la moyenne harmonique de Précision et Recall

$$\text{F1-Score} = 2 * \frac{\text{précision} * \text{recall}}{\text{Précision} + \text{recall}}$$

### 3.10. Comparaison des algorithmes :

Model	Précision	Recall	F1 -score
SVM	0,96	0,97	0.96
Naïve bayes	0,97	0,99	0.97
Les forêts aléatoires	0,96	0,97	0.96
Arbre de décision	0,94	0,95	0.94
KNN	0,96	0,96	0,94

Tableau 3.18 : les résultats des attributs d'évaluations pour différents modèles .

D'après le tableau ci-dessus le modèle Naïve Bayes qui obtenu les meilleurs résultats qui obtenu Précision qui égal 0.97 et Recall qui égal 0.99 et F1-score qui égal 0.97,

## Chapitre 3 Prédiction de cancer de sein par l'apprentissage automatique

---

D'autre part, nous remarquons que les deux modèles SVM et les forêts aléatoires qui obtenus les mêmes résultats qui obtenus Précisions qui égales 0.96 et le même Recall qui égales 0.97 et aussi le même F1-score qui égales 0.96.

De plus nous remarquons que le modèle KNN qui obtenu le Précision qui égal 0.96 et Recall qui égal 0.96 et le F1-score qui égal 0.94.

À la fin en remarquons que le modèle Arbre de décision obtenu le mauvaise résultats parmi les autres modèles, qui obtenu le Précision qui égal 0.94 et le Recall qui égal 0.95 et le F1-score qui égal 0.94.

C'est-à-dire que toutes les patients cancéreux

99 d'entre eux sont correctement classé à l'aide de mesure de diagnostiques médicales.

Nous choisirons le modèle « Naive bayes » comme le modèle le plus optimale et qui fonctionne mieux pour notre ensemble de données en raison de sa grande Précision et F1-score de Recall.

### 3.11.Conclusion :

Dans ce chapitre nous avons fait la description de dataset ,après nous allons appliquer cinq méthodes de performances ,ensuite nous comparons ces algorithmes entre eux .pour avoir un résultat les méthodes en affichant le rapport et la précision et f1 score de chaque modèle, nous permet de sélectionner le modèle "Naïve Bayes " comme le meilleur modèle qui a un il faut baser sur l'accuracy la plus adéquate.

### Conclusion générale

Le cancer de sein est l'une principales causes décès chez les femmes dans le monde ,en raison de croissance des cellules mammaires qui il ne peut pas être contrôlé.

Pour sa détection et sa , les médecins ont utilisé la mammographie pour établir un diagnostic sur leurs patientes. Cependant, la précision de la mammographie est moins impressionnante, de sorte que le besoin d'un meilleur facilitateur de prédiction est de plus en plus pressant.

De nombreux chercheurs ont utilisé les techniques d'apprentissage automatique et d'intelligence artificielle pour la prédiction et le diagnostic du cancer du sein.

Dans ce mémoire ,nous avons présenté des modèles d'apprentissage automatique pour la prédiction et diagnostic de cancer de sein sur l'ensemble des données (data set),nous avons mène à faire une comparaison entre cinq algorithmes d'apprentissage à savoir :support vector machine ,Random forest ,Naïve Bayes,k nearest neighbors,l'arbre de décision. Toutes ces techniques ML ont été par rapport à l'exactitude ,au rappel et à la précision .

L'efficacité de ce qui précède les algorithmes ont également été comparés à l'aide de Wisconsin. Nous avons constaté que la meilleure précision de Naïve Bayes et vectoriel support machine et était 97% et 96%,la meilleure précision de Naïve Bayes, et un rappel de 99%et 97%.

### Perspectives :

Les perspectives découlant de nos travaux sont nombreuses, les plus importantes sont les suivantes :

Notre travail futur consisterait à généraliser le modèle proposé à un plus large éventail de scénarios pratiques, tels que faciliter le diagnostic de plusieurs types de maladies. La quantité de données de vérité de base est généralement limitée en raison de la difficulté d'obtenir et d'annoter des données, et parce que les méthodes d'apprentissage automatique nécessitent de grandes quantités de données, nous envisageons également d'augmenter les données avec de nouvelles techniques.

## Conclusion Générale :

---

comment le jumeau numérique , une simulation de la santé d'un patient , pourrait conduire à la détection précoce du cancer à l'avenir grâce à la technologie de l'IA et à l'expertise médicale. L'intégration des données de santé permet des soins personnalisés avant même l'hospitalisation pour obtenir les meilleurs résultats possibles rapportés par les patients . Dans ce travail, un avenir possible de soins intelligents contre le cancer est démontré pour repousser les frontières de l'oncologie.

