République Algérienne Démocratique 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 : Mathématiques et Informatique

Mémoire de fin d'étude

Pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Option : Système information

Thème

IA pour la santé : cas d'étude « Prédiction de la durée de séjour hospitalier »

Présenté Par : Gadri Dhouha

Devant le jury

Dr. METROUH Abdelmalek	*MCA*	*Université Larbi Tébessi*	*Président*
Dr. SLIMI Hamda	*MAB*	*Université Larbi Tébessi*	*Examinateur*
Pr. LAOUAR Mohamed Ridda	*Pr*	*Université Larbi Tébessi*	*Encadreur*
Mlle TOUATI HAMAD Zaineb		*Université Larbi Tébessi*	*Co-Encadreur*

Date de soutenance : 08/ 06 / 2023


بسم الله الرحمن الرحيم

Sommaire

Liste des figures
Liste des Tableaux
Liste des Abréviations
Remerciement
Dédicace
Résumé
1. Introduction Générale I
2. Problématique III
3. Objectif de l'étude III
4. Choix & Intérêt du sujet IV
5.Organisation du mémoire V

Chapitre I : Généralités sur les notions d'hospitalisation

1. Introduction 6
2. Santé d'après Organisation mondiale de la santé 6
   2.1 Prévention 6
   2.2. Promotion et protection de la santé 7
   2.3. Éducation pour la santé 7
3. Hôpital 8
4. Durée de séjour hospitalier 8
   4.1 Importance de la DDSH 9
   4.2 Management des lits 9
   4.3 Facteurs qui influencent la DDSH 10
   4.5 Avantages et inconvénients de la DDSH 10
4. Programme de Médicalisation des Systèmes d'Information (PMSI) 11
   4.1. Définition du PMSI 11
   4.2. Utilisations du PMSI 12
   4.3. PMSI en Algérie 12
   4.4. Objectif du PMSI 13
   4.5. Types du PMSI 13
5. Systèmes d'Information Hospitaliers « SIH » 14
   5.1. Système Information 14
   5.2. Définition de SIH 14
   5.3. Sous-système d'information hospitalier 15

5.4 Définition des informations administratives 16
5.5. Professionnels de santé 17
6. Dossier du patient (DP) 17
6.1. Dossier médical du patient (DMP) 18
6.2. Rôle du dossier du patient 18
6.3. Contenu de DMP 19
6.4. Dossier patient électronique (DPE) 19
6.5. Suggestions pour la Transition numérique médicale 20
7. Donnée de santé 20
7.1. Catégories de données de santé 21
7.2. Propriétés des données médicales 22
7.3. Sources des données médicales des systèmes information hospitalier (SIH) 23
8. Conclusion 24

Chapitre II : Intelligence Artificiel et la prédiction de la durée de séjour hospitalier

1. Introduction 26
2. Intelligence Artificielle 26
2.1. Techniques de l'IA 27
3. Apprentissage automatique (Machine Learning) 27
3.1. Définition 27
3.2. Domaine d'application de l'apprentissage automatique 27
3.3 Types de l'apprentissage automatique 28
4. Apprentissage supervisé (Supervised Learning) 29
4.1. Fonctionnement de l'apprentissage supervisé 29
4.2. Types Apprentissage supervisé 30
4.3. Algorithme d'apprentissage supervisé les plus couramment utilisés 33
4.4. Comparaison d'algorithme de l'apprentissage automatique 36
5. Apprentissage non supervisé 38
5.1. Utilisation d'apprentissage non supervisé 38
5.2. Types Apprentissage non supervisé 38
5.3. Algorithmes d'apprentissage non supervisé 41
5.4. Comparaison d'algorithme de l'apprentissage non supervisé 44
6. Différence entre l'apprentissage supervisé et l'apprentissage non supervisé 44
7. Apprentissage automatique et la prédiction de la DDSH 46
8. Conclusion 47

Chapitre III : Etat de l'art de la prédiction de la DDSH

1. Introduction 49
2. Prédiction de la DDSH 49
3.Méthodes de prédiction de la DDSH 49
3.1. Méthodes classiques pour la prédiction de la DDSH 49
3.2. Méthodes intelligentes pour la prédiction de la DDSH 50
4. Etapes de modélisation de prédiction de la DDSH basées sur machine Learning 51
5. DataSets utilisés dans la prédiction de la durée de séjour hospitalier 52
6. Travaux connexes de la prédiction de la DDSH 55
7. Défis 61
8. Synthèse 61
9. Conclusion 62

Chapitre IV : Réalisation d'un modèle de prédiction de la DDSH

1. Introduction 64
2. Matériels utilisés 64
3. Outils de développement 64
3.1 L angages utilisés 64
3.2. Produit de Google Recherche 65
4.Choix et fonctionnement de l'algorithme 65
5. DataSet utilisé dans notre travail 66
5.1 Description de DataSet 66
6.Étendue de l'étude et Prédiction de la DDSH 68
7. Architecture des étapes de modélisation du la prédiction de la durée de séjour hospitalier 69
8.Étapes de modélisation de la prédiction de la DDSH 69
8.1 Collecte des données 69
8.2. Analyse et nettoyage des données 70
8.3 Prétraitement des données 72
8.4 Modélisation 74
8.5 Évaluation du modèle 74
9. Discussion et comparaison 76
9.1 Comparaison les résultat avec des travaux utilisent même DataSet avec notre travail…76
9.2 Comparaison avec des autres travaux qui utilisent des différentes DataSets 77
10. Conclusion 79
Conclusion générale et perspectives 81
Bibliographie 84

Liste des figures

Figure I.1. OMS……. 6
Figure I.2.Hopital…… 8
Figure I.3.Sejour hospitalier…… 9
Figure I.4.Systeme information hospitalier…… 16
Figure I.5.Dossier patient ……. 17
Figure I.6.Dossier patient électronique……. 20
Figure I.7.Donnees de sante…… 24
Figure II.1. Intelligence artificiel…… 26
Figure II.2. Techniques de IA…… 27
Figure II.3. Application Machine Learning…… 28
Figure II.4. Type ML……. 28
Figure II.5. Apprentissage supervise …… 29
Figure II.6. Matrice de confusion ……. 31
Figure II.7. Arbre décision …… 33
Figure II.8. Foret aléatoire …… 34
Figure II.9. Réseaux de neurones ……. 35
Figure II.10.SVM point de données…… 35
Figure II.11. Apprentissage non supervise ……. 38
Figure II.12. Clustering…….. 39
Figure II.13. Clustering hiérarchique…….. 40
Figure II.14. Clustering base sur la densité……. 40
Figure II.15. Clustering SVC ……. 41
Figure II.16. Algorithme K moyens…… 42
Figure II.17. Algorithme DBSCAN…… 42
Figure IV.1. Python……. 65
Figure IV.2. Architecture des étapes de la prédiction de la durée de séjour hospitalier 69
Figure IV.3. Durée de séjour et sexe……. 70
Figure IV.4. Durée de séjour et réadmission……. 70
Figure IV.5. Durée de séjour et fer à repasser…… 71
Figure IV.6. Durée de séjour et dépression……. 71
Figure IV.7. Corrélation entre les données de DataSet…….. 72
Figure IV.8. Histogramme Test Train…… 74

Figure IV.9. Performance des modèles…….. 74
Figure IV.10. Comparaison RMSE des modèles…… 75
Figure IV.11. Comparaison R2 Score des modèles…… 76

Liste des Tableaux

Tableau I.1	Avantages et inconvénients de la DDSH……..	11
Tableau I.2	Utilisation de PMSI……..	12
Tableau II.1	Comparaison algorithmes d'apprentissage supervisé……...	36
Tableau II.2	Comparaison algorithmes d'apprentissage non-supervisé……	44
Tableau II.3	Différence entre l'apprentissage supervisé et non supervisé……	45
Tableau III.1	Comparaison entre les DataSets (jeux de données) …….	53
Tableau III.2	Travaux connexes de la prédiction de la DDSH……...	57
Tableau IV.1	Fonctionnements des méthodes ……	66
Tableau IV.2	Description de DataSet …….	67
Tableau IV.3	Résultat de performance des méthodes utilisées……...	75
Tableau IV.4	Comparaison les résultats des travaux qui utilisent le même DataSet.....	77
Tableau IV.5	Comparaison résultats des travaux utilisent différentes DataSets…….	78

Liste des Abréviations

AD	Arbre décision
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DDS	Durée de séjour
DDSH	Durée de séjour hospitalier
DMP	Dossier médicale de patient
DP	Dossier patient
DPE	Dossier patient électronique
FA	Foret aléatoire
IA	Intelligence Artificiel
KNN	k-Nearest Neighbors (k-Plus Proches Voisins)
ML	Machine Learning Apprentissage automatique
MLRP	Multivariate Linear Regression with Panel Data
OMS	Organisation mondiale de la santé
PMSI	Programme de médicalisation système information
RB	Réseaux de Bayes
RG	Réalité Générative
RL	Régression linéaire
RMSE	Root Mean Square Error
RNA	Réseaux de neurones artificiel
SIH	Système information hospitalier
SVC	Support Vector Classifier
SVM	Machines à Vecteurs de Support
SVR	Support Vector Régression
XGBoost	eXtreme Gradient Boosting

REMERCIEMENT

*Au terme de ce travail, nous aimerions exprimer notre gratitude au **Dieu**. Nous sommes profondément reconnaissants de nous avoir accordé le courage, la volonté et la patience nécessaires pour mener à bien ce travail.*

*Nous tenons également à exprimer notre immense reconnaissance envers notre encadrant **LAOUAR Mohamed Ridda** et co-encadrent **TOUATI HAMAD Zaineb** pour son soutien inestimable, ses conseils avisés et sa présence constante. Votre expertise et votre dévouement ont été des sources d'inspiration pour nous.*

*Nous également remercier chaleureusement **les membres du jury « SLIMI Hamda et METROUH Abdelmalek »** , pour leur temps, leur expertise et leur évaluation minutieuse de notre travail.*

Enfin, nous exprimons notre gratitude envers toutes les personnes qui ont contribué de près ou de loin à l'élaboration de ce travail. Votre soutien, vos connaissances partagées, votre assistance technique et vos encouragements ont été essentiels à notre réussite.

DÉDICACE

À ma chère maman,

Je te suis extrêmement reconnaissante pour ton amour inconditionnel, ta bienveillance et tes sacrifices. Tu as été mon premier amour et ma première inspiration. Merci d'avoir toujours été là pour moi, de m'avoir soutenue et encouragée dans tous mes projets. Ce travail est dédié à toi, en témoignage de l'amour infini que j'ai pour toi. Ta lumière éclaire ma vie et ton soutien constant me guide. Je te suis profondément reconnaissante pour tout ce que tu fais. Je t'aime infiniment.

À mon cher père,

Ton soutien inébranlable et tes conseils avisés ont été une inspiration pour moi. Tu m'as montré la voie de l'intégrité, de la persévérance et de la détermination. Cette réalisation est dédiée à toi, en reconnaissance de tout ce que tu as fait pour moi.

À mes merveilleuses sœurs Isra et Baraa,

Vous êtes mes amies les plus proches, mes confidentes et mes alliées. Votre présence joyeuse et votre soutien indéfectible ont rendu ce voyage plus significatif. Cette réussite est dédiée à vous, en témoignage de notre lien familial et de notre amour inconditionnel.

À mon cher frère Mouafek,

À mes amis précieux,

Que cette dédicace témoigne de ma reconnaissance éternelle envers ma famille et mes amis qui ont été présents à chaque étape de ma vie, m'encourageant à atteindre de nouveaux sommets.

Avec amour et gratitude,

Dhouha

Résumé :

La prédiction de la durée de séjour hospitalisé est essentielle pour optimiser les soins, planifier les ressources et réduire les coûts. Ce projet se concentre sur l'utilisation de l'intelligence artificielle, en particulier l'apprentissage automatique (machine learning), pour cette prédiction. Nous commencerons par explorer les concepts d'apprentissage automatique, y compris les notions d'apprentissage supervisé et non supervisé. Ensuite, nous étudierons les méthodes classiques utilisées, en mettant en évidence leurs avantages et leurs limites. Par la suite, nous passerons en revue les méthodes intelligentes spécifiques telles que les réseaux de neurones, les arbres de décision et les méthodes de classification, en discutant de leurs avantages, de leurs limites et de leur pertinence dans le contexte de la prédiction de la durée de séjour hospitalisé. L'objectif principal est d'aider les professionnels de la santé à choisir la méthode la plus appropriée en fonction de leurs besoins et contraintes. Grâce à des modèles adaptés développés avec l'apprentissage automatique, des prédictions précises ont été réalisées, facilitant la collecte d'informations sur les patients et l'évaluation des durées d'hospitalisation et de traitement.

Les meilleurs résultats de la prédiction de la durée de séjour hospitalier ont été obtenus avec l'algorithme de forêt aléatoire, avec un RMSE (Root Mean Square Error) de 0.24 et un coefficient de détermination (R^2) de 0.94. Ces performances élevées démontrent l'efficacité de l'algorithme de forêt aléatoire dans la prédiction précise de la durée de séjour hospitalier.

Mots clés : **Prédiction de la durée de séjour hospitalier, Intelligence Artificielle, Apprentissage supervisé, Méthodes intelligentes.**

Résumé

Abstract:

The prediction of hospital length of stay is crucial for optimizing care, resource planning, and cost reduction. This project focuses on the use of artificial intelligence, particularly machine learning, for this prediction. We will begin by exploring the concepts of machine learning, including supervised and unsupervised learning. Next, we will review the conventional methods used, highlighting their advantages and limitations. Subsequently, we will examine specific intelligent methods such as neural networks, decision trees, and classification methods, discussing their strengths, limitations, and relevance in the context of hospital length of stay prediction. The main objective is to assist healthcare professionals in selecting the most appropriate method based on their needs and constraints. Through tailored models developed with machine learning, accurate predictions have been achieved, facilitating the collection of patient information and the assessment of hospitalization and treatment durations.

The best results for predicting the duration of hospital stay were achieved with the Random Forest algorithm, yielding a Root Mean Square Error (RMSE) of 0.24 and a coefficient of determination (R^2) of 0.94. These high performances demonstrate the effectiveness of the Random Forest algorithm in accurately predicting the duration of hospital stay.

Keywords: **Hospital length of stay prediction, Artificial Intelligence, supervised learning, Intelligent methods.**

Résumé

ملخص:

إن توقع مدة الإقامة في المستشفى ضروري لتحسين الرعاية الصحية وتخطيط الموارد وتقليل التكاليف. يركز هذا المشروع على استخدام الذكاء الاصطناعي، وبالتحديد التعلم الآلي. لهذا التوقع سنبدأ باستكشاف مفاهيم التعلم الآلي، بما في ذلك مفهومي التعلم المشرف وغير المشرف. ثم سنستعرض الطرق التقليدية المستخدمة، مسلطين الضوء على مزاياها وقيودها. بعد ذلك، سنستعرض الطرق الذكية المحددة مثل الشبكات العصبية وأشجار القرار وطرق التصنيف، ونناقش مزاياها وقيودها وملاءمتها في سياق توقع مدة الإقامة في المستشفى. الهدف الرئيسي هو مساعدة العاملين في مجال الرعاية الصحية على اختيار الطريقة الأكثر ملاءمة وفقًا لاحتياجاتهم وقيودهم. من خلال النماذج المعدة بشكل جيد التي تم تطويرها باستخدام التعلم الآلي، تم تحقيق توقعات دقيقة، مما يسهل جمع معلومات المرضى وتقييم مدة الإقامة والعلاج.

تم الحصول على أفضل النتائج لتوقع مدة الإقامة في المستشفى باستخدام خوارزمية الغابات العشوائية حيث أعطت قيمة خطأ المربع المتوسطة (RMSE) بمقدار 0.24 ومعامل التحديد (R^2) بمقدار 0.94. تلك النتائج العالية تُظهر فعالية خوارزمية الغابات العشوائية في التنبؤ الدقيق لمدة الإقامة في المستشفى.

الكلمات المفتاحية: توقع مدة الإقامة في المستشفى، الذكاء الاصطناعي، التعلم تحت الإشراف، الطرق الذكية.

INTRODUCTION GÉNÉRALE

1. Introduction Générale

La numérisation de l'information médicale est un processus essentiel dans la modernisation des soins de santé. Il s'agit de convertir les dossiers médicaux des patients en fichiers électroniques, afin de faciliter leur stockage, leur partage et leur utilisation dans différents établissements de santé.

L'objectif principal de la numérisation des informations médicales est d'améliorer la qualité des soins de santé et de réduire les coûts associés à la gestion des dossiers médicaux papier. Cette technologie offre de nombreux avantages, tels qu'un accès instantané aux informations médicales des patients, une réduction des erreurs médicales et la possibilité de partager facilement des informations entre différents prestataires de soins de santé.

Ces dernières années, l'IA a été largement appliquée dans de nombreux domaines, y compris la santé. L'un des domaines est d'application de l'IA dans la santé est la prédiction de la durée de séjour hospitalisé.

L'intelligence artificielle (IA) est une technologie de pointe qui permet à des machines d'apprendre, de raisonner et de prendre des décisions autonomes. C'est une branche de l'informatique qui vise à développer des systèmes informatiques capables de faire des tâches qui nécessitent normalement l'intelligence humaine, comme la reconnaissance de la parole, la traduction de langues, la reconnaissance d'images et la prédiction.

La prédiction de la durée de séjour hospitalisé est un enjeu important pour les professionnels de la santé, les patients et les établissements de santé. Elle permet de planifier les soins et les ressources nécessaires, de mieux anticiper les coûts et d'améliorer la qualité des soins.

La prédiction de la durée de séjour hospitalier à l'aide de l'IA montre une croissance importante de la recherche dans ce domaine. Les méthodes de prédiction traditionnelles ont été remplacées par des techniques plus avancées basées sur l'apprentissage automatique et l'IA.

Les méthodes traditionnelles de prédiction de la durée de séjour hospitalier comprennent des modèles statistiques simples, tels que la régression linéaire et les analyses de survie. Ces modèles ont été largement utilisés dans le passé, mais leur précision est limitée car ils ne prennent pas en compte la complexité et la variabilité des patients.

La prédiction de séjour hospitalisé avec intelligence artificielle est un domaine en développement dans lequel des algorithmes d'IA sont utilisés pour prédire la durée probable de l'hospitalisation d'un patient en fonction de divers facteurs tels que l'état de santé du patient, les antécédents médicaux, les diagnostics, les traitements et les interventions médicales, son âge, son sexe, et sa maladie.

L'objectif de cette approche est d'aider les professionnels de la santé à prendre des décisions plus informées en ce qui concerne la planification des soins pour les patients hospitalisés et de maximiser l'efficacité des ressources hospitalières en prédisant la durée probable de l'hospitalisation.

L'utilisation de l'intelligence artificielle apporte une réelle valeur ajoutée dans ce domaine, les méthodes traditionnelles de prédiction de la durée du séjour à l'hôpital reposent sur des modèles statistiques qui peuvent être limités par la complexité des données de santé. En revanche, l'IA peut analyser de grandes quantités de données en temps réel et détecter des modèles complexes, ce qui peut améliorer la précision des prédictions.

En fin de compte, l'application de l'IA à la prédiction de la durée de séjour hospitalisé peut améliorer l'efficacité des soins de santé, réduire les coûts et améliorer l'expérience des patients. Cependant, il est important de garantir que les modèles d'IA sont transparents, équitables et respectent la confidentialité des patients pour assurer une utilisation éthique de cette technologie.

En résumé, l'état de l'art de la prédiction de la durée de séjour hospitalier montre que l'IA et l'apprentissage automatique offrent des méthodes plus précises et plus avancées pour prédire la durée de séjour des patients.

2. Problématique

Les services hospitaliers ont essayé d'optimiser leurs ressources pour améliorer leurs performances. L'un des indicateurs les plus importants de l'efficacité des systèmes de santé est la durée du séjour à l'hôpital, qui correspond à l'intervalle de temps entre l'admission et la sortie du patient. Afin de se débarrasser de la charge supplémentaire sur le travail des professionnelles en soins, notamment face aux besoins de santé croissants de la population.

- Comment l'utilisation de l'intelligence artificielle peut-elle améliorer la précision de la prédiction de la durée de séjour hospitalier**?**
- Quels sont les avantages de l'utilisation de l'intelligence artificielle dans ce domaine, par rapport aux méthodes de prédiction classiques **?**
- Quel est la meilleure méthode utilisée pour la prédiction de la durée de séjour hospitalisé à l'aide de l'intelligence artificielle **?**

En répondant à ces questions, la problématique permettra de mieux comprendre l'utilisation de l'intelligence artificielle dans la prédiction de la durée de séjour hospitalisé, et de déterminer les conditions dans lesquelles cette technologie peut apporter une réelle valeur ajoutée à la prise en charge des patients.

3. Objectif de l'étude

La prédiction de la durée de séjour hospitalier est une tâche complexe en raison de la variabilité des patients, des maladies et des procédures médicales. La plupart des hôpitaux utilisent des méthodes traditionnelles pour prédire la durée de séjour des patients, qui ne sont pas toujours précises.

- Développer un modèle d'IA pour prédire la durée de séjour hospitalier de manière plus précise et plus efficace.
- Comparaison les méthodes classiques de prédiction de la durée de séjour à l'hôpital avec des méthodes intelligentes utilisant l'intelligence artificielle (IA) pour déterminer l'impact de l'IA sur la précision des prédictions.
- Identifier les ensembles de données les plus appropriés pour prédire la durée du séjour à l'hôpital en utilisant l'IA.
- Choisir la méthode intelligente la plus appropriée pour cette étude.
- Examinez les avantages et les inconvénients de l'utilisation de différents types d'algorithmes d'IA (réseaux de neurones, arbres de décision, etc.) pour prédire la durée du séjour à l'hôpital.

4. Choix & Intérêt du sujet

4.1. Intérêt de travail

Le développement d'algorithmes d'intelligence artificielle pour la prédiction de la durée de séjour hospitalisé est un enjeu important pour les professionnels de santé et les gestionnaires hospitaliers. En effet, une meilleure prédiction de la durée de séjour permettrait d'optimiser la gestion des ressources et des lits d'hospitalisation, ainsi que de réduire les coûts associés à une hospitalisation prolongée.

Les résultats de la prédiction de la durée de séjour hospitalisé peuvent également aider les professionnels de santé à planifier la sortie des patients et à organiser leur suivi post-hospitalisation, ce qui contribue à améliorer la qualité des soins et la satisfaction des patients.

4.2. Intérêt scientifique

La prédiction de la durée de séjour hospitalisé est un domaine de recherche important en santé publique, qui peut bénéficier de l'utilisation de l'intelligence artificielle. Les méthodes classiques de prédiction, basées sur des modèles statistiques, peuvent être limitées par la complexité des données de santé et la variabilité des facteurs de risque. L'intelligence artificielle, en revanche, permet d'analyser de grandes quantités de données en temps réel et de détecter des modèles complexes, ce qui peut améliorer la précision de la prédiction.

L'intelligence artificielle peut également contribuer à la recherche sur les facteurs de risque de prolongation de la durée de séjour hospitalisé, en identifiant des variables prédictives qui n'ont pas été considérées dans les modèles classiques.

4.3. Intérêt académique

La prédiction de la durée de séjour hospitalisé est un sujet d'intérêt pour les étudiants et chercheurs en santé publique, en médecine, en informatique et en statistiques. Ce sujet permet de développer des compétences en analyse de données, en modélisation statistique et en programmation informatique.

L'utilisation de l'intelligence artificielle dans la prédiction de la durée de séjour hospitalisé soulève également des questions éthiques et juridiques, qui peuvent intéresser les étudiants et chercheurs en droit, en éthique médicale et en sciences sociales.

5.Organisation du mémoire

Le mémoire proposé est organisé en 4 chapitres :

Chapitre 1 :

- Définition de quelques concepts que nous avons jugé nécessaire sur la prédiction de la durée de séjour hospitalisé à l'aide de l'intelligence artificielle.
- Citer les facteurs qui influencent de la DDSH.
- Avantages et des inconvénients de la DDSH.

Chapitre 2 :

- Introduction à l'intelligence artificielle et à ses techniques.
- Apprentissage automatique et ses deux types : Apprentissage supervisé et non supervisé et ses algorithmes
- Différences entre l'apprentissage supervisé et l'apprentissage non supervisé
- Apprentissage supervisé et la prédiction de la DDSH

Chapitre 3 :

- Méthodes de prédiction du la DDSH (classiques et intelligentes)
- DataSets utilisées dans les travaux connexes de la prédiction de la DDSH

Chapitre 4 :

- Matériels et outils de développement utilisées
- Description de DataSet utilisée
- Choix de l'algorithme
- Architecture de contribution
- Etapes de prédiction du la DDSH

CHAPITRE I

Généralités sur les notions d'hospitalisation

1. Introduction :

La durée de séjour hospitalier est un indicateur clé dans l'évaluation de la qualité des soins de santé et de l'efficacité des établissements de santé. Une durée de séjour optimale peut contribuer à améliorer la qualité des soins, à réduire les coûts et à optimiser l'utilisation des ressources. Cependant, la gestion de la durée de séjour est un défi complexe pour les établissements de santé, car elle dépend de nombreux facteurs, tels que la gravité de la maladie, les comorbidités, les protocoles de traitement et les ressources disponibles. Dans ce contexte, il est essentiel de surveiller et d'analyser la durée de séjour hospitalier afin d'identifier les domaines qui nécessitent des améliorations et de développer des stratégies pour optimiser la qualité des soins et des services offerts aux patients.

2. Santé d'après Organisation mondiale de la santé :

Le bien-être global ne se limite pas à l'absence de maladie ou de handicap, mais englobe un état de santé physique, mental et social optimal. [1]

Figure I. 1. OMS [2]

Trois grands concepts de la définition de la santé : la prévention, de la promotion (La protection) et de l'éducation pour la santé, voici leurs définitions comme suit :

2.1 Prévention :

Selon l'OMS en 1948 : « l'ensemble des mesures visant à éviter ou réduire le nombre et la gravité des maladies, des accidents et des handicaps ».

- ✓ La prévention est l'ensemble des actions visant à :
  - Réduire l'impact des déterminants des maladies ;
  - Et/ou à éviter la survenue des maladies ;
  - Arrêter leur propagation et/ou à limiter leurs conséquences. [3]

2.2. Promotion et protection de la santé

Elle comprend cinq principaux domaines d'intervention, à savoir l'élaboration de politiques en faveur de la santé, la création d'environnements propices, le renforcement de l'engagement communautaire, le développement de compétences individuelles et la réorientation des services de santé. « Donner aux individus davantage de maîtrise de leur propre santé et davantage de moyens de l'améliorer ». [4]

La protection de la santé : Il s'agit d'un ensemble de mesures qui englobent des domaines tels que la santé, l'économie, le social, l'éducation et l'écologie, dans le but de réduire ou d'éliminer les risques sanitaires. Ces risques peuvent provenir de facteurs héréditaires, de l'alimentation, du comportement humain ou de l'environnement. L'objectif de ces mesures est de préserver la santé tant au niveau individuel que collectif. [5]

2.3. Éducation pour la santé

L'éducation en matière de santé est un moyen de prévention qui vise à réduire les décès prématurés causés par des maladies ou des accidents, en grande partie liés aux comportements et aux modes de vie. Les actions d'éducation en matière de santé contribuent à **:** [6]

- Améliorer le bien-être individuel et collectif ;
- Intégrer la santé dans le développement local ;
- Impliquer les individus dans les décisions concernant leur santé.

L'État, les collectivités locales et les institutions éducatives collaborent afin de fournir une éducation en santé qui contribue au bien-être de la population. L'objectif est de permettre à chacun d'acquérir les connaissances nécessaires pour prendre soin de sa santé. Cela inclut des sujets tels que la prévention des maladies, l'adoption de modes de vie sains et la compréhension des risques sanitaires. L'éducation en santé vise à autonomiser les individus en les informant et en les encourageant à prendre des décisions éclairées pour leur bien-être notamment en matière : [7]

- Hygiène individuelle et collective ;
- Nutrition saine et équilibrée ;
- Promotion de la santé bucco-dentaire ;
- Prévention en matière de santé mentale ;
- Consommation des médicaments…etc.

3. Hôpital

C'est une institution qui est construite, dotée en personnel et équipée pour le diagnostic de la maladie ; pour le traitement médical et chirurgical des malades et des blessés ; et pour leur logement pendant ce processus. [8]

Les hôpitaux complètent et amplifient l'efficacité de nombreuses autres parties du système de santé, offrant une disponibilité continue des services pour les affections aiguës et complexes. Ils concentrent les ressources rares dans des réseaux de référence bien planifiés pour répondre efficacement aux besoins de santé de la population. Ils sont un élément essentiel de la Couverture Santé Universelle (CSU). [9]

Figure I. 2 : Hôpital [10]

4. Durée de séjour hospitalier

Lorsqu'un individu est admis à l'hôpital, que ce soit dans un établissement public ou une clinique privée, la durée de sa présence à l'hôpital est désignée par le terme "séjour". Pendant cette période, le patient peut être pris en charge dans différents services, qui sont actuellement appelés des "unités médicales". Pendant le séjour, le patient peut être examinée, diagnostiquée, traitée et surveillée par des professionnels de la santé. Le séjour comporte une date d'entrée et une date de sortie, qui définissent la durée du séjour. [11]

DDS est un aspect important à considérer dans la planification et la gestion des ressources hospitalières, nous en parlerons : C'est la période pendant laquelle un patient reste hospitalisé dans un établissement de santé pour recevoir des soins médicaux ou chirurgicaux. Cette période commence à la date d'admission du patient dans l'établissement et se termine à la date de sa sortie, qu'elle soit planifiée ou non. [13]

Cette durée peut varier considérablement en fonction de nombreux facteurs, tels que la nature et la gravité de la maladie ou de l'affection, les traitements requis, la réponse du patient

au traitement, la présence de complications ou de comorbidités, l'âge du patient et la disponibilité des lits d'hospitalisation. [14]

Elle est aussi un indicateur clé dans les systèmes de santé car elle peut être utilisée pour évaluer l'efficacité des soins de santé et la qualité des services hospitaliers. [15]

Figure I. 3: Séjour hospitalier [12]

4.1 Importance de la DDSH

La durée moyenne de séjour à l'hôpital est souvent considérée comme un indicateur de l'efficacité des services de santé. En général, un séjour hospitalier de courte durée permet de réduire les coûts par patient et de transférer les soins vers des établissements de santé moins coûteux. Les séjours prolongés peuvent indiquer une mauvaise coordination des soins, ce qui entraîne des attentes inutiles à l'hôpital pour l'organisation de soins de rééducation ou de soins de longue durée. Cependant, il est également possible que certains patients soient libérés trop tôt, alors qu'un séjour plus long aurait peut-être pu améliorer leur état de santé ou réduire le risque de réadmission à l'hôpital. [16]

4.2 Management des lits

Il vise à optimiser les ressources d'un établissement de santé en exploitant au maximum les capacités de prise en charge des patients en attente de soins. Bien que la DDS soit un indicateur clé, la gestion des lits est un outil utilisé pour rechercher l'optimisation, en respectant des critères de qualité tels que la qualité des soins, les délais, la durée et la fiabilité de la programmation. Son objectif est de fournir une prise en charge efficace tout en assurant la meilleure utilisation des ressources disponibles. [17]

4.3 Facteurs qui influencent la DDSH :

DDSH est influencée par de nombreux facteurs qui peuvent varier d'un patient à l'autre. Voici quelques-uns des principaux facteurs qui peuvent influencer la durée du séjour hospitalier : [18]

- **Âge et sexe** :

  En ce qui concerne l'âge, les personnes âgées ont tendance à nécessiter des soins plus complexes et plus longs que les plus jeunes en raison de leur vulnérabilité accrue aux maladies et aux complications. [19]

- **Gravité de la maladie ou de l'affection :**

  La nature et la gravité de la maladie peuvent également affecter la durée du séjour hospitalier, car certaines maladies ou affections nécessitent plus de temps pour guérir ou nécessitent des traitements plus complexes.

  Les patients atteints de maladies graves ou qui nécessitent une intervention chirurgicale complexe ont souvent besoin d'une hospitalisation plus longue pour récupérer. [20]

- **Facteurs socio-économiques :** les patients qui ont un accès limité aux soins de santé ou qui ont un niveau socio-économique plus faible peuvent avoir besoin d'une hospitalisation plus longue pour une meilleure prise en charge. [21]

- **Facteurs psychologiques** : les patients qui souffrent de troubles psychologiques tels que l'anxiété ou la dépression peuvent nécessiter une hospitalisation plus longue pour leur permettre de se rétablir avant de rentrer chez eux. [22]

En fin de compte, DDSH dépend de plusieurs facteurs et varie d'un patient à l'autre. Les professionnels de la santé travaillent à optimiser DDS pour fournir des soins efficaces tout en minimisant les risques d'infections nosocomiales et les coûts pour le système de santé.

4.5 Avantages et inconvénients de la DDSH :

Le tableau I.1 représente les avantages et des inconvénients de la durée de séjour dans les hôpitaux.

Il est important de noter que les avantages et les inconvénients peuvent varier en fonction de la durée de séjour spécifique, de la nature de la maladie ou de l'affection traitée, des ressources disponibles dans le système de santé, des politiques de remboursement des assurances et d'autres facteurs. Il est donc essentiel de prendre en compte chaque situation individuelle pour déterminer la durée de séjour appropriée pour un patient donné.

Tableau. I. 1: Avantages et inconvénients de la DDSH. [23]

Avantages	Inconvénients
Meilleure qualité de soin.	Coûts plus élevés pour les patients et les systèmes de santé.
Meilleur évaluation et traitements des problèmes.	Risques d'infection nosocomiale accrue.
Accès aux soins plus rapides pour les autres patients.	Diminution de la qualité de vie pour les patients hospitalisés.
Elle peut réduire le risque de réadmissions précoces.	Retards dans l'accès aux soins pour les autres patients en attente de traitement.

4. Programme de Médicalisation des Systèmes d'Information (PMSI)

PMSI est un outil essentiel pour la gestion et l'optimisation des données médicales dans les établissements de santé voici la définition et les utilisations de PMSI :

4.1. Définition du PMSI

Modèle introduit en France suite aux travaux de Robert Fetter en 1986. Il permet de concilier les perspectives médicales et économiques sur les coûts de la santé. Le PMSI se distingue par sa collecte d'informations de qualité, en définissant chaque cas de patient hospitalisé en fonction de caractéristiques spécifiques et en évaluant le coût de sa pathologie en fonction de son état de santé individuel. Ainsi, il fournit des données précieuses pour l'analyse et la gestion des ressources de santé. [24].

4.2. Utilisations du PMSI :

Voici quelques exemples d'utilisations du PMSI comme le montre le tableau I.2 :

Tableau. I. 2: Utilisations du PMSI [25]

Utilisation	Description
Évaluation des pathologies les plus fréquentes.	Tableau des diagnostics principaux et secondaires des patients hospitalisés, ainsi que leur fréquence d'occurrence.
Analyse des pratiques médicales.	Tableau des actes médicaux réalisés sur les patients hospitalisés, tels que les interventions chirurgicales, les examens d'imagerie médicale, les prises de sang…etc.
Évaluation des coûts des soins.	Tableau des coûts associés à chaque patient hospitalisé, tels que les frais d'hospitalisation, les coûts des actes médicaux, les frais de médicaments…etc.
Suivi des durées de séjour.	Tableau des durées de séjour des patients hospitalisés, ainsi que des indicateurs tels que le taux de séjours courts et longs.
Évaluation de la qualité des soins.	Tableau des indicateurs de qualité tels que les taux de réadmission, les complications post-opératoires, les infections nosocomiales…etc.

Ce tableau permet de visualiser et d'analyser les données médicales des patients hospitalisés, afin de comprendre les tendances, les écarts par rapport aux pratiques habituelles, les coûts associés à chaque pathologie et les opportunités d'amélioration de la qualité des soins et de l'efficacité de la gestion des établissements de santé.

4.3. PMSI en Algérie :

En Algérie, le PMSI est un outil de gestion de l'activité hospitalière qui a été mis en place progressivement depuis les années 2000. Il est géré par le Ministère de la Santé et est obligatoire pour tous les établissements hospitaliers publics. Le PMSI algérien vise à améliorer la qualité des soins, la sécurité des patients et l'efficacité de la gestion des établissements de santé en fournissant des données précises sur l'activité médicale. Cependant, il reste encore des défis à relever pour assurer une généralisation efficace de cet outil dans l'ensemble des établissements de santé en Algérie. [26]

4.4. Objectif du PMSI

C'est un dispositif **médico-économique** qui vise à optimiser l'organisation des soins dans les établissements de santé en collectant, traitant et diffusant des informations médicales relatives aux patients hospitalisés. Il permet également **d'améliorer l'organisation économique** en évaluant les coûts associés à chaque pathologie et en identifiant les pratiques les plus efficientes en termes de ressources. Le but du PMSI était de donner aux hôpitaux une méthode qui leur permette d'estimer et de mesurer leur activité. [27]

Voici quelques objectifs de PMSI :

- ✓ Collecter, traiter et diffuser des informations médicales relatives aux patients hospitalisés ;
- ✓ Améliorer l'organisation des soins dans les établissements de santé.
- ✓ Optimiser l'organisation économique en évaluant les coûts associés à chaque pathologie ;
- ✓ Identifier les pratiques les plus efficientes en termes de ressources ;
- ✓ Évaluer la qualité des soins et les opportunités d'amélioration ;
- ✓ Permettre une meilleure coordination entre les différents acteurs de la santé ;
- ✓ Contribuer à la recherche et à la planification de la santé publique.

4.5. Types du PMSI

Il existe plusieurs types de PMSI qui sont adaptés aux différentes formes d'hospitalisation [28]:

a) **PMSI-MCO** : est utilisé pour les patients hospitalisés dans les services de médecine, chirurgie et obstétrique.

b) **PMSI-SSR** : est utilisé pour les patients hospitalisés dans les services de soins de suite et de réadaptation, qui ont pour objectif de favoriser la récupération après une hospitalisation en médecine, chirurgie ou obstétrique.

c) **PMSI-HAD** : est utilisé pour les patients qui sont hospitalisés à domicile pour des soins médicaux.

d) **PMSI-PSY** : est utilisé pour les patients hospitalisés dans les services de psychiatrie.

Ces différents types de PMSI permettent de collecter des informations spécifiques en fonction des pathologies et des types de soins dispensés, et sont utilisés pour l'analyse et l'optimisation de l'organisation des soins dans les établissements de santé.

5. Systèmes d'Information Hospitaliers « SIH »

En intégrant les processus de soins, de gestion administrative et de prise de décision, les SIH contribuent à améliorer la qualité des soins, à réduire les coûts et à renforcer la sécurité des patients.

5.1. Système Information

Le SI est composé d'éléments tels que des informations, des processus de traitement, des règles d'organisation, ainsi que des ressources humaines et techniques. Les ensembles d'informations constituent des représentations partielles des faits pertinents pour une institution, une organisation ou une entreprise. [29]

5.2. Définition de SIH

Le SI est conçu pour simplifier la gestion complète des données médicales et administratives d'un établissement hospitalier. Il permet de gérer l'ensemble des données relatives à la prise en charge des patients dans un établissement de santé, qu'il s'agisse d'un hôpital, d'une clinique ou d'un centre de soins. Il permet la gestion des informations médicales et administratives des patients, la planification des rendez-vous et des interventions chirurgicales, la gestion des ressources humaines, la gestion des stocks de médicaments et de matériel médical, la facturation et le suivi de la qualité des soins. [30]

L'un des avantages les plus importants est la réduction des erreurs médicales. Les dossiers médicaux électroniques permettent aux professionnels de santé d'accéder rapidement et facilement aux antécédents médicaux des patients, ce qui réduit les risques d'erreurs médicales.