**Références**

**Bibliographies**

## Références

---

- [1] <https://www.who.int/fr/news-room/fact-sheets/detail/breast-cancer>
- [2] <http://www.santemaghreb.com/algerie/depcansein.htm#:~:text=Le%20taux%20d'incidence%20du,dix%20ans%20sont%20de%2050%25>.
- [3] [https://www.aps.dz/sante-science-technologie/63491-cancer-du-sein-l-oms-recommande-le-depistage-precoce-dans-les-pays-en-developpement#:~:text=En%20Alg%C3%A9rie%2C%20le%20cancer%20du,publique%20\(INSP\)%20de%202015](https://www.aps.dz/sante-science-technologie/63491-cancer-du-sein-l-oms-recommande-le-depistage-precoce-dans-les-pays-en-developpement#:~:text=En%20Alg%C3%A9rie%2C%20le%20cancer%20du,publique%20(INSP)%20de%202015).
- [4] <https://www.lillyoncologie.fr/cancer-du-sein/definitionfbclid=IwAR2zERY2zsK6OwWBF1IDm5H8CU-nlVYup0Zs4-eWNfy8BkDfHun44knJ2oI>.
- [5] [https://www.e-cancer.fr/Dictionnaire/T/tumeur-malignefbclid=IwAR16RTQQPSuVmy6woVV49IsZRTUvrvHk4HSM\\_-wqtP5UKvoHvqRdfpRwL0](https://www.e-cancer.fr/Dictionnaire/T/tumeur-malignefbclid=IwAR16RTQQPSuVmy6woVV49IsZRTUvrvHk4HSM_-wqtP5UKvoHvqRdfpRwL0).
- [6] [https://www.e-cancer.fr/Dictionnaire/T/tumeur-benignefbclid=IwAR0NwcoqhMBucndhkxwRiujLAIGLW1c\\_I4DIeXC4-N8DzkdZyNib-Yq5A](https://www.e-cancer.fr/Dictionnaire/T/tumeur-benignefbclid=IwAR0NwcoqhMBucndhkxwRiujLAIGLW1c_I4DIeXC4-N8DzkdZyNib-Yq5A).
- [7] [https://www.prostate.fr/cancer-de-la-prostate/mecanismes-origine/fbclid=IwAR1qU0UIuILrGmv-K4aP-TI-ichNM4pWnPvFNzW1sM0LQL5rWyniq49X\\_eQ](https://www.prostate.fr/cancer-de-la-prostate/mecanismes-origine/fbclid=IwAR1qU0UIuILrGmv-K4aP-TI-ichNM4pWnPvFNzW1sM0LQL5rWyniq49X_eQ).
- [8] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Anatomie-du-seinfbclid=IwAR375sNtOvsxfTBRtMx2TSPvKT0-jDXvBL5uiCzqIJdenlvcfudOeS-DBvo>.
- [9] <https://www.biron.com/fr/centre-du-savoir/petit-guide-biron/cancer-du-sein/?fbclid=IwAR1X-q4kMf-eRQBRvJiDOQR8WgggVoJWJxWbsy7ZdQIyRiC9mM1Azi22jl4> .
- [10] [https://www.arcagy.org/infocancer/localisations/cancers-feminins/cancer-du-sein/formes-de-la-maladie/la-stadification.html/?fbclid=IwAR1qU0UIuILrGmv-K4aP-TI-ichNM4pWnPvFNzW1sM0LQL5rWyniq49X\\_eQ](https://www.arcagy.org/infocancer/localisations/cancers-feminins/cancer-du-sein/formes-de-la-maladie/la-stadification.html/?fbclid=IwAR1qU0UIuILrGmv-K4aP-TI-ichNM4pWnPvFNzW1sM0LQL5rWyniq49X_eQ).