SIH est conçu pour : [31]

a. **Faciliter la centralisation, l'enregistrement, l'organisation et la diffusion des informations médicales et administratives des patients** : cela favorise une coordination accrue entre les divers professionnels de santé impliqués dans les soins des patients, ce qui peut entraîner une amélioration de la qualité des prestations médicales ;
b. **Améliorer l'efficacité de la gestion des ressources :** le SIH peut aider à optimiser la gestion des ressources humaines, des équipements médicaux, des médicaments et autres fournitures nécessaires aux soins de santé, ce qui peut réduire les coûts et améliorer l'efficacité des processus ;
c. **Améliorer la sécurité des patients :** le SIH peut aider à éviter les erreurs médicales en garantissant que les informations médicales sont complètes, précises et accessibles en temps voulu ;

d. **Stocker et de traiter des données provenant de diverses sources**, telles que les dossiers médicaux électroniques, les résultats d'examens, les ordonnances, les données financières et les données de gestion des ressources humaines. Ces données peuvent être utilisées pour identifier les tendances, les risques et les opportunités d'amélioration dans la qualité des soins de santé ;

e. **Améliorer la pertinence** et la continuité de la prise en charge sans rupture pour le patient ;

f. **Harmoniser** les services numériques des établissements de santé.

5.3. Sous-système d'information hospitalier

Les trois sous-systèmes principaux du SIH sont :

a. Sous-système administratif :

Il gère les informations relatives à la gestion administrative et financière de l'hôpital, telles que la gestion des rendez-vous, des admissions, des factures, des ressources humaines, etc. [32]

b. Sous-système clinique :

Il gère les informations cliniques des patients, telles que les dossiers médicaux, les ordonnances, les résultats d'examens, les antécédents médicaux, etc. Ce sous-système permet de faciliter la prise de décision médicale en mettant à disposition des professionnels de santé les informations nécessaires pour un traitement efficace des patients. [33]

c. Sous-système de gestion des ressources :

Il gère les ressources de l'hôpital, telles que la gestion des stocks, des médicaments, des équipements médicaux…etc. Ce sous-système permet une gestion optimale des ressources et une maîtrise des coûts. [33]

La figure I.4. Représente **schéma** des trois sous-systèmes du SIH, comme nous pouvons le voir, chaque sous-système à ses propres fonctions qui sont interconnectés avec les autres pour former un système global de gestion des informations de santé dans un établissement hospitalier.

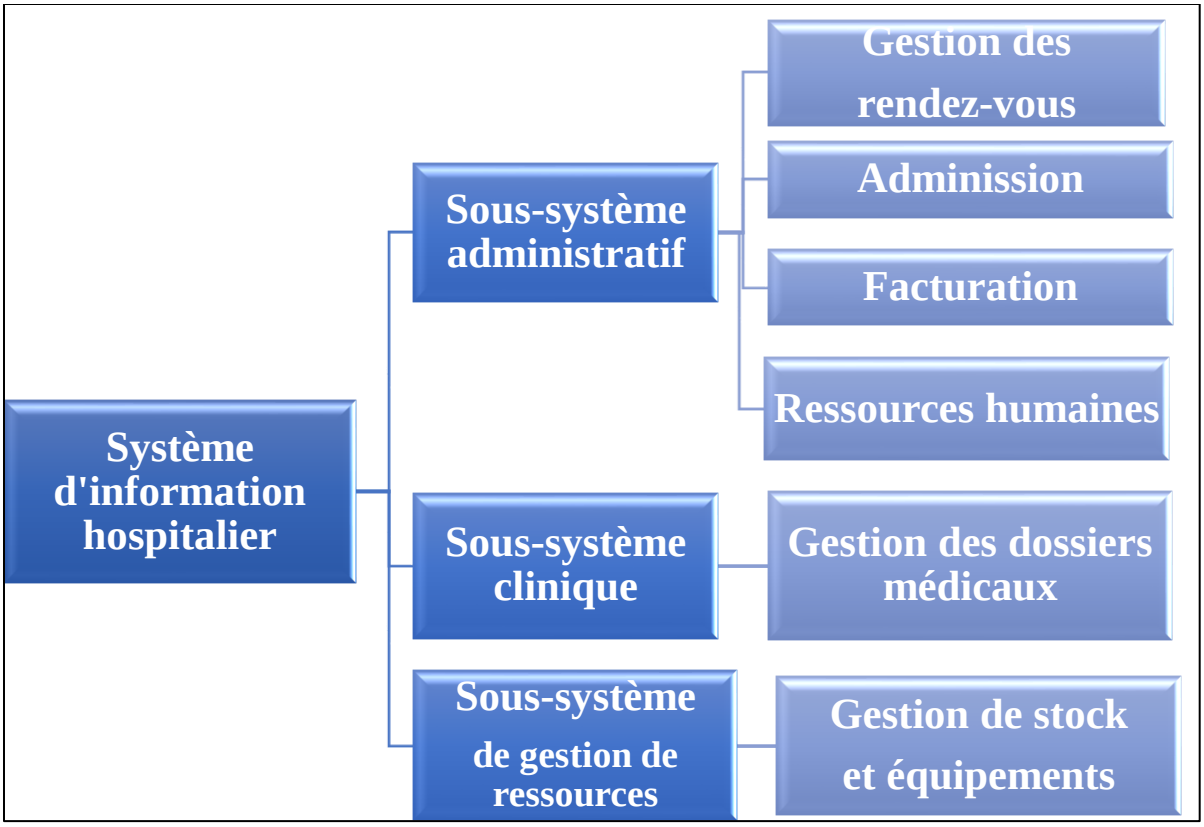


Figure I. 4 : SIH

5.4 Définition des informations administratives

Pour chaque individu bénéficiant de soins dans un établissement de santé, l'administration hospitalière est chargée de constituer un dossier administratif distinct du dossier médical. Ce dossier administratif contient des informations telles que l'identification du patient et des données sociodémographiques qui viennent enrichir le dossier médical. Il est essentiel de garantir l'authenticité des informations administratives collectées et de les maintenir régulièrement à jour, notamment en ce qui concerne l'état civil, la couverture sociale, le statut matrimonial, les employeurs, et autres données pertinentes. [34]

5.5. Professionnels de santé

Les professionnels de santé sont des acteurs clés du système de santé, chargés de prodiguer des soins de qualité aux patients.

a. Définition du professionnel de santé

Ce sont des personnes formées et autorisées à fournir des soins médicaux, tels que des médecins, des infirmières, des dentistes, des psychologues, des physiothérapeutes, des pharmaciens, des techniciens de laboratoire…etc.

Ils ont reçu une formation spécialisée dans leur domaine et sont responsables de diagnostiquer, traiter et prévenir les maladies et les problèmes de santé, ainsi que de prodiguer des soins médicaux et des traitements appropriés aux patients. Les professionnels de santé sont réglementés par des organismes de réglementation professionnels qui supervisent leur pratique et assurent que les normes de soins de santé sont respectées. [35]

b. Informations des professionnels de santé :

Les informations des professionnels de santé font référence à toutes les informations et connaissances pertinentes pour les professionnels de santé dans leur pratique professionnelle. Cela peut inclure des informations sur les maladies, les traitements, les médicaments, les procédures médicales, les avancées technologiques, les recherches cliniques, les normes de pratique, les réglementations et les politiques de santé.

6. Dossier du patient (DP)

Le DP contient les informations essentielles sur sa santé, être en papier ou informatisé. C'est un pilier essentiel de la prise en charge du patient, rassemblant de manière écrite les données cliniques, biologiques, diagnostiques et thérapeutiques d'un individu. Il constitue à la fois un enregistrement individuel et collectif, constamment actualisé, des informations concernant le patient. [36]

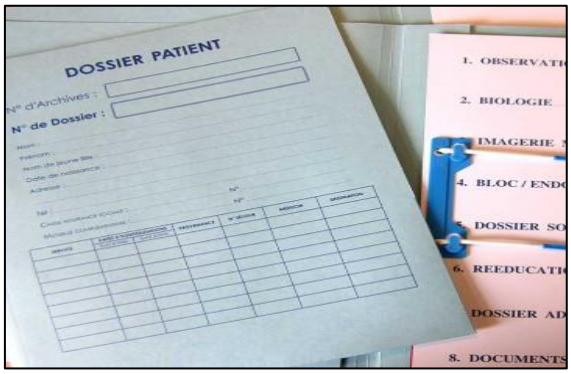


Figure I. 5 : DP [37]

6.1. Dossier médical du patient (DMP)

C'est l'endroit où les informations administratives sont collectées et préservées. Il est établi lors du premier contact du patient avec l'établissement de soins, que ce soit lors d'une consultation externe ou d'une hospitalisation. Par la suite, il est enrichi avec les détails des interventions effectuées par les différents professionnels de santé rencontrés par le patient tout au long de son parcours dans l'établissement de santé. À la fin de chaque interaction, le dossier est classé et archivé pour une conservation ultérieure. [38]

- Fournir les informations nécessaires et pertinentes pour assurer une prise en charge efficace et un suivi adéquat du patient.
- Assurer une traçabilité complète des soins et des actions entreprises à l'égard du patient.
- Garantir la continuité des soins tout au long du parcours du patient.
- Servir de lieu pour obtenir le consentement éclairé du patient, analyser les avantages et les risques, et enregistrer les décisions prises.
- Évaluer la qualité des soins prodigués et la précision de la documentation médicale.
- Contribuer à l'enseignement et à la recherche médicale. [39]

6.2. Rôle du dossier du patient

Le **DP** joue en effet un rôle important en tant que mémoire du patient et des professionnels, de communication et de coordination : [40]

a) Mémoire du patient et des professionnels : Le dossier du patient est une source de référence pour les professionnels de santé qui suivent le patient, car il contient toutes les informations pertinentes sur son état de santé, son historique médical, ses allergies, les traitements qu'il a reçus et les résultats de ses examens. Il permet également au patient de connaître et de suivre son propre parcours de soins.

b) Communication : Le dossier du patient est un moyen important de communication entre les différents professionnels de santé impliqués dans les soins du patient. Il permet de transmettre les informations pertinentes à chaque intervenant et de s'assurer que tous les professionnels sont sur la même page quant à l'état de santé du patient.

c) Coordination : Le dossier du patient est un outil essentiel pour la coordination des soins. Il permet aux différents professionnels de santé de travailler ensemble pour élaborer un plan de soins cohérent et adapté aux besoins du patient. Le dossier du patient peut également faciliter le transfert du patient entre les différents services de santé, ce qui est particulièrement important dans les cas où le patient nécessite des soins complexes ou spécialisés.

6.3. Contenu de DMP

- Informations administratives ;
- Informations des professionnels de santé.

a. Informations administratives du DMP [41]

Le dossier de patient contient plusieurs informations administratives, notamment :

- ✓ Informations d'identification du patient : nom, prénom, date de naissance, sexe, adresse, numéro de téléphone, adresse e-mail…etc ;
- ✓ Les informations de contact d'urgence : nom, prénom, relation avec le patient, numéro de téléphone, adresse e-mail…etc ;
- ✓ Les informations sur l'assurance maladie : numéro d'assurance maladie, nom de l'assureur, type de couverture…etc ;
- ✓ Les informations sur les antécédents médicaux : maladies préexistantes, allergies, interventions chirurgicales antérieures, médicaments pris régulièrement…etc ;
- ✓ Les informations sur les rendez-vous médicaux : date, heure, lieu, nom du médecin, nature de la consultation…etc ;
- ✓ Les informations sur les hospitalisations précédentes : dates, noms des établissements de santé, nature des traitements reçus…etc.

Ces informations administratives sont essentielles pour identifier le patient de manière fiable, communiquer avec lui et avec les membres de sa famille, établir un plan de soins approprié, suivre l'évolution de son état de santé, respecter ses choix en matière de traitement et assurer une coordination efficace des soins entre les différents professionnels de santé impliqués.

b. Informations professionnelles de santé du DMP

Les informations sur les professionnels de la santé qui ont accès au **DMP** indiquent qui peut consulter les informations de santé du patient dans le **DMP**. Les informations de santé confidentielles ne peuvent être partagées qu'entre les professionnels de la santé impliqués dans la prise en charge du patient et ne sont pas divulguées au patient sans autorisation.

6.4. Dossier patient électronique (DPE) :

Au début des années 2000, la gestion des informations liées à la prise en charge des patients a connu une évolution majeure. Auparavant, ces données étaient conservées et présentées sous la forme de dossiers médicaux papier distincts du dossier infirmier. Cependant, grâce à la numérisation et aux nouveaux outils technologiques dans le domaine de

la santé, les informations relatives aux patients ont été centralisées et regroupées au sein d'un dossier patient informatisé. [42]

Figure I. 6: DPE [43]

Le dossier patient informatisé est spécifiquement élaboré pour faciliter la création et la gestion électronique d'un dossier exhaustif sur chaque patient. Ce dossier regroupe de manière centralisée toutes les informations administratives et médicales relatives au patient, telles que son identité, ses antécédents médicaux et les traitements en cours, dans un format électronique. [44]

6.5. Suggestions pour la Transition numérique médicale :

a. Former et impliquer les professionnels de santé : Il est essentiel de former les professionnels de santé à l'utilisation des outils numériques et de les impliquer dans le processus de déploiement. Les formations devraient être adaptées aux besoins de chaque utilisateur et inclure une formation sur l'utilisation du système.

b. Sécurité des données médicales La gestion de l'information de santé est une question de sécurité et de confidentialité. Il est donc important de mettre en place des solutions de sauvegarde et de sécurité efficaces pour protéger les informations sensibles des patients.

En résumé, pour assurer la réussite de l'implémentation d'un dossier patient informatisé, il est primordial de se préparer de manière adéquate, de gérer efficacement le processus de changement, et d'offrir une formation et un soutien appropriés aux acteurs de la santé.

7. Donnée de santé :

Ce sont des informations concernent la santé physique ou mentale d'une personne, qu'elle soit passée, présente ou future. [45]. Les données à caractère personnel relatives à la santé incluent les informations sur la santé physique ou mentale d'une personne, qu'elles concernent

le passé, le présent ou l'avenir, y compris les services de soins de santé, et qui révèlent des détails sur son état de santé. [46]

Les données de santé font référence à toutes les informations relatives à la santé d'un individu ou d'une population, qu'elles soient collectées, enregistrées, stockées ou traitées de manière électronique ou non. Ces données peuvent inclure les informations médicales, les données d'assurance maladie, les données de recherche, les données de santé publique, les données génétiques, les données de télémédecine…etc.

7.1. Catégories de données de santé

Peuvent être classées en plusieurs catégories en fonction de leur nature et de leur usage, notamment :

a) Données personnelles de santé :

L'article 2 de la loi "Informatique et Libertés" définit le cadre juridique relatif à la protection des données. Selon cette loi, toute information permettant d'identifier directement ou indirectement une personne physique est considérée comme une donnée à caractère personnel. [47]

Les données personnelles de santé il s'agit des données relatives à l'état de santé, aux antécédents médicaux, aux traitements, aux résultats d'examens médicaux, etc., d'un individu. Ces données sont souvent collectées dans le cadre de la prise en charge médicale d'un patient ou permettent son identification en cas de poursuites judiciaires.

b) Données de recherche clinique : englobent les informations recueillies dans le cadre d'études cliniques qui évaluent l'efficacité et la sécurité des médicaments, dispositifs médicaux, thérapies…etc. Dans le domaine de la recherche clinique, une base de données regroupe un ou plusieurs groupes de patients partageant une caractéristique commune, formant ainsi une cohorte de patients. L'objectif principal de ces bases de données est d'améliorer la prise en charge des patients, d'aider au diagnostic des maladies, de contribuer à la surveillance sanitaire et de soutenir la recherche médicale. Les informations recueillies sont déterminées lors de la création de la base de données et peuvent être modifiées au fil du temps. [48]

c) Données de santé publique :

La santé publique se caractérise par son aspect organisationnel et préventif. Bien que la santé individuelle ait un lien direct avec celle des populations, l'approche de la santé publique est axée sur la dimension collective. [49]

- **Données de santé publique :** il s'agit des données relatives à la santé d'une population,

telles que les statistiques de mortalité, les taux de prévalence de maladies, les facteurs de risque…etc. Il est important de noter que les données de santé sont considérées comme des données sensibles et nécessitent donc une protection particulière en matière de sécurité et de confidentialité, conformément aux réglementations nationales et internationales en vigueur.

7.2. Propriétés des données médicales

Les données médicales sont des informations sensibles et confidentielles qui nécessitent une protection rigoureuse. Leur précision et leur intégrité sont essentielles pour garantir la qualité des soins et la sécurité des patients, voici quelques propriétés courantes des données médicales :

a) **Traitement des données médicales :** Étant interdit par le droit européen, il ne devrait pas y avoir de danger pour les personnes concernées.: « Les données, qui sont susceptibles par leur nature de porter atteinte aux libertés fondamentales ou à la vie privée, ne devraient pas faire l'objet d'un traitement… » [50]

Le Règlement Général sur la Protection des Données (RGPD), qui est en vigueur depuis le 25 Mai 2018, a pour but de simplifier l'accès aux données personnelles des individus et de renforcer leurs droits sur ces données. [51]

b) **Confidentialité :** selon l'article 4 du Règlement Général sur la Protection des Données (RGPD) de l'Union Européenne : « les données relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne » sont définies comme définis à caractère personnel. Ces donnes doivent donc être protégées et une politique et une démarche de sécurité de ces données doivent être définies pour les protéger. [52]

Chaque patient est appelé à disposer à terme d'un seul dossier, et non de plusieurs comme actuellement, les données médicales doivent être traitées avec le plus grand respect pour la vie privée du patient. Les informations ne doivent être accessibles qu'aux personnes autorisées, telles que les professionnels de la santé impliqués dans les soins du patient, données à caractère personnel. [53]

c) **Disponibilité :** les données doivent être accessibles. [54]

Les données médicales doivent être disponibles pour les professionnels de la santé autorisés qui en ont besoin pour prendre des décisions éclairées sur les soins du patient. Cela peut inclure les **DPE** et les résultats de tests.

d) **Intégrité :** les données doivent être fiables, exactes, complètes et à jour. [55]

e) **Sécurité :** Dans le contexte du **RGPD (Règlement général sur la protection des données)**, il est impératif de garantir la confidentialité des données de santé des patients. Cela nécessite la mise en place de mesures de sécurité appropriées pour prévenir tout accès non autorisé, la destruction accidentelle, la perte, etc. Ces mesures peuvent inclure l'utilisation d'un mot de passe personnel, l'adoption d'un système de chiffrement fiable, et d'autres dispositifs de protection des données. [56]

Les données médicales doivent être stockées et transmises de manière sécurisée pour éviter toute violation de la confidentialité ou de l'intégrité des données. Cela peut inclure des mesures de sécurité telles que des pares-feux, des protocoles de chiffrement et des politiques d'accès.