## Références

---

- [11] <https://dumas.ccsd.cnrs.fr/dumas-01109531/document?fbclid=IwAR1X-q4kMf-eRQBRvJiDOQr8WggBV0JWJxWbsy7ZdQIyRiC9mM1Azi22jl4>.
- [12] <https://www.europadonna.fr/le-cancer-du-sein/le-cancer-du-sein/differents-types-de-cancer/fbclid=IwAR17fMNoskVn51FzRfO-yX4zwNbptRzL9x-ygg6aHLqn5n0G7XLGfXIWPkE>.
- [13] [http://www.depistagesein.ca/types-de-cancer-du-sein/fbclid=IwAR0NwcoqhMBucndhkxwRiuJLAIGLW1c\\_I4DIeXC4-N8DzkdZyNib-Yq5A#.YqMX1nbP3ce](http://www.depistagesein.ca/types-de-cancer-du-sein/fbclid=IwAR0NwcoqhMBucndhkxwRiuJLAIGLW1c_I4DIeXC4-N8DzkdZyNib-Yq5A#.YqMX1nbP3ce).
- [14] <https://www.vidal.fr/maladies/cancers/cancer-sein/causes-prevention.html?fbclid=IwAR1jWd45-xaUxL0DARMZBNyQvvSYw30ZX0boJNaiNGHqnL07XPo-FdZiHEw>.
- [15] [https://www.fondation-arc.org/cancer/cancer-sein/facteurs-risque-cancerfbclid=IwAR2GSIgzYbSkSWNoaO3QYFziiuZ0maN\\_9RXWmu47tQVXhf5Z4mmlql0z-zk](https://www.fondation-arc.org/cancer/cancer-sein/facteurs-risque-cancerfbclid=IwAR2GSIgzYbSkSWNoaO3QYFziiuZ0maN_9RXWmu47tQVXhf5Z4mmlql0z-zk).
- [16] <https://m.facebook.com/SephireTN/photos/a.2074244229536872/2673618599599429/?type=3>.
- [17] [https://rubanrose.org/minformer/a-propos-du-cancer-du-sein/signes-et-symptomes-du-cancer-du-sein/fbclid=IwAR1EJ0rd4lnds5un9FnkLyAyEa5vPRzHmJfyeQ0LbG-weyImERx\\_vHzWBTs](https://rubanrose.org/minformer/a-propos-du-cancer-du-sein/signes-et-symptomes-du-cancer-du-sein/fbclid=IwAR1EJ0rd4lnds5un9FnkLyAyEa5vPRzHmJfyeQ0LbG-weyImERx_vHzWBTs).
- [18] [https://rubanrose.org/blogue/prenez-la-sante-de-vos-seins-en-mains/?fbclid=IwAR1SpBSUC1D9sKsOs5x3J9QCETHvrBx\\_Lu3MtOOuthNozMollG4HET\\_Ghqs](https://rubanrose.org/blogue/prenez-la-sante-de-vos-seins-en-mains/?fbclid=IwAR1SpBSUC1D9sKsOs5x3J9QCETHvrBx_Lu3MtOOuthNozMollG4HET_Ghqs).
- [19] [https://www.e-cancer.fr/Comprendre-prevenir-depister/Se-faire-depister/Depistage-du-cancer-du-sein?fbclid=IwAR11Zb\\_JdtKRLT9TuEgaOtGI48FUj8s1cvqIZmLEQ\\_K8vyTMAvC27IHhpTk](https://www.e-cancer.fr/Comprendre-prevenir-depister/Se-faire-depister/Depistage-du-cancer-du-sein?fbclid=IwAR11Zb_JdtKRLT9TuEgaOtGI48FUj8s1cvqIZmLEQ_K8vyTMAvC27IHhpTk).
- [20] [https://mon-cancer.com/articles-blog/lautopalpation-des-seins-parlons-en/fbclid=IwAR11Zb\\_JdtKRLT9TuEgaOtGI48FUj8s1cvqIZmLEQ\\_K8vyTMAvC27IHhpTk](https://mon-cancer.com/articles-blog/lautopalpation-des-seins-parlons-en/fbclid=IwAR11Zb_JdtKRLT9TuEgaOtGI48FUj8s1cvqIZmLEQ_K8vyTMAvC27IHhpTk)