7.3. Sources des données médicales des systèmes information hospitalier (SIH) :

Il y'a plusieurs sources des données médicales de SIH voici quelque source :

a) **Données administratives :** les informations d'identification du patient, les informations de facturation et les informations d'assurance, peuvent également être stockées dans le SIH pour faciliter la gestion des activités administratives de l'hôpital.

b) **DPE** : contiennent toutes les informations sur la santé et les antécédents médicaux du patient, les diagnostics, les traitements, les médicaments prescrits, les résultats d'examens médicaux, les images médicales…etc. Les données sont saisies par les professionnels de santé qui fournissent les soins.

c) **Enquêtes et les recherches cliniques :** peuvent également fournir des données importantes pour le SIH. Ces études peuvent inclure des enquêtes sur la satisfaction des patients, des études épidémiologiques sur les taux de maladies dans la population locale ou des études de recherche pour évaluer l'efficacité des nouveaux traitements.

d) **Systèmes de laboratoire** : Les résultats des tests de laboratoire, tels que les analyses sanguines, les tests d'imagerie médicale, les tests de diagnostic moléculaire…etc, sont stockés dans les SIH.

e) **Systèmes de facturation et de gestion financière** : Les informations financières, telles que les factures, les paiements, les coûts de traitement, les remboursements, etc., sont stockées dans les SIH pour aider à la gestion financière de l'établissement de santé.

f) **Systèmes de gestion de la qualité :** Les données de qualité, telles que les indicateurs de performance, les audits, les enquêtes de satisfaction des patients, les rapports d'incidents, etc., sont stockées dans les SIH pour aider à améliorer la qualité des soins

et des services.

En résumé, les données des SIH proviennent de diverses sources qui sont intégrées pour fournir une vue complète des informations sur les patients, les soins médicaux et les ressources hospitalières. Les données sont utilisées pour améliorer la qualité des soins, pour aider à la prise de décision et pour la gestion de l'établissement de santé.

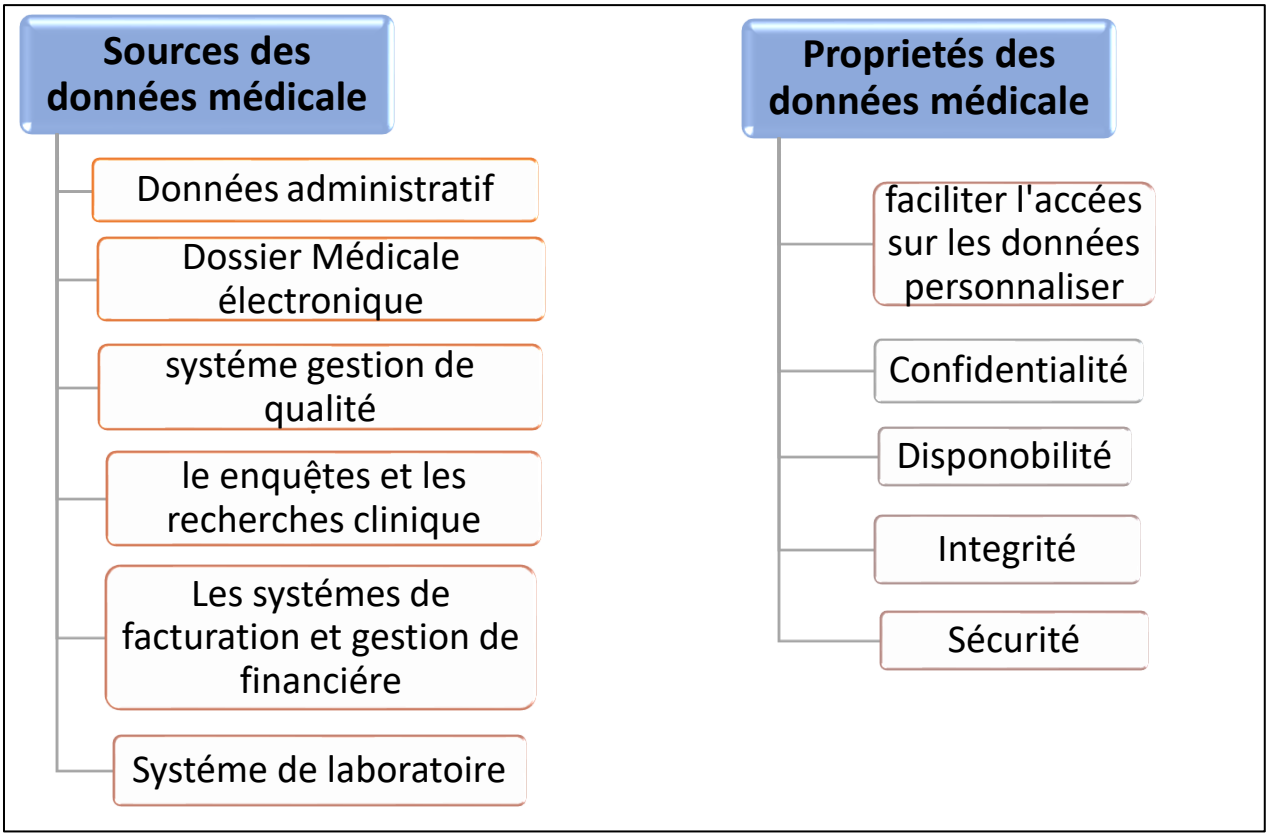


Figure I. 7: Données de santé

8. Conclusion :

En conclusion, la durée de séjour hospitalier est un indicateur important de la qualité des soins de santé et de l'efficacité des établissements de santé. Il permet de mesurer la performance des établissements de santé et de suivre l'évolution de leur activité au fil du temps. Une durée de séjour optimale peut contribuer à améliorer la qualité des soins, à réduire les coûts et à optimiser l'utilisation des ressources. Il est donc essentiel de surveiller et d'analyser la durée de séjour hospitalier afin d'identifier les domaines qui nécessitent des améliorations et d'élaborer des stratégies pour optimiser la qualité des soins et des services offerts aux patients.

CHAPITRE II

Intelligence Artificiel et la prédiction de la durée de séjour hospitalier

1. Introduction :

L'utilisation de l'intelligence artificielle (IA) et de l'apprentissage automatique a révolutionné la prédiction de la durée de séjour hospitalier. Les modèles statistiques traditionnels étaient limités en termes de précision, de mise à jour des modèles et de gestion des données. L'IA et l'apprentissage automatique offrent des méthodes plus précises, actualisables et automatisées pour prédire la durée de séjour à l'hôpital. Dans ce chapitre, nous explorerons l'apprentissage supervisé et non supervisé pour prédire la durée de séjour, en utilisant des méthodes basées sur des données historiques et des caractéristiques cliniques. Ces avancées permettent aux cliniciens de mieux planifier les soins et les ressources, et d'identifier les patients à haut risque de rester plus longtemps à l'hôpital.

2. Intelligence Artificielle :

Le terme IA est appliqué lorsqu'un appareil imite des fonctions cognitives, telles que l'apprentissage et la résolution de problèmes [57]

IA Un domaine émergent d'enseignement et de recherche scientifique est apparu, ayant pour objectif de développer des machines capables d'apprendre de leurs erreurs, de comprendre leur environnement et le monde en général, de s'adapter aux changements, d'anticiper et de prédire l'avenir, et d'agir en vue d'améliorer la condition humaine. Bien que cette ambition ait longtemps été considérée comme farfelue ou irréaliste, elle persiste aujourd'hui. [58]

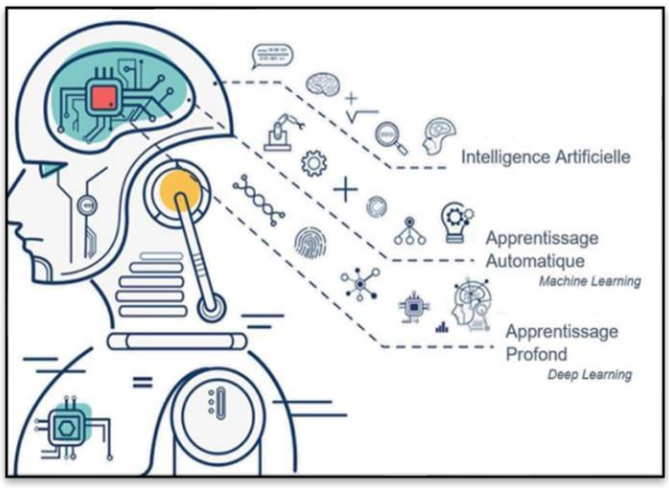


Figure II. 1: IA [59]

2.1. Techniques de l'IA

IA est une branche de l'informatique qui consiste à créer des machines capables de réaliser des tâches qui, si elles étaient accomplies par un être humain, nécessiteraient de l'intelligence pour les exécuter. Il s'agit d'une technologie qui permet à des machines de percevoir leur environnement, d'apprendre à partir de l'expérience, de raisonner et de prendre des décisions autonomes en fonction des données qu'elles ont collectées [60]. La figure II.2 montre ces techniques

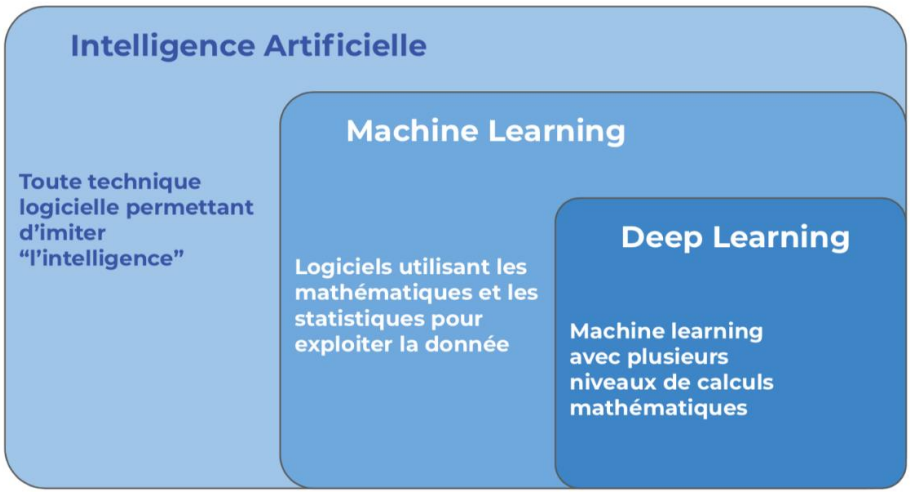


Figure II. 2 : Techniques de l'IA. [61]

3. Apprentissage automatique (Machine Learning)

Il est fréquent de constater que les définitions de "Machine Learning" (apprentissage automatique) varient selon les sources et les contextes, tout comme pour la plupart des termes utilisés dans le domaine de l'IA, l'un des sous-domaines de l'intelligence artificielle.

3.1. Définition

L'apprentissage automatique joue un rôle essentiel dans le domaine en plein essor de la science des données. En utilisant des méthodes statistiques, les algorithmes sont entraînés à effectuer des classifications, des prédictions et à découvrir des informations importantes dans les projets d'exploration de données. Ces informations sont ensuite utilisées pour prendre des décisions dans le cadre d'applications et d'entreprises. [62]

3.2. Domaine d'application de l'apprentissage automatique

L'apprentissage automatique est utilisé dans de nombreuses applications telles que : la reconnaissance d'images, la traduction automatique, la recommandation de produits, la détection de fraudes, la prédiction de résultats, et l'optimisation de processus. Il est également

utilisé dans des applications plus spécifiques, comme la détection de spams, la reconnaissance vocale, ou l'analyse de données, le traitement du langage naturel (traduction automatique, compréhension du langage, synthèse de texte…etc), l'optimisation de processus (optimisation de la production, planification de la chaîne d'approvisionnement…etc).

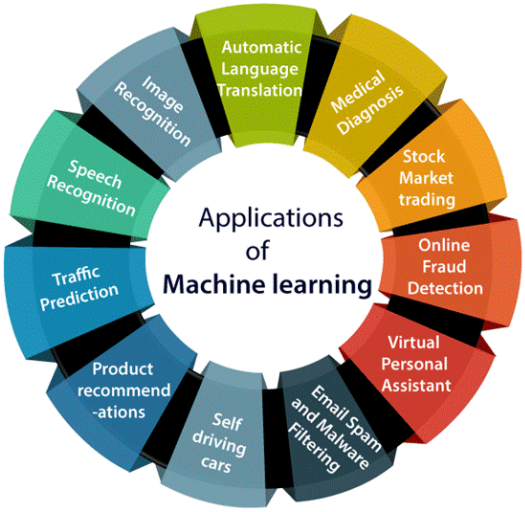


Figure II. 3: Applications ML [63]

3.3 Types de l'apprentissage automatique :

L'apprentissage automatique implique de présenter à une machine une grande quantité de données pour qu'elle puisse apprendre et faire des prédictions, trouver des motifs ou classer des données. Les trois types d'apprentissage automatique sont supervisé, non supervisé et par renforcement, voici le schéma suivant :

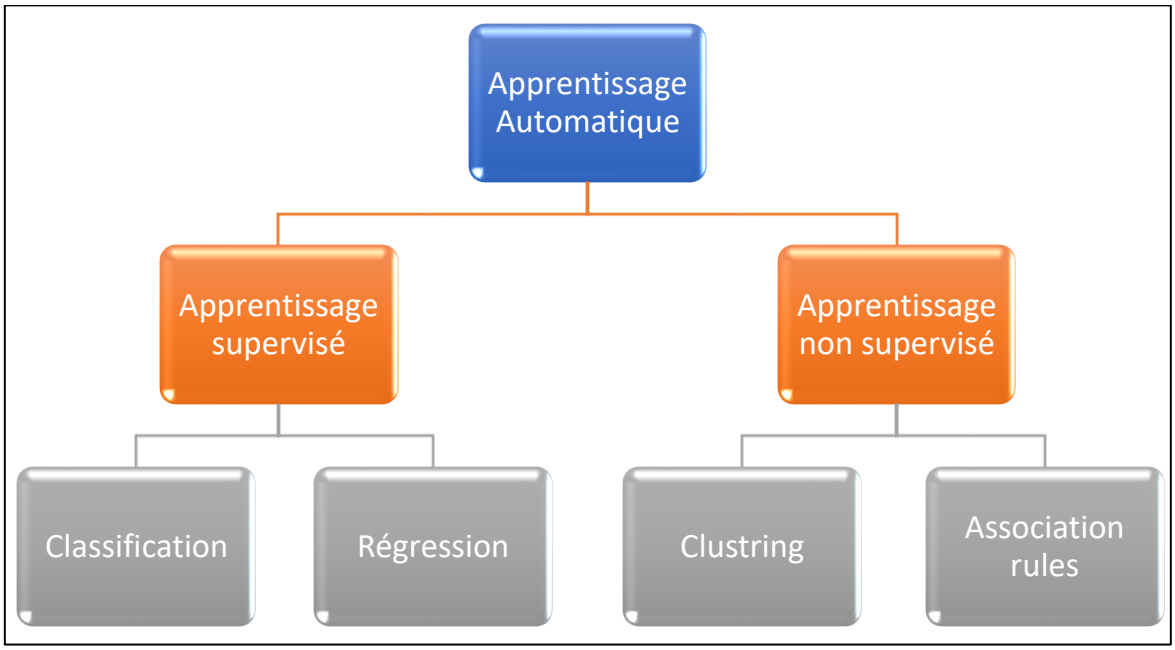


Figure II.4: Type ML

Selon le schéma, nous avons deux types D'apprentissage :

- Apprentissage supervisé
- Apprentissage non supervisé

4. Apprentissage supervisé (Supervised Learning) :

L'apprentissage supervisé est une technique d'apprentissage automatique où un modèle est entraîné à partir de données qui sont annotées afin de prédire les annotations des nouvelles données. En d'autres termes, le modèle est entraîné sur un ensemble de données contenant des entrées et des sorties attendues, et il est capable de prédire la sortie correspondante pour de nouvelles entrées qu'il n'a jamais vues auparavant.

L'apprentissage supervisé est le paradigme dominant en Machine Learning et en Deep Learning. Ce concept implique de guider l'apprentissage de la machine en lui fournissant des exemples (des données) de la tâche à réaliser, comme son nom l'indique. [64]

4.1. Fonctionnement de l'apprentissage supervisé [65]

L'apprentissage supervisé se déroule généralement en quatre étapes :

- ✓ Importer un ensemble de données (DataSet) contenant les exemples sur lesquels nous allons travailler.
- ✓ Développer un modèle initial avec des paramètres aléatoires.
- ✓ Définir une fonction de coût qui mesure les erreurs entre le modèle et les données de l'ensemble de données.
- ✓ Utiliser un algorithme d'apprentissage pour ajuster les paramètres du modèle de manière à minimiser la fonction de coût.

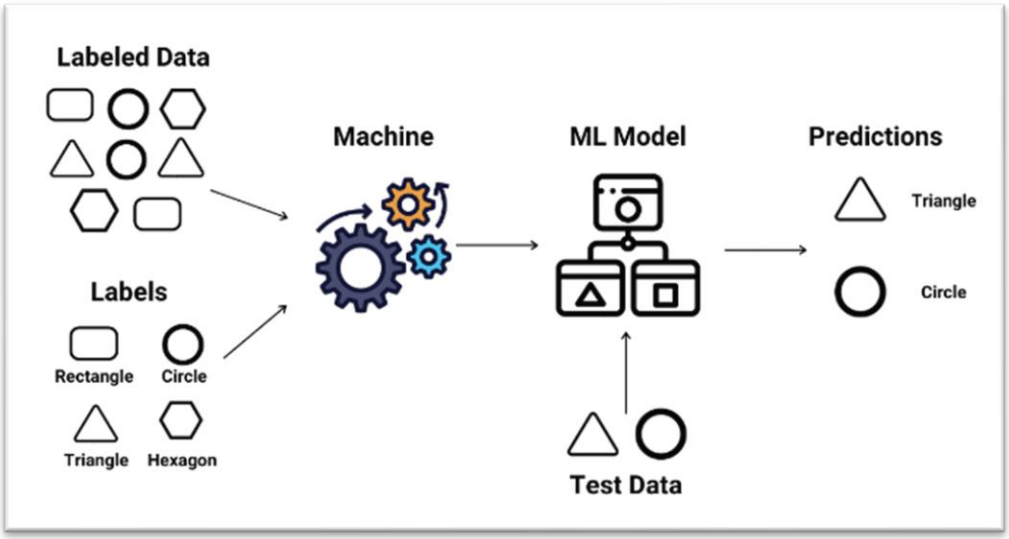


Figure II. 5 : Apprentissage supervisé [66]

4.2. Types Apprentissage supervisé :

a) Classification :

Les modèles de classification sont des modèles d'apprentissage automatique qui visent à prédire la classe ou la catégorie à laquelle appartient une observation, en fonction de ses caractéristiques ou de ses variables d'entrée. Partie de l'apprentissage supervisé, ils nécessitent des données étiquetées pour entraîner le modèle. Le but de la classification est de prédire la classe d'un nouvel exemple, en fonction de la similarité de ses caractéristiques avec les données d'entraînement préalablement étiquetées. Il existe plusieurs types de modèles de classification;

- Voici quelques-unes des mesures de classification les plus couramment utilisées et sont formules : [67]

- **Précision :** La précision mesure la proportion de prédictions positives qui sont correctes. Cela signifie que la mesure de la précision est le nombre de vrais positifs divisé par le nombre total de prédictions positives.

$$Precision = \frac{TP}{TP + FP}$$

- **Rappel (sensibilité) :** Le rappel mesure la proportion de cas positifs réels qui ont été correctement prédits. Cela signifie que la mesure de rappel est le nombre de vrais positifs divisé par le nombre total de vrais positifs et de faux négatifs.

$$Sensitivity = \frac{TP}{TP + FN} = 1 - Type\ 2\ error$$

- **F1-score** : L'indice F1 est une mesure de la précision globale du modèle qui combine les mesures de précision et de rappel. Il est calculé comme une moyenne harmonique de la précision et du rappel.

$$F1score = 2 * \frac{Sensitivity * Precision}{Sensitivity + Precision}$$

- **Spécificité :** La spécificité est une mesure qui indique le nombre de prédictions négatives correctes faites par un modèle. Cela se calcule en divisant le nombre de prédictions négatives correctes par le nombre total de prédictions négatives.:

$$Specificity = \frac{TN}{TN + FP} = 1 - Type\ 1\ error$$

- **AUC-ROC :** L'aire sous la courbe ROC (Receiver Operating Characteristic) mesure la capacité du modèle à distinguer les classes positives et négatives. La courbe ROC est un graphique qui trace le taux de vrais positifs en fonction du taux de faux positifs à

différents seuils de classification. Cette courbe représente deux paramètres :

- TPR (True Positive Rate) est le taux de vrais positifs (également appelé sensibilité).
- FPR (False Positive Rate) est le taux de faux positifs.

- **Matrice de confusion :** est un tableau qui présente le nombre de prédictions correctes et incorrectes pour chaque classe dans un modèle de classification. Elle est couramment utilisée pour évaluer et visualiser les performances d'un modèle en comparant les données réelles d'une variable cible avec les prédictions générées par le modèle. Cela permet de quantifier et d'analyser les erreurs de classification effectuées par le modèle.[68]

		Reality	
	Confusion matrix	**Negative : 0**	**Positive : 1**
Prediction	**Negative : 0**	**True Negative : TN**	**False Negative : FN**
	Positive : 1	**False Positive : FP**	**True Positive : TP**

Figure II. 6 : Matrix Confusion [69]

TP : True Positives (vrais positifs)

FP : False Positives (faux positifs)

FN : False Négatives (faux négatifs)

TN : True Négatives (vrais négatifs)

b) Régression : La régression est utilisée pour déterminer la relation entre une variable et une ou plusieurs autres variables. En apprentissage automatique, l'objectif de la régression est d'estimer une valeur numérique de sortie en se basant sur les valeurs d'un ensemble de caractéristiques en entrée. [70]

$$\hat{y}(x) = f(x_1, x_2, ..., x_m)$$

- **Mesures de régression [71]** : Sont des indicateurs statistiques utilisés pour évaluer la qualité de l'ajustement d'un modèle de régression à un ensemble de données. Elles permettent de mesurer à quel point le modèle est capable de prédire les valeurs de la variable dépendante en fonction des variables indépendantes, Voici quelques-unes des mesures de régression les plus courantes **:**

- **Coefficient de détermination (R^2) :** il mesure la proportion de la variance de la variable dépendante expliquée par le modèle. Il varie entre 0 et 1, et plus il est élevé, plus le modèle est considéré comme étant performant.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- **Erreur quadratique moyenne (EQM) :** elle mesure la moyenne des carrés des écarts entre les valeurs prédites par le modèle et les valeurs observées dans les données. Plus l'EQM est faible, plus le modèle est considéré comme étant performant.

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (y_i - \hat{y}_i)^2.$$

- **Erreur absolue moyenne (EAM) :** elle mesure la moyenne des valeurs absolues des écarts entre les valeurs prédites par le modèle et les valeurs observées dans les données. Plus l'EAM est faible, plus le modèle est considéré comme étant performant.

$$\text{MedAE}(y, \hat{y}) = \text{median}(| y_1 - \hat{y}_1 |, \ldots, | y_n - \hat{y}_n |).$$

- **Root Mean Squar Error (RMSE) :** est une mesure couramment utilisée pour évaluer la performance d'un modèle de régression. La RMSE est calculée en prenant la racine carrée de la moyenne des carrés des écarts entre les prédictions et les vraies valeurs. Elle est une métrique de performance qui fournit une mesure de la dispersion entre les prédictions et les vraies valeurs.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

- **Coefficient de corrélation (r) :** il mesure la force et la direction de la relation linéaire entre les variables indépendantes et la variable dépendante. Il varie entre -1 et 1, et plus sa valeur est proche de 1 ou de -1, plus la relation entre les variables est forte.

r = cov (x,y)/(sx*sy)

Où cov(x,y) est la covariance entre les variables x et y, sx est l'écart-type de x et sy est l'écart-type de y.

4.3. Algorithme d'apprentissage supervisé les plus couramment utilisés :

a. Régression linéaire : c'est une méthode de régression qui cherche à établir une relation linéaire entre une variable d'entrée et une variable de sortie continue. Sert à trouver une relation d'une variable de sortie (continue) par rapport à une autre. [72]

$$\hat{y}(x) = \theta_0 + \theta_1 x_1$$

b. Arbre de décision : est une structure de données qui représente les résultats potentiels d'une série de choix interconnectés. Il permet à une personne ou à une organisation d'évaluer différentes actions envisageables en fonction de leurs coûts, de leurs probabilités et de leurs bénéfices. L'arbre de décision commence généralement par un nœud initial à partir duquel plusieurs résultats possibles se développent. Chacun de ces résultats conduit à d'autres nœuds, générant ainsi d'autres possibilités. Elles permettent de prendre des décisions en utilisant une séquence de questions oui/non. Les arbres de décision sont couramment utilisés en apprentissage supervisé pour la classification et la régression. [73]
Les arbres de décision sont des classifieurs utilisés pour traiter des ensembles de données comprenant des attributs et leurs valeurs correspondantes. Un arbre de décision est composé des éléments suivants :[74]

- Nœud racine qui sert de point d'accès à l'arbre et effectue un test sur un attribut.
- Branches qui représentent les différentes valeurs de l'attribut testé dans le nœud parent. Feuilles (ce sont les nœuds terminaux de l'arbre) qui indiquent la classe résultante.
- Nœuds enfants, qui sont les descendants du nœud parent et effectuent d'autres tests sur les attributs.
- Feuilles, qui sont les nœuds terminaux de l'arbre et indiquent la classe résultante.

La figure.2.1. Représentent un exemple d'arbre de décision de la maladie :

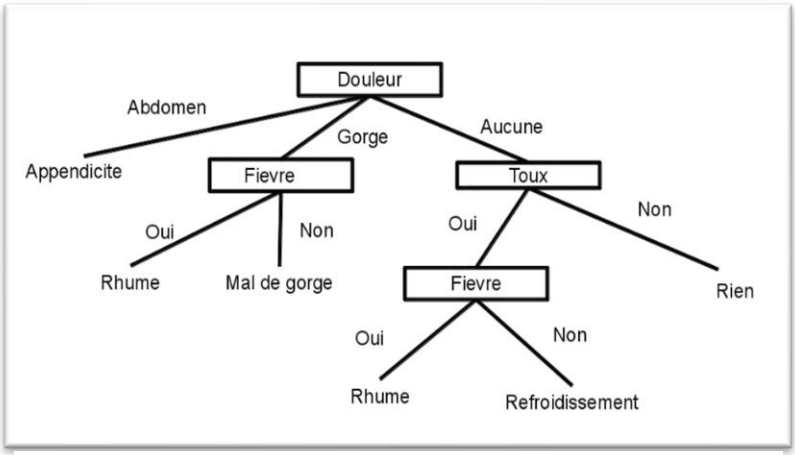


Figure II. 7: Arbre décision [75]

c. **Forêts aléatoires :** est une structure de données utilisée pour visualiser les résultats potentiels d'une série de choix interconnectés. Il offre la possibilité à une personne ou une organisation d'évaluer différentes actions envisageables en tenant compte de leurs coûts, probabilités et bénéfices respectifs. L'arbre de décision débute généralement par un nœud initial d'où se développent plusieurs résultats possibles. Chaque résultat mène à d'autres nœuds, générant ainsi de nouvelles possibilités. [76]

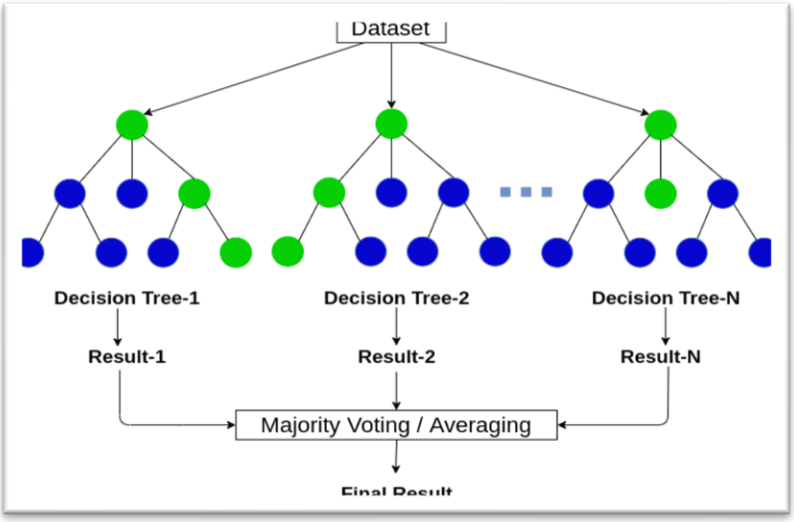


Fig.II.8. Forêt aléatoire [77]

d. **Réseaux de neurones :** Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires, les neurones formels. Les principaux réseaux se distinguent par l'organisation du graphe (en couches, complets. . .), c'est-à-dire leur architecture, son niveau de complexité (le nombre de neurones, présence ou non de boucles de rétroaction dans le réseau), par le type des neurones (leurs fonctions de transition ou d'activation) [78]. Les réseaux de neurones sont des modèles mathématiques qui s'inspirent du fonctionnement du cerveau humain. Les réseaux de neurones sont utilisés en apprentissage supervisé pour des tâches telles que la classification, la reconnaissance de formes et la prédiction.

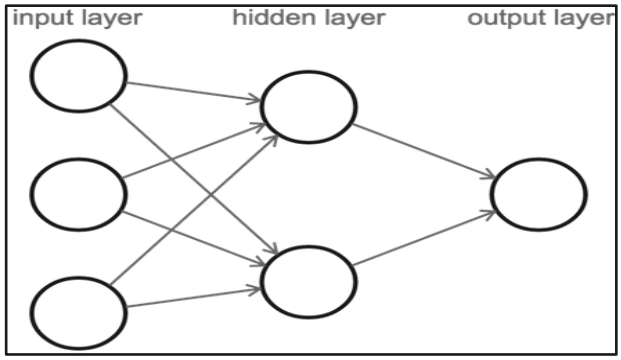


Figure II. 9 : Réseau de neurones [79]

e. SVM : Les SVM sont des algorithmes de Machine Learning utilisés pour la classification, la régression et la détection d'anomalies. Ils ont été développés dans les années 1990 par Vladimir Vapnik et Alexey Chervonenkis. Les SVM sont populaires pour des tâches telles que l'attribution de catégories, l'analyse des sentiments et la détection de spams. Ils utilisent des hyperplans pour séparer les données dans un espace de dimension supérieure, permettant une meilleure généralisation et prédiction. Les SVM sont appréciés pour leur capacité à gérer des données non linéaires grâce à l'utilisation de noyaux. [80]

SVM une méthode relativement récente de résolution de problèmes de classification (trier des individus en fonction de leurs caractéristiques) [81]

SVM effectuent un mappage des données vers un espace d'attributs de dimension supérieure afin de les classifier, même lorsque les données ne sont pas séparables de manière linéaire. Un séparateur entre les catégories est identifié, puis les données sont transformées pour représenter ce séparateur sous forme d'hyperplan. En utilisant les caractéristiques des nouvelles données transformées, il est possible de prédire le groupe auquel un nouvel enregistrement appartient. [82]

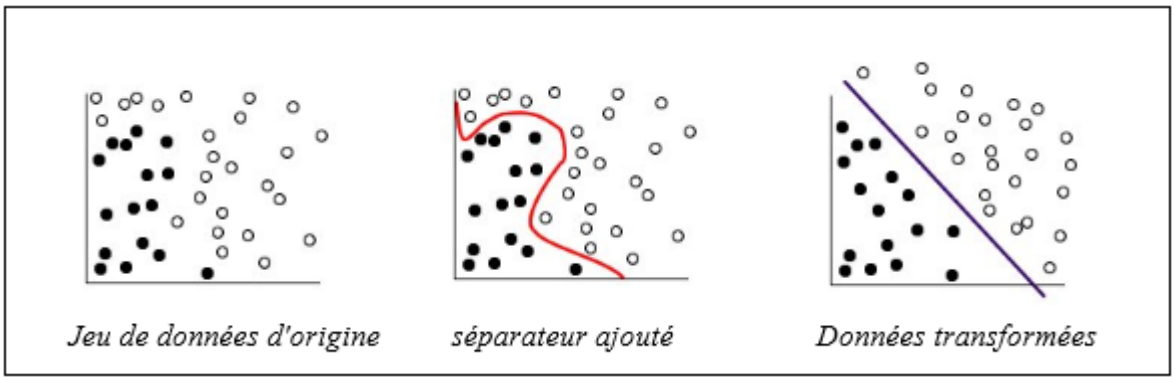


Figure II. 10 : SVM points de données

g. KNN (k-NN) : L'algorithme des k plus proches voisins est une méthode d'apprentissage supervisé non paramétrique utilisée pour la classification et la régression. Il se base sur la proximité des données pour effectuer des prédictions. En identifiant les k voisins les plus proches d'un point donné, il est capable d'estimer sa classe ou sa valeur. C'est une approche simple et flexible qui repose sur l'idée que des points similaires tendent à être regroupés. [83] KKN est basé sur une intuition simple et intuitive, ce qui en fait l'un des algorithmes de Machine Learning supervisé les plus simple : [84]

- **Étape 1** : Sélectionnez le nombre K de voisins
- **Étape 2** : Calculez la distance
- **Étape 3** : Prenez les K voisins les plus proches selon la distance calculée.
- **Étape 4** : Parmi ces K voisins, comptez le nombre de points appartenant à chaque catégorie.
- **Étape 5** : Attribuez le nouveau point à la catégorie la plus présente parmi ces K voisins ;
- **Étape 6** : Notre modèle est prêt.

4.4. Comparaison d'algorithme de l'apprentissage automatique :

Voici un tableau comparatif de quelques-uns des algorithmes d'apprentissage supervisé, ce tableau résume tous les algorithmes ci-dessus :

Tableau II. 1 : Comparaison algorithmes de l'apprentissage supervisé

Algorithme	Type	Avantages	Inconvénients
Régression linéaire	Régression	-Facile à comprendre et à mettre en œuvre, rapide à entraîner	-Fonctionne mieux avec des relations linéaires simples
Arbre de décision	Classification/ Régression	-Facile à comprendre et à visualiser. -Traiter des données mixtes (numériques et catégoriques). - Peut capturer des relations non linéaires.	- Sensible aux valeurs aberrantes. - Surajout (overfitting) de variables inutiles ou redondantes

Forêts aléatoires	Classification/ Régression	-Bonne précision, peut traiter des données mixtes (numériques et catégoriques). - capturer des relations non linéaires, résiste bien à la sur-adaptation.	- Coût en ressources. - Difficile à interpréter
SVM	Classification/ Régression	- Peut capturer des relations non linéaires. - résiste bien à la sur-adaptation	- Sensible aux choix des paramètres. - Difficulté à interpréter les résultats.
Réseau de neurones	Classification/ Régression	Peut capturer des relations non linéaires complexes. - résiste bien à la sur-adaptation.	- Complexité et difficulté de la conception -Requiert beaucoup de données.
KNN	Classification/ régression selon que la variable cible est discrète ou continue.	-facile à comprendre et à mettre en œuvre. -Il ne nécessite pas de phase d'apprentissage coûteuse, car il stocke simplement toutes les données d'entraînement. -KNN peut être utilisé pour des problèmes de classification multi-classe.	- lent et nécessite beaucoup de mémoire pour stocker toutes les données d'entraînement. - Le choix du paramètre K peut être difficile et avoir une grande influence sur les performances de l'algorithme.

5. Apprentissage non supervisé

L'apprentissage non supervisé est une discipline du machine learning qui se concentre sur l'analyse et la classification de données qui ne sont pas étiquetées. Les algorithmes utilisés dans ce domaine sont conçus pour découvrir des motifs ou des regroupements au sein des données, sans nécessiter une intervention humaine significative. [85]

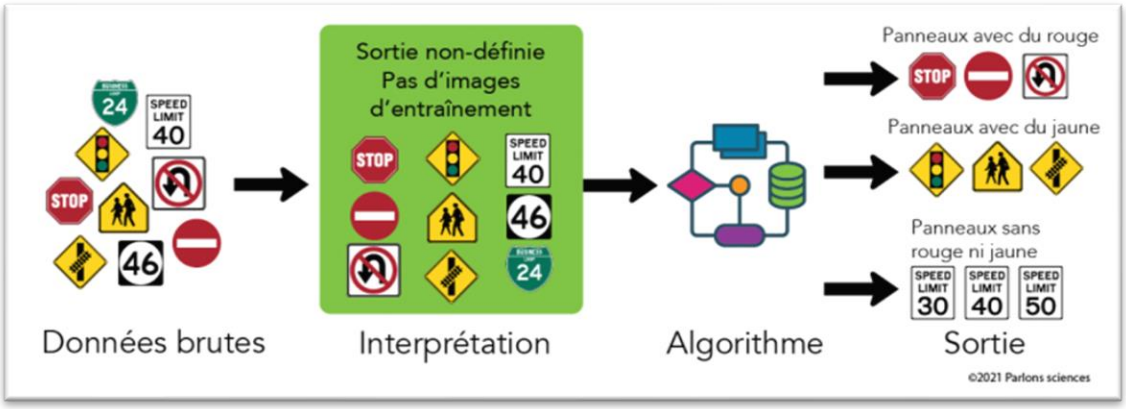


Figure II.12: Apprentissage non-supervisé. [86]

5.1. Utilisation d'apprentissage non supervisé :

a. Classement des données : cela implique de regrouper les données en fonction de leur similarité. Les algorithmes de clustering cherchent à trouver des groupes de données qui sont similaires les uns aux autres, tout en étant différents des autres groupes.

b. Calcul approximatif de la densité de distribution : cela consiste à estimer la densité de probabilité des données dans un espace donné. Cette technique est souvent utilisée pour identifier les zones de concentration de données, appelées "clusters" dans le cadre de l'analyse de cluster.

c. Réduction des dimensions : cela implique de réduire le nombre de variables dans les données en préservant autant d'informations que possible. Les algorithmes de réduction de dimensionnalité sont utilisés pour simplifier les données et faciliter leur analyse tout en réduisant les coûts de calcul et de stockage. [87]

5.2. Types Apprentissage non supervisé :

Y'a deux types de modèles **:**

a) Clustering :

Le clustering est une technique utilisée pour regrouper des objets au sein d'un ensemble de données, de manière précise et efficace. Ce processus est largement employé dans le domaine du marketing pour regrouper les clients en fonction de certaines caractéristiques. L'un des algorithmes les plus couramment utilisés pour le clustering est le K-means. Bien

que parfois complexe à appréhender pour les humains, cette méthode permet à la machine de regrouper les données de manière pertinente. [88]

Clustering consiste à regrouper des points de données similaires en fonction de certains critères de similarité.