## Références

---

- [30] <https://datascience.eu/fr/apprentissage-automatique/quest-ce-que-lapprentissage-automatique-une-definition/>.
- [31] <https://datascience.eu/fr/apprentissage-automatique/top-6-des-techniques-dapprentissage-machine/>.
- [32] <https://course.elementsofai.com/fr/4/1>.
- [33] [https://projeduc.github.io/intro\\_apprentissage\\_automatique/introduction.html](https://projeduc.github.io/intro_apprentissage_automatique/introduction.html).
- [34] [https://fr.m.wikipedia.org/wiki/Apprentissage\\_supervis%C3%A9](https://fr.m.wikipedia.org/wiki/Apprentissage_supervis%C3%A9).
- [35] <https://datascientest.com/apprentissage-supervise>.
- [36] [https://fr.m.wikipedia.org/wiki/Apprentissage\\_non\\_supervis%C3%A9](https://fr.m.wikipedia.org/wiki/Apprentissage_non_supervis%C3%A9).
- [37] [https://people.minesparis.psl.eu/fabien.moutarde/ES\\_MachineLearning/Slides/coursFM\\_UNSUPERVISED-learning\\_2pp.pdf](https://people.minesparis.psl.eu/fabien.moutarde/ES_MachineLearning/Slides/coursFM_UNSUPERVISED-learning_2pp.pdf).
- [38] [https://analyticsinsights-io.cdn.ampproject.org/v/s/analyticsinsights.io/apprentissage-supervise-vs-non-supervise/amp/?amp\\_gsa=1&amp\\_js\\_v=a9&usqp=mq331AQKKAFQArABIICAw%3D%3D#amp\\_tf=Source%2%A0%3A%20%251%24s&aoh=16548795301167&referrer=https%3A%2F%2Fwww.google.com&ampshare=https%3A%2F%2Fanalyticsinsights.io%2Fapprentissage-supervise-vs-non-supervise%2F](https://analyticsinsights-io.cdn.ampproject.org/v/s/analyticsinsights.io/apprentissage-supervise-vs-non-supervise/amp/?amp_gsa=1&amp_js_v=a9&usqp=mq331AQKKAFQArABIICAw%3D%3D#amp_tf=Source%2%A0%3A%20%251%24s&aoh=16548795301167&referrer=https%3A%2F%2Fwww.google.com&ampshare=https%3A%2F%2Fanalyticsinsights.io%2Fapprentissage-supervise-vs-non-supervise%2F).
- [39] [https://www-lemagit-fr.cdn.ampproject.org/v/s/www.lemagit.fr/conseil/Machine-Learning-les-9-types-dalgorithmes-les-plus-pertinents-en-entreprise?amp=1&amp\\_gsa=1&amp\\_js\\_v=a9&usqp=mq331AQKKAFQArABIICAw%3D%3D#amp\\_tf=Source%2%A0%3A%20%251%24s&aoh=16548795829833&referrer=https%3A%2F%2Fwww.google.com&ampshare=https%3A%2F%2Fwww.lemagit.fr%2Fconseil%2FMachine-Learning-les-9-types-dalgorithmes-les-plus-pertinents-en-entreprise](https://www-lemagit-fr.cdn.ampproject.org/v/s/www.lemagit.fr/conseil/Machine-Learning-les-9-types-dalgorithmes-les-plus-pertinents-en-entreprise?amp=1&amp_gsa=1&amp_js_v=a9&usqp=mq331AQKKAFQArABIICAw%3D%3D#amp_tf=Source%2%A0%3A%20%251%24s&aoh=16548795829833&referrer=https%3A%2F%2Fwww.google.com&ampshare=https%3A%2F%2Fwww.lemagit.fr%2Fconseil%2FMachine-Learning-les-9-types-dalgorithmes-les-plus-pertinents-en-entreprise).
- [40] <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501879-machine-a-vecteurs-de-support-svm-definition-et-cas-d-usage/>.
- [41] [https://www.researchgate.net/figure/Classification-of-data-by-support-vector-machine-SVM\\_fig8\\_304611323](https://www.researchgate.net/figure/Classification-of-data-by-support-vector-machine-SVM_fig8_304611323).

## Références

---

- [42] <https://www.faepa.br/kgyo.aspx?cname=python+svm+image+classification&cid=66>.
- [43] <https://databusiness-ai.com/support-vector-machines-svm/>.
- [44] <https://penseeartificielle.fr/tout-pour-bien-debuter-en-machine-learning-4/hyperplan-svm/>.
- [45] <https://blent.ai/modele-machine-learning-populaires/>.
- [46] <https://data.scientest.com/knn>.
- [47] <https://medium.com/@kenzaharifi/bien-comprendre-lalgorithme-des-k-plus-proches-voisins-fonctionnement-et-impl%C3%A9mentation-sur-r-et-a66d2d372679>.
- [48] [https://www.imo.universite-parisaclay.fr/~goude/Materials/ProjetMLF/rf\\_web.html](https://www.imo.universite-parisaclay.fr/~goude/Materials/ProjetMLF/rf_web.html).
- [49] <https://datapeaker.com/fr/Big-Data/for%C3%AAts-al%C3%A9atoires-dans-1%27apprentissage-automatique/>.
- [50] <https://datascientest.com/random-forest-definition>.
- [51] [https://fr.m.wikipedia.org/wiki/Arbre\\_de\\_d%C3%A9cision](https://fr.m.wikipedia.org/wiki/Arbre_de_d%C3%A9cision).
- [52] <https://scaron.info/doc/intro-arbres-decision/>.
- [53] [https://fr.m.wikipedia.org/wiki/Classification\\_na%C3%AFve\\_bay%C3%A9sienne](https://fr.m.wikipedia.org/wiki/Classification_na%C3%AFve_bay%C3%A9sienne).
- [54] [https://fr.slideshare.net/ALAIN\\_TASSY/application-de-lintelligence-artificielle-la-prdiction-et-au-diagnostic-en-cancrologie](https://fr.slideshare.net/ALAIN_TASSY/application-de-lintelligence-artificielle-la-prdiction-et-au-diagnostic-en-cancrologie).
- [55] <https://geekflare.com/fr/understanding-al-ml-dl/>.