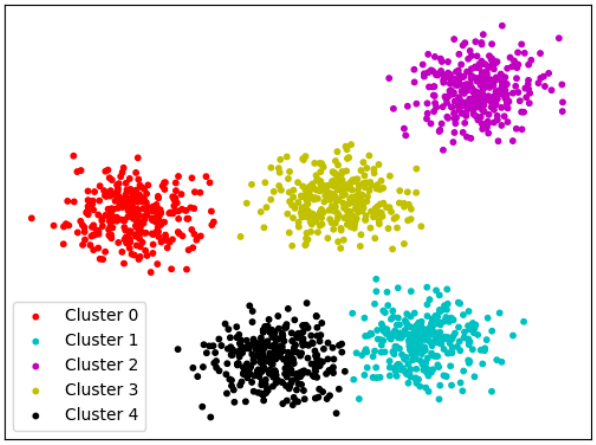


Figure II. 13: Clustering [89]

- **Types de clustering :**

Il existe plusieurs types de clustering, chacun avec ses propres méthodes et algorithmes. Voici quelques-uns des types de clustering les plus courants, Vous trouverez ci-dessous une brève discussion autour de trois approches courantes :

a. Le clustering hiérarchique : crée une arborescence de clusters. Il n'est pas surprenant que le clustering hiérarchique soit bien adapté aux données hiérarchiques, telles que les taxonomies. [90]

Le regroupement hiérarchique, également connu sous le nom de clustering, implique la création d'une structure arborescente de clusters pour représenter les données. À l'intérieur de cette hiérarchie, chaque groupe ou "nœud" est lié à deux groupes ou plus qui lui succèdent. Les groupes sont imbriqués les uns dans les autres et organisés sous la forme d'un arbre. Chaque nœud de l'arbre contient un ensemble de données similaires, et les nœuds sont regroupés en fonction de leurs similarités. [91]

Technique d'apprentissage automatique permettant de regrouper des chaînes de données par distance ou par similarité. [92]

C’est une Méthode de clustering non supervisée qui vise à diviser un ensemble de données en un certain nombre de groupes ou de clusters. Contrairement à l'algorithme K-means qui divise directement les données en K clusters prédéfinis, le clustering hiérarchique peut être utilisé pour trouver un nombre optimal de clusters en construisant une hiérarchie de clusters imbriqués. Méthode de classification non supervisée

rassemble un ensemble d'algorithmes d'apprentissage dont le but est de regrouper entre elles des données non étiquetées présentant des propriétés similaires. [93]

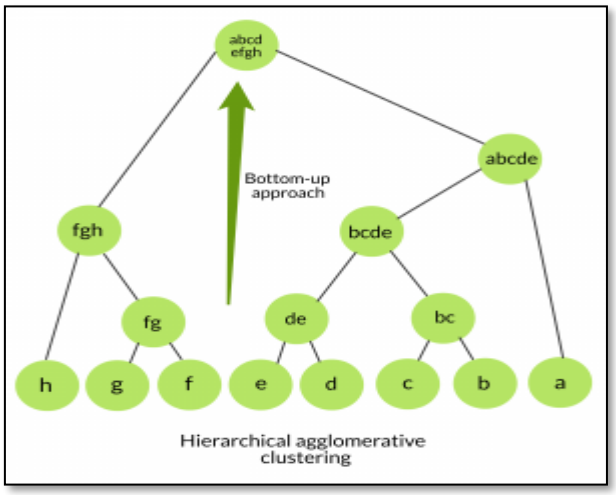


Figure II. 14: Clustering hiérarchique [94]

b. Clustering basé sur la densité : est une technique de regroupement de données qui repose sur la densité des points dans l'espace, technique de clustering efficace pour des ensembles de données complexes avec des formes de clusters arbitraires et du bruit.

Ce type de clustering utilise la densité plutôt que la distance pour regrouper les données. Dans cette approche, on considère qu'un point est dense s'il a un nombre de voisins supérieur à un seuil spécifié. De plus, deux points sont considérés comme voisins s'ils sont à une distance inférieure à une valeur prédéterminée. Ainsi, au lieu de se concentrer uniquement sur les distances entre les points, le clustering basé sur la densité prend en compte la densité locale des points pour former des clusters.

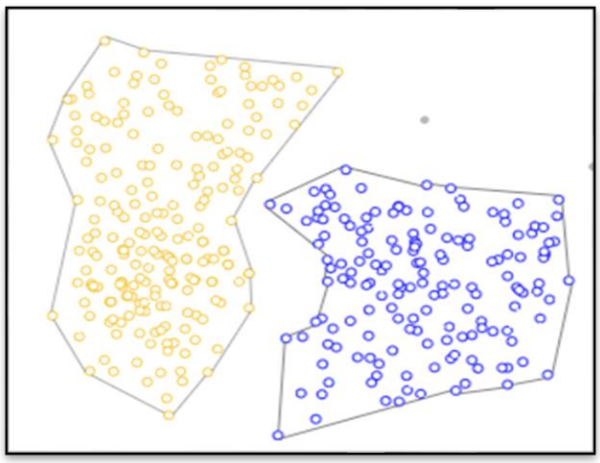


Figure II. 15 : Clustering basé sur la densité [95]

c. C**lustering par vecteurs de support (SVC) :** Le clustering par vecteurs de support (SVC) est une méthode de clustering dérivée de la méthode des machines à vecteurs supports (SVM). Elle regroupe les données dans des zones connues par la fonction de décision, en maximisant la marge entre les groupes de données. Cette méthode est adaptée aux ensembles de données complexes et non linéaires, car elle peut détecter des clusters de formes arbitraires. Cependant, les paramètres tels que le noyau et la régularisation peuvent influencer les résultats du clustering. En résumé, le SVC est une méthode de clustering efficace pour les ensembles de données complexes. [96]

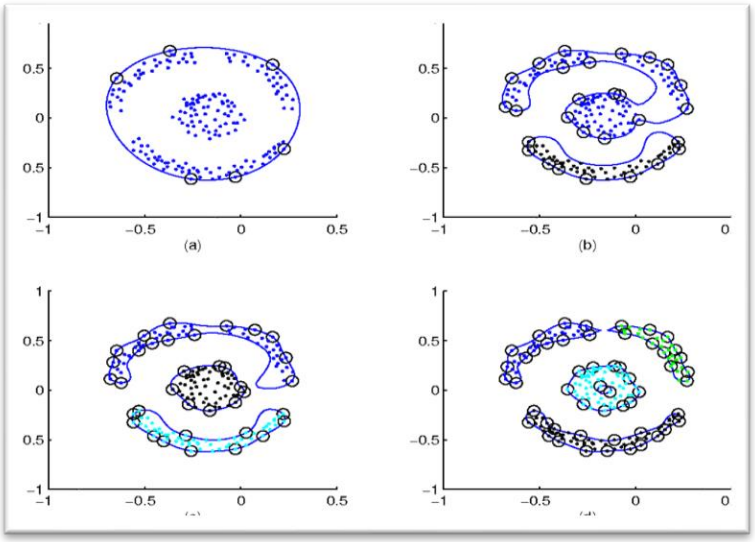


Figure II. 16: Clustering SVC [97]

b) **Règles d'association (Association Rules) :** méthode d'apprentissage non supervisé qui vise à identifier les relations entre les éléments d'un ensemble de données. Cette méthode est souvent utilisée pour l'analyse de paniers d'achat dans les magasins de détail, mais elle peut être appliquée à tout ensemble de données contenant des ensembles d'éléments.

L'objectif principal des règles d'association est d'identifier des règles implicites ou des modèles de corrélation entre les éléments d'un ensemble de données. Elle identifie les associations fréquentes « si-alors », qui constituent elles-mêmes les règles d'association. Une règle d'association comprend deux parties : une antécédente (si) et une conséquente (alors). [98]

5.3. Algorithmes d'apprentissage non supervisé :

a. **Méthode des K-means (k-moyennes) :** est un algorithme d'apprentissage non supervisé utilisé pour résoudre des problèmes de regroupement en apprentissage automatique ou en science des données. Dans ce sujet, nous allons apprendre ce qu'est l'algorithme de clustering K-Means, comment l'algorithme fonctionne, ainsi que la mise en

œuvre en Python du clustering K-Means **[99],** utilisé pour la segmentation de données, c'est-à-dire la division d'un ensemble de données en plusieurs groupes ou clusters homogènes. K-means est une méthode de clustering très courante, qui fonctionne en itérant pour assigner chaque point de données au cluster le plus proche en termes de distance euclidienne, puis en recalculant le centre de chaque cluster. C'est un exemple d'algorithme d'apprentissage non supervisé car il ne nécessite pas de données étiquetées ou de supervision humaine pour former le modèle **[100].**

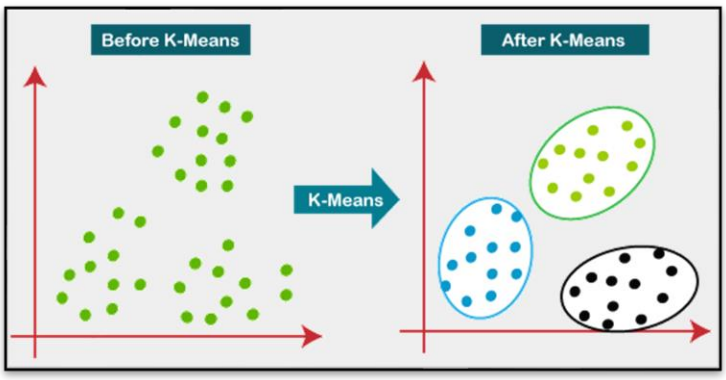


Figure II. 17: Algorithme K-moyennes [101]

b. **Algorithme DBSCAN [102] :** est un algorithme de clustering non supervisé largement utilisé. Son objectif est de diviser un ensemble de points en k groupes, appelés clusters, qui sont homogènes et compacts. Il utilise l'estimation de la densité locale pour définir ces clusters.
Dans le DBSCAN on utilise généralement la distance euclidienne, soient p = (p1,….,pn) et q = (q1,….,qn) :

$$d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_i-q_i)^2+\cdots+(p_n-q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i-q_i)^2}.$$

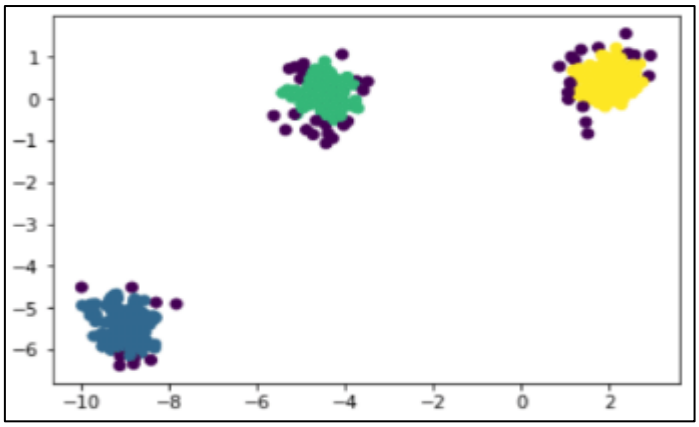


Figure II. 18: Algorithme DBSCAN [103]

c. **Algorithme Clustering hiérarchique :** un algorithme de clustering qui commence par considérer chaque point comme un cluster, puis fusionne les clusters les plus proches les uns des autres jusqu'à ce qu'un nombre prédéfini de clusters soit atteint. [104]

d. **Algorithme Apriori :** L'algorithme le plus couramment utilisé pour générer des règles d'association est l'algorithme Apriori, est une méthode populaire d'exploration de données pour découvrir des relations intéressantes entre des ensembles d'articles dans une base de données. Il est souvent utilisé pour l'analyse de paniers d'achat dans le commerce électronique ou pour la recommandation de produits dans des systèmes de filtrage collaboratif. [105]

Étape algorithme apriori [106]:

1-Prétraitement des données : doivent être nettoyées et transformées en un format approprié pour l'analyse.

2-Détermination du support minimum : est le seuil minimum de fréquence nécessaire.

3-Génération de l'ensemble de candidats : L'algorithme génère toutes les combinaisons possibles d'articles qui satisfont au support minimum.

4-Calcul du support : Le nombre de transactions contenant chaque ensemble de candidats est compté pour déterminer leur support.

5-Élimination des ensembles de candidats non pertinents : Les ensembles de candidats ayant un support inférieur au support minimum sont éliminés.

6-Génération de règles : Une règle d'association est une implication de la forme "Si A, alors B".

7-Calcul de la confiance : est la probabilité que B soit vrai, étant donné que A est vrai.

8-Évaluation des règles

5.4. Comparaison d'algorithme de l'apprentissage non supervisé :

Voici un tableau comparatif entre les algorithmes d'apprentissage non supervisé :

Tableau II. 2.Comparaison algorithmes de l'apprentissage non-supervisé

Algorithme	Type d'algorithme	Avantages	Inconvénients
K-means	Clustering de Partitionnement	- Efficace pour les grands ensembles de données. - Facile à comprendre et à mettre en œuvre	- La performance dépend de la sélection initiale des centres de cluster. - Ne convient pas aux données de forme irrégulière ou de densité variable
DBSCAN	Clustering de Densité	- Peut détecter des clusters de forme arbitraire - N'a pas besoin de spécifier le nombre de clusters à l'avance	- Ne convient pas aux ensembles de données de haute dimension - La performance dépend de la sélection de l'epsilon et de la densité minimum
Clustering hiérarchique	Clustering Hiérarchique	- Peut visualiser la structure hiérarchique des clusters	- N'est pas adapté aux ensembles de données volumineux en raison de sa complexité temporelle élevée
Apriori	Association de règles	- Peut découvrir des relations intéressantes entre des ensembles d'articles	- Peut produire un grand nombre de règles d'association, dont certaines peuvent être redondantes ou non significatives

6. Différence entre l'apprentissage supervisé et l'apprentissage non supervisé :

L'apprentissage supervisé et l'apprentissage non supervisé sont deux approches distinctes de l'apprentissage automatique. Dans l'apprentissage supervisé, les données d'entraînement sont étiquetées, ce qui permet au modèle d'apprendre à partir des exemples avec des réponses connues. L'objectif est de prédire ou de classer de nouvelles données en fonction de ces étiquettes. En revanche, dans l'apprentissage non supervisé, les données d'entraînement ne sont

pas étiquetées, et le modèle cherche à découvrir des structures ou des modèles cachés dans les données. [107]

Ce tableau résumé les principales différences entre l'apprentissage supervisé et l'apprentissage non supervisé :

Tableau II. 3 : Différence entre apprentissage supervisé et apprentissage non supervisé [107]

	Apprentissage supervisé	**Apprentissage non supervisé**
Objectif	Prédire les résultats basés sur de nouvelles données	Obtenir de nouveaux insights à partir de grandes quantités de nouvelles données
Données d'entrée	Données avec des étiquettes ou des résultats connus	Données sans étiquettes ou résultats connus
Algorithme principal	Régression, classification, arbres de décision, réseaux neuronaux	Clustering, réduction de dimensionnalité, règles d'association
Supervision	Nécessite des données de sortie étiquetées pour l'entraînement	Nécessite uniquement des données d'entrée pour l'entraînement
Prédiction	Est capable de prédire de nouveaux résultats	Ne fait pas de prédiction explicite, se concentre sur l'identification de motifs
Connaissance a priori	L'utilisateur connaît les résultats attendus	L'utilisateur espère découvrir de nouvelles informations
Évaluation	Mesure de l'erreur de prédiction sur un ensemble de données de test.	Mesure de la qualité de la structure de clustering ou de la réduction de dimensionnalité.

Avantages	Bonne performance de prédiction pour les tâches supervisées.	Découverte de structures ou de relations cachées entre les données.
Limitations	Dépendance à la qualité et à la quantité des données d'entraînement étiquetées.	Sensible aux valeurs aberrantes et aux données bruyantes.

7. Apprentissage automatique et la prédiction de la DDSH

a) Apprentissage supervisé : consiste à entraîner un modèle à partir d'un ensemble de données d'entraînement qui contient à la fois des données d'entrée (caractéristiques du patient) et des données de sortie. Le modèle est ensuite utilisé pour prédire la **DDS** pour de nouveaux patients.

b) Apprentissage non supervisé : ne nécessite pas de données de sortie pour l'entraînement du modèle. Au lieu de cela, il utilise des algorithmes de regroupement pour identifier des modèles dans les données d'entrée. Les algorithmes d'apprentissage non supervisé couramment utilisés pour ce type de tâche incluent la classification ascendante hiérarchique, la méthode k-moyennes.

c) Importance de l'apprentissage supervisé dans la prédiction de la DDSH :

L'apprentissage supervisé est essentiel pour la prédiction de la **DDSH** car il permet d'obtenir des estimations précises en utilisant des données d'entraînement comprenant des exemples de durées de séjour antérieures et d'autres variables pertinentes. Cela aide les professionnels de la santé à planifier les ressources, à fournir des soins adaptés et à améliorer l'efficacité des opérations hospitalières. [108]

7.1. Rôle de l'apprentissage automatique dans la prédiction de la DDSH :

- **Capacité à analyser de grandes quantités** de données complexes et à identifier des modèles ou des relations qui ne sont pas facilement perceptibles par les méthodes traditionnelles d'analyse statistique. [109]
- **Identification des facteurs prédictifs :** peut aider à identifier les facteurs qui influencent la DDSH. Par exemple, il peut révéler que certaines conditions médicales, certains résultats de tests ou certains médicaments sont associés à une durée de séjour plus longue. Cette connaissance peut être utilisée pour optimiser les protocoles de traitement et les décisions cliniques [109].

- **Scalabilité des processus d'analyse de données :** L'un des principaux avantages du Machine Learning réside dans sa capacité à traiter des ensembles de données vastes et complexes de manière efficace. L'application d'algorithmes de Machine Learning à l'analyse de données à grande échelle peut mettre en évidence des modèles et des relations que les analystes humains pourraient ne pas déceler. Cette capacité permet d'utiliser les données pour prendre des décisions et effectuer des prévisions avec une précision accrue. [109]
- **Modélisation des relations complexes :** Les facteurs qui influencent la **DDSH** sont souvent multiples et interdépendants. Les modèles d'apprentissage automatique peuvent prendre en compte de nombreuses variables simultanément et apprendre les relations complexes entre elles. Ils peuvent détecter des interactions non linéaires ou des modèles subtils qui pourraient être négligés par les méthodes traditionnelles de modélisation. [110]

8. Conclusion :

En conclusion, l'IA et le Machine Learning jouent un rôle important dans la prédiction de la durée de séjour hospitalier, offrant des performances prometteuses. Ces modèles de prédiction peuvent aider les cliniciens à planifier les soins, optimiser les flux de patients et identifier les patients à haut risque de rester plus longtemps à l'hôpital. Cependant, il est essentiel de procéder à une évaluation rigoureuse des modèles et de les intégrer efficacement dans les pratiques cliniques pour garantir leur utilité et leur précision. En continuant à développer et à améliorer ces modèles, nous pourrons mieux comprendre et gérer la durée de séjour hospitalier, ce qui pourrait avoir un impact significatif sur les soins de santé.

CHAPITRE III

Etat de l'art

de la prédiction de la DDSH

1. Introduction

Dans ce chapitre, nous examinons l'état de l'art de la prédiction de la durée de séjour hospitalier, en nous intéressant aux ensembles de données utilisés, aux travaux connexes ainsi qu'aux méthodes classiques et intelligentes employées dans ce domaine. La prédiction de la durée de séjour joue un rôle crucial dans la planification des ressources et l'amélioration des soins médicaux. Les avancées récentes dans les techniques d'apprentissage automatique ont ouvert de nouvelles perspectives pour développer des modèles prédictifs précis.

2. Prédiction de la DDSH

Revêt une importance primordiale à plusieurs égards. Tout d'abord, cela permet d'optimiser l'utilisation des ressources hospitalières en planifiant les admissions, les lits et le personnel en fonction des besoins prévus. Ensuite, cela contribue à améliorer la gestion des soins en permettant une coordination efficace entre les différents départements et professionnels de santé. De plus, en prévoyant **DDS**, les patients et leurs proches peuvent mieux se préparer et s'organiser, favorisant ainsi une expérience plus positive. Enfin, ces prédictions aident également à évaluer les risques potentiels de complications et à mettre en place des mesures préventives adéquates, ce qui peut réduire les coûts de santé et améliorer les résultats cliniques. [111]

3.Méthodes de prédiction de la DDSH :

Les méthodes classiques de prédiction de la **DDSH** se basent sur des modèles statistiques qui peuvent être limités par la complexité des données de santé et la variabilité des facteurs de risque. En revanche, l'intelligence artificielle permet d'analyser de grandes quantités de données en temps réel et de détecter des modèles complexes, ce qui peut améliorer la précision de la prédiction.

3.1. Méthodes classiques pour la prédiction de la DDSH :

Ces méthodes impliquent l'utilisation de modèles statistiques pour prédire la **DDS** d'un patient en se basant sur les données cliniques disponibles telles que l'âge, le sexe, le diagnostic, les antécédents médicaux, les traitements précédents, les examens de laboratoire et d'imagerie…etc.

Les méthodes classiques pour la prédiction de la **DDSH** peuvent varier selon les établissements de santé et les pays, mais voici quelques-unes des méthodes les plus couramment utilisées : [112]

A. **Les méthodes basées sur les règles :** ces méthodes utilisent des règles cliniques prédéfinies pour prédire la durée de séjour à l'hôpital en fonction du diagnostic, des

Traitements et des antécédents médicaux du patient. Ces règles peuvent être établies par les médecins ou les professionnels de la santé en fonction de leur expérience clinique et de leur expertise.

B. **Les méthodes basées sur l'expertise humaine :** ces méthodes impliquent la prise de décision clinique par les professionnels de la santé en se basant sur leur expérience et leur expertise. Les médecins et les professionnels de la santé peuvent utiliser leur jugement clinique pour prédire la **DDS** à l'hôpital en fonction de l'état de santé du patient et de leur expérience passée.

Il est important de noter que ces méthodes classiques ont des limites et peuvent ne pas être suffisamment précises pour prédire la **DDS** à l'hôpital. C'est pourquoi l'utilisation de l'intelligence artificielle pour la prédiction de la durée de séjour à l'hôpital est de plus en plus étudiée et considérée comme une alternative prometteuse.

3.2. Méthodes intelligentes pour la prédiction de la DDSH :

Les méthodes intelligentes pour la prédiction de la **DDS** à l'hôpital impliquent l'utilisation de (**IA**) pour prédire la **DDS** d'un patient en se basant sur les données cliniques disponibles. Voici quelques-unes des méthodes utilisées : [113]

A. **Réseaux de neurones artificiels :** sont une méthode intelligente couramment utilisée pour prédire la DDSH. Ils utilisent des modèles d'apprentissage automatique inspirés du fonctionnement du cerveau humain. En se basant sur les caractéristiques des patients, les RNA apprennent à trouver les relations complexes entre ces caractéristiques et la DDS. Une fois entraînés, ils peuvent prédire la DDS pour de nouveaux patients. Les RNA sont capables de capturer des relations non linéaires et sont puissants pour la prédiction de la DDS, mais nécessitent suffisamment de données d'entraînement et une expertise en modélisation. Cependant, avec une mise en œuvre adéquate, les RNA peuvent fournir des prédictions précises et améliorer la prise de décision en gestion des séjours hospitaliers. [114]

B. **Arbres de décision :** sont une méthode intelligente utilisée pour prédire la DDSH. Ils utilisent une structure d'arbre pour prendre des décisions basées sur les caractéristiques des patients. Les arbres de décision sont simples et interprétables, ce qui les rend utiles pour comprendre les facteurs qui influencent la DDS. Ils peuvent être utilisés seuls ou en combinaison avec d'autres techniques pour améliorer les prédictions. En résumé, les arbres de décision sont une approche intuitive et interprétable pour prédire la DDSH. [115]

C. **Forêts aléatoires :** sont une méthode intelligente couramment utilisée pour prédire la DDSH. Elles combinent les prédictions de plusieurs arbres de décision pour obtenir une prédiction finale plus précise. Les FA capturent les relations complexes entre les caractéristiques des patients et la DDS, sont capables de gérer de grandes quantités de données, et fournissent des mesures d'importance des variables. Elles offrent une précision élevée et une interprétabilité, ce qui en fait une méthode efficace pour la prédiction de la DDSH. [116]

D. **SVM :** Sont une méthode intelligente couramment utilisée pour prédire la DDSH. Les SVR sont branche de SVM des algorithmes d'apprentissage supervisé efficaces pour la classification et la régression. Le SVR fonctionne en cherchant à trouver une fonction qui s'ajuste au mieux aux données d'entraînement tout en minimisant l'erreur de prédiction. Il utilise des vecteurs de support pour déterminer cette fonction et construire une frontière de décision optimale. [117]

E. **Extrême Gradient Boosting Model :** (XGboost) est considéré comme l'un des meilleurs dans le domaine de l'apprentissage supervisé. Il s'agit d'une version améliorée du GBM qui intègre des paramètres pour améliorer le modèle de prédiction. Sa grande efficacité réside dans sa capacité à tirer parti du parallélisme offert par tous les micro-processeurs de la machine, ce qui réduit considérablement le temps de calcul. De plus, il intègre une technique de régularisation qui prévient le sur-apprentissage, rendant ainsi le modèle généralisable et évitant les erreurs dues à une sur-spécialisation. Un autre avantage de cet algorithme est sa capacité à traiter les données manquantes et à capturer les relations non linéaires entre les différentes variables. Enfin, il utilise le calcul de la deuxième dérivée dans la réduction du taux d'erreur de la fonction de perte, ce qui permet de guider plus efficacement la direction du gradient lors de l'optimisation du modèle. [111]

4. Etapes de modélisation de prédiction de la DDSH basées sur machine Learning :

Description du problème : La prédiction précise de la **DDSH** est essentielle pour les professionnels de la santé et les établissements médicaux. Elle permet une meilleure planification des traitements, une allocation efficace des ressources et un contrôle des coûts plus efficace. Des séjours hospitaliers plus courts réduisent les coûts pour les patients, les compagnies d'assurance et les établissements médicaux, tout en améliorant la qualité des soins et la satisfaction des patients.

- Les étapes de modélisation de prédiction de la Durée De Séjour Hospitalier (DDSH) basées sur la machine learning sont généralement les suivantes [118] :

a) Collecte des données : La première étape consiste à collecter les données pertinentes pour la prédiction de la DDSH. Cela peut inclure des informations sur les patients (âge, sexe, antécédents médicaux, etc.), les diagnostics, les procédures médicales, les médicaments administrés, les résultats des tests, etc. Ces données peuvent être extraites à partir des dossiers médicaux électroniques ou d'autres sources de données médicales.

b) Prétraitement des données : Une fois les données collectées, elles doivent être prétraitées pour les rendre adaptées à l'analyse par les modèles de machine learning. Cela peut impliquer le nettoyage des données en supprimant les valeurs manquantes ou aberrantes, la normalisation des variables, la transformation des données catégoriques en variables numériques, etc.

c) Division des données : Les données sont divisées en ensembles d'apprentissage et de test. L'ensemble d'apprentissage est utilisé pour entraîner le modèle de prédiction, tandis que l'ensemble de test est utilisé pour évaluer les performances du modèle.

d) Choix du modèle de machine Learning : Différents modèles de machine learning peuvent être utilisés pour prédire la DDSH, tels que les réseaux de neurones, les arbres de décision, les méthodes ensemblistes (comme les forêts aléatoires et les boosters), ou les méthodes de régression. Le choix du modèle dépend des caractéristiques des données et des objectifs spécifiques de prédiction.

e) Entraînement du modèle : Dans cette étape, le modèle de machine learning est entraîné sur l'ensemble d'apprentissage en utilisant les caractéristiques sélectionnées. Le modèle apprend à partir des données pour établir une relation entre les caractéristiques et la DDSH.

f) Évaluation du modèle : Une fois le modèle entraîné, il est évalué sur l'ensemble de test pour évaluer ses performances de prédiction. Cela peut être mesuré à l'aide de différentes métriques telles que l'exactitude, la précision, le rappel, le score F1…etc.

5. DataSets utilisés dans la prédiction de la durée de séjour hospitalier :

A. MIMIC-III « Centre d'Information Médicale pour les Soins Intensifs » : est une grande base de données à centre unique comprenant des informations relatives aux patients admis dans les unités de soins intensifs d'un grand hôpital de soins tertiaires. Les données comprennent les signes vitaux, les médicaments, les mesures de laboratoire, les observations et les notes consignées par les fournisseurs de soins, l'équilibre hydrique, les codes de procédure, les codes de diagnostic, les rapports d'imagerie, la durée du séjour à l'hôpital, les données de survie et plus encore. [119]

B. **Base de données de recherche collaborative eICU** : est une base de données multi centres comprenant des données de santé désidentifiées associées à plus de 200 000 admissions dans des unités de soins intensifs à travers les États-Unis entre 2014 et 2015. La base de données comprend des mesures de signes vitaux, de la documentation de plan de soins, des mesures de gravité de la maladie, des informations de diagnostic et des informations de traitement. Les données sont collectées via le programme Philips eICU, un programme de télésanté de soins intensifs qui fournit des informations aux soignants au chevet des patients. [120]

C. **APACHE-IV (Évaluation de la Physiologie Aiguë et de la Santé Chronique-IV) :** est un système de scoring clinique utilisé pour prédire la mortalité et la durée de séjour des patients admis en soins intensifs. Ce système utilise des données cliniques, telles que des mesures de signes vitaux, des résultats de laboratoire et des antécédents médicaux, pour attribuer un score à chaque patient en fonction de sa gravité de maladie. [121]

D. **Investissements dans le domaine de la santé et durée du séjour à l'hôpital :** Les données ont été obtenues à partir de la base de données de l'OCDE. La recherche se limite aux pays de l'OCDE où toutes les données pour la période 1990-2018 sont disponibles simultanément dans la base de données. [122]

Voici un tableau comparatif entre les jeux de données (DataSet) utilisé dans ce domaine :

Tableau III. 1 : Comparaison entre les jeux de données (DataSet)

Jeu de donnée	Taille	Source	Année	Class
MIMIC-III [119]	Ensemble de données volumineux et complexe, plus de 40 000 patients dans des unités de soins intensifs	Boston, Massachusetts Etats-Unis *archives des systèmes d'information sur les soins intensifs. *bases de données des dossiers de santé	2001-2012	Urgent Elective Emergency

		électroniques des hôpitaux. *Fichier principal des décès de l'administration de la sécurité sociale. **Disponible**		
eICU **[120]**	Plus de 200 000 patients	Base de données de patients hospitalisées États-Unis	2014-2015	N'est pas structurée avec une classe spécifique pour la prédiction de la durée de séjour hospitalier.
APACHE-IV [121]	Système de classification et de prédiction pour évaluer la gravité des patients et prédire leur risque de décès.	Base de données de patients hospitalisées **Pas disponible**	Publié 2006	risk of death length of stay
Investissements dans le domaine de la santé et durée du séjour à l'hôpital [122]	Compte tenu des données médicales sur différents pays **518** entries	OECD DataSet **Disponible**	1990-2018	/

Durée de séjour à l'hôpital avec Microsoft R Server[123]	Ensemble de donnée **100000** points de données de patient admis	Microsoft open source **Disponible**	2012-2016	le nombre réel de jours de DDS
Data-driven Approach to Predict Hospital Length of Stay [124]	Un modèle prédictif pour la durée du séjour (LOS) basé sur un grand ensemble de données, en utilisant des indicateurs couramment disponibles lors du processus d'hospitalisation. En se basant sur une approche de régressionn. **15253 Patient**	A Portuguese Case Study **Pas disponible**	2000-2013	le nombre réel de jours de DDS

6. Travaux connexes de la prédiction de la DDSH :

On peut les appeler des études ou des travaux antérieurs car ce sont des recherches menées précédemment sur le même sujet (prédiction de la DDSH en utilisant des techniques de Machine Learning) et qui ont été publiées dans des revues scientifiques.

A. **Utilisation de modèles d'apprentissage automatique pour prédire la durée de séjour dans un environnement hospitalier :** Dans cette étude comparative, nous avons utilisé les algorithmes Random Forest (RF) et Gradient Boosting (GB) sur le jeu de données "Lenghtofstay.csv" de Microsoft. Après prétraitement des données, le RF a démontré une meilleure performance en termes de (MAE) par rapport au GBM, soulignant ainsi sa supériorité dans la prédiction de la durée du séjour. [125]

B. **Prédiction de la durée de séjour et de la mortalité des patients :** Vise à améliorer les performances des hôpitaux en réduisant les décès des patients en unités de soins intensifs

(USI). Les chercheurs utilisent des méthodes d'apprentissage automatique telles que (DT), (RF), (ANN), les réseaux bayésiens et (SVM) pour prédire des résultats mesurables tels que le risque de complications, la mortalité et la durée du séjour à l'hôpital. Ils se basent sur le jeu de données MIMIC II. Le document fournit une revue complète des différentes méthodes utilisées pour prédire la mortalité et la durée de séjour en soins intensifs, ainsi que des facteurs qui influencent ces prédictions. Il souligne également les défis et les lacunes dans ce domaine et conclut qu'il n'y a pas de technique fiable unique pour la prédiction. [126]

C. **Utilisation des techniques d'exploration de données pour déterminer et prédire la durée de séjour des patients cardiaques :** Utilise des techniques d'exploration de données pour prédire la DDS des patients atteints de maladies coronariennes. Les chercheurs ont collecté les données de 4 948 patients et utilisé 36 variables d'entrée avec trois algorithmes de classification : l'arbre de décision, les machines à vecteurs de support (SVM) et (ANN). L'algorithme SVM s'est révélé le plus précis avec une précision de 96,4%. Les patients souffrant de troubles pulmonaires ou respiratoires et d'hypertension artérielle avaient tendance à avoir des séjours plus longs. Les données ont été stockées dans une base de données Microsoft SQL Server et divisées en ensembles d'entraînement (80%) et de test (20%). **[127]**

D. **Prédiction de la durée de séjour à l'hôpital à l'aide de réseaux neuronaux :** Les chercheurs ont utilisé le package neuralnet en R pour entraîner un modèle basé sur les caractéristiques démographiques telles que l'âge, le sexe, l'état matrimonial, l'ethnicité et la religion. Les données utilisées provenaient de patients de l'unité de soins intensifs (USI) de la base de données MIMIC III. Les résultats ont montré que le modèle était capable de prédire avec précision, environ 79% du temps, si un séjour serait court (≤ 5 jours) ou long (> 5 jours) après le départ du patient de l'USI. Cette approche de prédiction offre une solution plus générale par rapport aux études antérieures. [128]

E. **Prédiction de la durée de séjour du patient au moment de son admission à l'aide de l'apprentissage automatique** : ce travail est pour la prédiction de la durée de séjour des patients à l'admission dans le service à partir des données d'admission aux urgences semble réalisable en utilisant des algorithmes d'apprentissage automatique. De plus, les résultats montrent que différents modèles d'apprentissage automatique peuvent être utilisés pour cette tâche, et que le groupe de patients utilisé dans ce jeu de données était assez large, ce qui suggère un potentiel pour utiliser ces algorithmes dans des paramètres plus généralisés dans les hôpitaux. [130]

F. L'identification d'une durée de séjour prolongée pour les patients subissant une intervention chirurgicale : Une étude a développé un modèle de prédiction de la durée de séjour prolongée chez les patients subissant une opération en utilisant des techniques d'apprentissage supervisé. Les résultats ont montré que la méthode de forêt aléatoire était le modèle de prédiction le plus précis et le plus stable. Les auteurs ont souligné l'utilité des techniques d'apprentissage supervisé pour les pronostics des patients et l'amélioration de la sécurité des patients. Cependant, des recherches supplémentaires sont nécessaires pour valider ces résultats. [129]

G. Application d'un modèle de réseau neuronal BP pour prédire la durée du séjour à l'hôpital : Cette étude utilise une approche de fouille de données basée sur les réseaux de neurones de rétropropagation (BP) pour créer un modèle de prédiction de la durée de séjour à l'hôpital pour les patients atteints de cholécystite. Le modèle atteint une précision d'environ 80%. L'étude souligne l'importance de la durée de séjour pour évaluer les coûts médicaux et l'utilisation des ressources, et suggère que la fouille de données peut être un outil précieux pour identifier les facteurs potentiels de la durée de séjour. [135]

Tableau III. 2: Tableau des travaux connexes

Nom de l'article	Année	Modèles	Méthode	DataSet	Résultat
[125]	April 2020	FA & RG	Apprentissage automatique « ML »	Lenghtofstay.csv Microsoft	Le MAE obtenu par la RF est meilleur que celui obtenu par le GBM.
[126]	2017 Mar 22	AD FA RB SVM	Apprentissage automatique « ML »	MIMIC II	/
[127]		AD SVM RNA	Apprentissage automatique « ML »	Microsoft structured query language	-SVM Précision : 96.4%

	JUIN (2013)			(MS-SQL) database. Train : 80% Test : 20%	-ANN Précision : 84.5% DT précision : 83.5%
[128]	**2017**	RNA	Apprentissage automatique « ML »	Mimic III Train : 90% Test : 10%	Performance de modèle d'environ 80%
[129]	**2015 - 2016**	SVM AD FA	Apprentissage automatique « ML »	Collecte de 913 dossiers complets de patients hospitalisés ayant subi des interventions chirurgicales entre janvier 2006 et décembre 2012 dans le sud de Taiwan.	FA est plus précis et le plus stable avec 0.877 SVM=0.816 AD=0.758
[130]	**2019**	FA. SVM. RNA. AD. RG model (GBM.)	Apprentissage automatique ML	N'est pas disponible Train :70% Test : 30%	FA avec précision = 0.75
Un modèle d'attention profonde pour prévoir la durée	**2021**	Réseau de neurones à attention	Deep Learning	MIMIC III Train :70% Test : 30%	-HAN-LoS (AUROC): 0.82

du séjour et la mortalité à l'hôpital dès l'admission à partir des codes ICD et des données démographiques [131]		hiérarchique (HAN)			-Micro-F1 score of 0.24 -HAN-Mor achieved AUROC of 0.87
Prédiction de la durée du séjour à l'hôpital au stade de l'admission pour les patients en cardiologie à l'aide d'un réseau neuronal artificiel [132]	**2016**	Régression linéaire (RL) Réseaux de neurones artificiels (RNA)	Apprentissage automatique ML	Données cliniques et administratives ont été obtenues pour les patients cardiaques hospitalisés entre le 1er octobre 2010 et le 31 décembre 2011 dans un hôpital de Taipei, Taiwan	- Précision : 88,07% à 89.95%. 1<MAE<1.11 0.44<MRE<0.47
Comparaison des algorithmes d'apprentissage automatique dans la prédiction des patients hospitalisés atteints de	**2020**	AD AdaBoost FA RB SVM k-NN	Apprentissage automatique ML	11 884 dossiers d'admission électroniques correspondant à 6 933 patients présentant différents	Meilleur résultat : précision Forêt aléatoire= 72.7% AUC :79.6%

schizophrénie. [135]				troubles de santé mentale. Ces dossiers appartiennent aux unités de soins aigus de 11 hôpitaux publics d'une région d'Espagne.	Precision: 72.8% F1-Score: 72.7%. Recall: 72.7%,
Prédiction de la durée de séjour en unité de soins intensifs d'un hôpital en utilisant des caractéristiques générales à l'admission [134]	**2021**	RB SVM KNN	Apprentissage automatique ML	Données collectées à partir des dossiers médicaux de l'unité de soins intensifs (USI) de l'hôpital spécialisé de l'Université du canal de Suez en Égypte. Période de deux ans, de septembre 2015 à septembre 2017	SVM précision : 69% KNN précision : 57%

Application d'un modèle de réseau neuronal BP pour prédire la durée du séjour à l'hôpital [135]	**2010**	Réseau de neurones à propagation arrière (BP)	Machine learning	Données comprenaient 921 patients d'un hôpital chinois entre 2003 et 2007.sélection nées à partir de la base de données Oracle du Système d'Information Hospitalier	Précision d'environ 80%

D'après cette étude, il a été observé que les algorithmes d'apprentissage automatique, en particulier ceux basés sur l'apprentissage supervisé, sont couramment utilisés pour prédire la durée de séjour à l'hôpital.

7. Défis :

Défis de la prédiction de la durée de séjour hospitalier avec l'intelligence artificielle :

- Collecte et intégration des données provenant de différentes sources.
- Variabilité des facteurs influençant la durée de séjour.
- Modélisation des interactions complexes entre les variables.
- Intégration des modèles d'IA dans les workflows cliniques.

8. Synthèse :

Synthèse de la prédiction de la durée de séjour hospitalier avec l'intelligence artificielle :

- Amélioration de la planification des traitements et de l'allocation des ressources.
- Utilisation de l'IA pour fournir des prédictions précises.
- Amélioration des données et de l'intégration de variables avancées.
- Contribution à une meilleure gestion des soins, à une utilisation efficace des ressources et à une qualité accrue des services médicaux.

9. Conclusion :

En conclusion, ce chapitre a fourni un aperçu de l'état de l'art de la prédiction de la durée de séjour hospitalier, en se concentrant sur les ensembles de données utilisés, les travaux connexes et les méthodes classiques et intelligentes employées. Les avancées récentes dans les techniques d'apprentissage automatique ont ouvert de nouvelles perspectives pour la prédiction précise de la durée de séjour. Toutefois, il est important de considérer les caractéristiques et les limitations spécifiques de chaque méthode, ainsi que la disponibilité de données de haute qualité. Cette revue de la littérature nous permettra de mieux comprendre les approches existantes et de poser les bases de notre propre travail de recherche sur la prédiction de la durée de séjour hospitalier, en combinant les avantages des méthodes classiques et intelligentes pour obtenir des résultats précis et fiables.

CHAPITRE IV

REALISATION

d'un modèle de prédiction de la DDSH

1. Introduction

Ce chapitre se concentre sur la réalisation et l'implémentation de notre système de prédiction de la durée de séjour hospitalier. Nous aborderons les différents aspects techniques de notre projet, y compris le matériel utilisé, les outils de développement, le choix de l'algorithme et du fonctionnement, ainsi que le jeu de données sélectionné. Nous décrirons également l'architecture des étapes de modélisation pour la prédiction et discuterons des résultats obtenus, en les comparant à d'autres travaux similaires..

2. Matériels utilisés :

Le développement du modèle est réalisé via ordinateur portable ayant les caractéristiques suivantes :

Marque	**DELL Inspiron1525**
Processeur	Intel Core™ 2 Duo avec CPU IntelT1600 (1.66GHz)
RAM	4GO
Disque dur	500 GO
Système d'exploitation	Microsoft Windows 10 professionnel

3. Outils de développement:

3.1 Langages utilisés :

a. Python : est un langage de programmation interprété open source, polyvalent et convivial, propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels. Il est largement utilisé dans le domaine du développement logiciel, de l'analyse de données, de l'apprentissage automatique (machine learning) et de l'intelligence artificielle. Python se distingue par sa syntaxe claire et concise, ce qui facilite la lecture et l'écriture du code. Il offre également une vaste bibliothèque standard et de nombreuses bibliothèques tierces spécialisées qui facilitent le développement d'applications et la résolution de problèmes variés. Grâce à sa popularité et à sa communauté active, Python est devenu l'un des langages de programmation les plus couramment utilisés et est apprécié pour sa flexibilité et sa facilité d'utilisation. [136]

Figure IV. 1 Python [137]

3.2. Produit de Google Recherche :

a. **Google colab :** est un environnement de développement basé sur le cloud développé par Google Recherche. Il offre un accès gratuit à des ressources de calcul, y compris des processeurs graphiques, permettant aux utilisateurs d'exécuter du code Python via leur navigateur. Colab est largement utilisé dans le domaine du Machine Learning, de l'analyse de données et de l'éducation, offrant un environnement pratique et collaboratif pour l'exécution de projets Python, l'exploration de données, l'entraînement de modèles de machine Learning, et bien plus encore. [138]

4.Choix et fonctionnement de l'algorithme :

a. **Choix de l'algorithme :** Les algorithmes **SVM(SVR**), **RL(MLPR), FA** et **AD** jouent un rôle important dans la prédiction de la DDSH. Les **SVR** sont efficaces pour gérer des ensembles de données complexes, tandis que **MLRP** est une technique de régression linéaire multivariée utilisée pour analyser les relations entre plusieurs variables indépendantes et une variable dépendante dans des données de panel. Les **FA** combinent les prédictions de plusieurs arbres de décision pour obtenir des résultats plus précis, tandis que les **AD** offrent une interopérabilité en identifiant les facteurs importants. Chaque algorithme présente ses avantages spécifiques et le choix dépendra des caractéristiques des données et des objectifs de prédiction. En utilisant ces algorithmes, les professionnels de la santé peuvent améliorer la précision des prédictions de DDS et ainsi soutenir la prise de décision clinique et la gestion des ressources hospitalières.

b. **Fonctionnement de l'algorithme :** le fonctionnement simplifié des algorithmes **SVM(SVR), AD, FA et ANN(MLPR)** dans un tableau :

Tableau IV. 1: Fonctionnement des méthodes

Algorithme	Initialisation	Construction du modèle	Prédiction
SVM (SVR)	Sélection des paramètres.	Entraînement sur les données d'apprentissage en trouvant l'hyperplan optimal.	Classification des nouvelles instances en fonction de leur position par rapport à l'hyperplan.
Arbre de Décision	Construction de l'arbre à partir des données d'apprentissage.	Séparation des données en fonction des caractéristiques pour former un arbre.	Prédiction en suivant les chemins de l'arbre basée sur les caractéristiques des nouvelles instances.
Forêt Aléatoire	Sélection des paramètres.	Construction d'un ensemble d'arbres de décision aléatoires.	Agrégation des prédictions de chaque arbre pour obtenir une prédiction finale.
RL (MLPR)	Initialisation des poids et des biais.	Entraînement sur les données d'apprentissage en ajustant les poids et les biais.	Prédiction en propageant les entrées à travers le réseau pour obtenir une sortie.

5. DataSet utilisé dans notre travail :

5.1 Description de DataSet :

Le jeu de données "**LengthOfStay**" a été rendu open source par Microsoft. Il comprend des informations sur 100 000 patients hospitalisés, telles que des indicateurs sur leur état de santé et la durée de leur séjour à l'hôpital. Ces données sont précieuses pour prédire la durée du séjour à l'hôpital des patients. Le fichier "LengthOfStay.csv" contient les détails spécifiques de chaque patient dans cet ensemble de données. [139]

Tableau IV. 2: Description DataSet [14]

Index	Data Field	Type	Description
01	eid	Entier	Identifiant unique de l'admission à l'hôpital.
02	vdate	String	Date de visite
03	rcount	Entier	Nombre de réadmissions au cours des 180 derniers jours
04	gender	String	Genre du patient - M or F
05	dialysisrenalendstage	String	Indicateur de maladie rénale pendant l'épisode de soins
06	asthma	String	Indicateur d'asthme pendant l'épisode de soins
07	irondef	String	Indicateur de carence en fer pendant l'épisode de soins
08	pneum	String	Indicateur de pneumonie pendant l'épisode de soins
09	substancedependence	String	Indicateur de dépendance à une substance pendant l'épisode de soins
10	psychologicaldisordermajor	String	Indicateur de trouble psychologique majeur pendant l'épisode de soins
11	depress	String	Indicateur de dépression pendant l'épisode de soins
12	psychother	String	Indicateur d'autres troubles psychologiques pendant l'épisode de soins
13	fibrosisandother	String	Indicateur de fibrose pendant l'épisode de soins
14	malnutrition	String	Indicateur de malnutrition pendant l'épisode de soins
15	hemo	String	Indicateur de trouble sanguin pendant l'épisode de soins
16	hematocrit	Float	Valeur moyenne de l'hématocrite pendant l'épisode de soins (g/dL)
17	neutrophils	Float	Valeur moyenne des neutrophiles pendant l'épisode de soins (cellules/µL)
18	sodium	Float	Valeur moyenne du sodium pendant l'épisode de soins (mmol/L)

19	glucose	Float	Valeur moyenne de la glycémie pendant l'épisode de soins (mmol/L)
20	bloodureanitro	Float	Valeur moyenne de l'azote uréique sanguin pendant l'épisode de soins (mg/dL)
21	creatinine	Float	Valeur moyenne de la créatinine pendant l'épisode de soins (mg/dL)
22	bmi	Float	Indice de masse corporelle moyen pendant l'épisode de soins (kg/m^2)
23	pulse	Float	Pouls moyen pendant l'épisode de soins (battements/min)
24	respiration	Float	Respiration moyenne pendant l'épisode de soins (respirations/min)
25	secondarydiagnosisnonicd9	Entier	Indicateur indiquant si un diagnostic non formaté en ICD-9 a été codé comme un diagnostic secondaire
26	discharged	String	Date de sortie
27	facid	Entier	Identifiant de l'établissement où l'épisode de soins s'est produit
28	lengthofstay	Entier	Durée du séjour pour l'épisode de soins

6.Étendue de l'étude et Prédiction de la DDSH

Dans le cadre de notre thème, la mesure de la **DDSH** sera calculée en termes de nombre de jours. Cette approche de mesure en jours offre une perspective plus pratique et adaptée à notre réalité clinique, car la **DDSH** est généralement évaluée et gérée en fonction des jours complets d'hospitalisation.

En utilisant cette échelle de mesure en jours, il nous devient plus facile de planifier nos ressources hospitalières, d'organiser nos rendez-vous médicaux et nos procédures, ainsi que de coordonner les soins post-hospitaliers.

7. Architecture des étapes de modélisation du la prédiction de la durée de séjour hospitalier

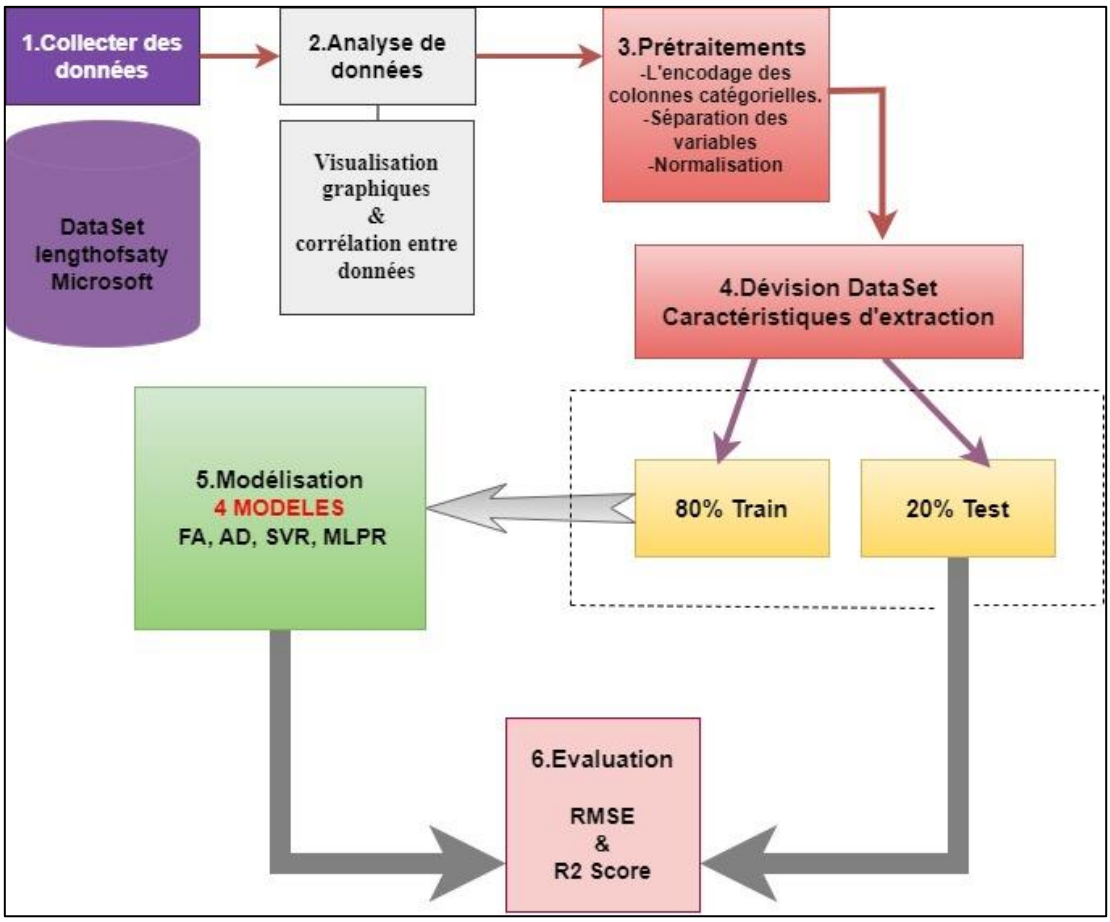


Figure IV.2. Architecture des étapes de prédiction de la DDSH

8.Étapes de modélisation de la prédiction de la DDSH

La prédiction des DDSH est calculée selon les étapes suivantes. Une première étape permet de collecter les données nécessaires issues du DataSet "lenghtofstay.csv". Les données sont récupérées à partir de ce fichier contenant les informations sur les séjours hospitaliers. Ensuite, les étapes d'analyse des données, de nettoyage, de sélection des variables, de prétraitement des données et de modélisation sont réalisées. Ces étapes permettent de construire des modèles de prédiction de la DDS basés sur les caractéristiques des patients. Les modèles ainsi développés sont évalués pour leur précision et leur performance afin d'améliorer la prise de décision en matière de gestion des séjours hospitaliers.

8.1 Collecte des données :

Les données nécessaires sont collectées à partir du DataSet "lenghtofstay.csv" qui contient les informations sur les séjours hospitaliers.

8.2. Analyse et nettoyage des données :

Les données sont analysées pour comprendre leur structure, les types de variables présentes et identifier d'éventuelles erreurs ou valeurs manquantes. Les données sont nettoyées en traitant les valeurs manquantes, en supprimant les enregistrements erronés ou incomplets, et en corrigeant les erreurs éventuelles.

- Afficher nombre d'occurrences de chaque valeur unique dans chaque colonne de DataSet.
- **Sélection des variables :** Les variables pertinentes pour la prédiction de la DDS sont sélectionnées en fonction de leur importance et de leur relation avec la variable cible.

  Visualisation des données :

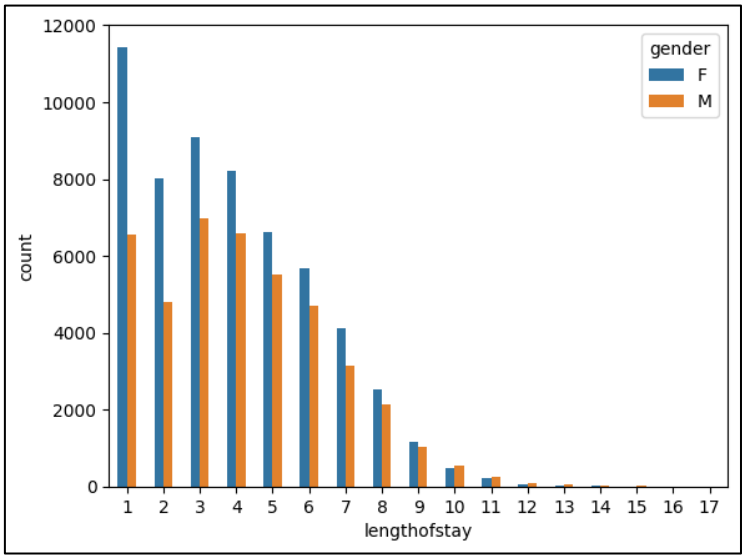


Figure IV.3. Durée de séjour et sexe

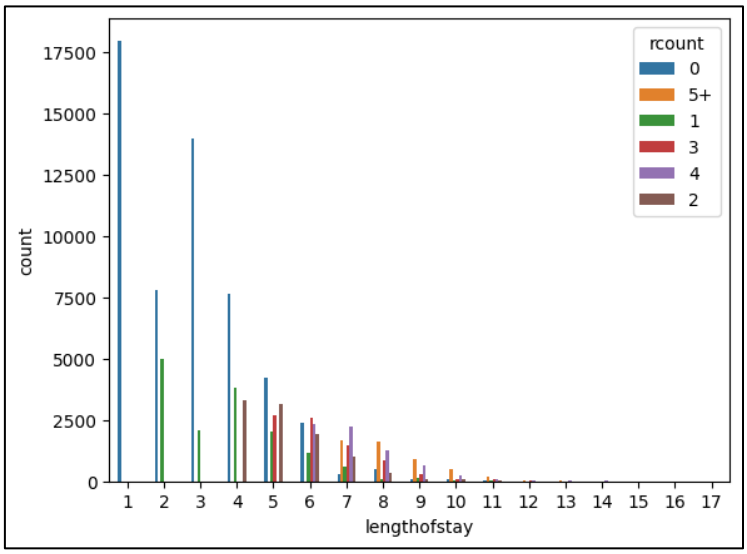


Figure IV.4. Durée de séjour et réadmission

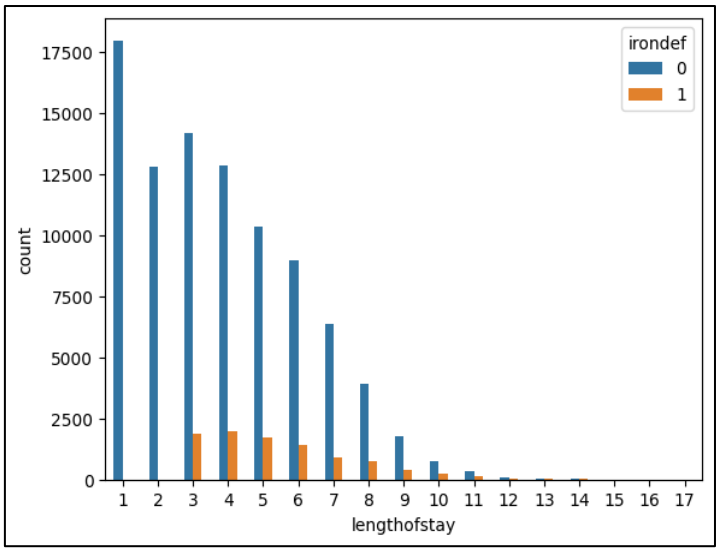
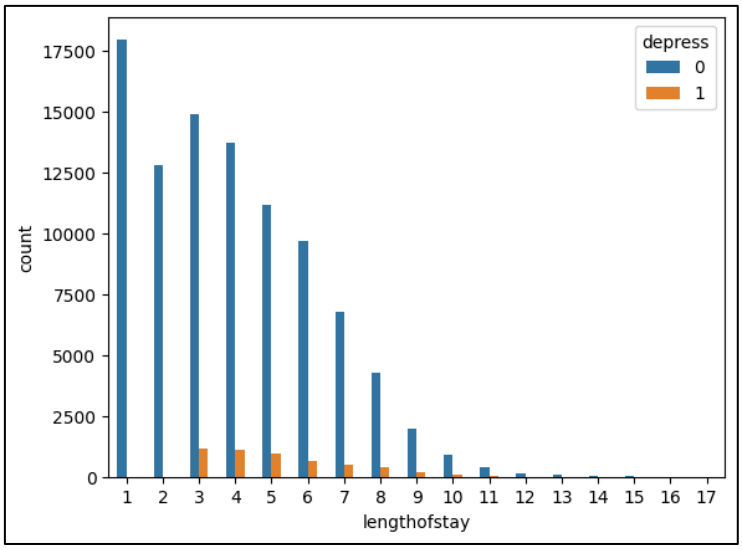


Figure IV.5. Durée de séjour et fer à repasser



Figures IV.6. Durée de séjour et dépression

- **Afficher les corrélations entre toutes les variables du DataSet**

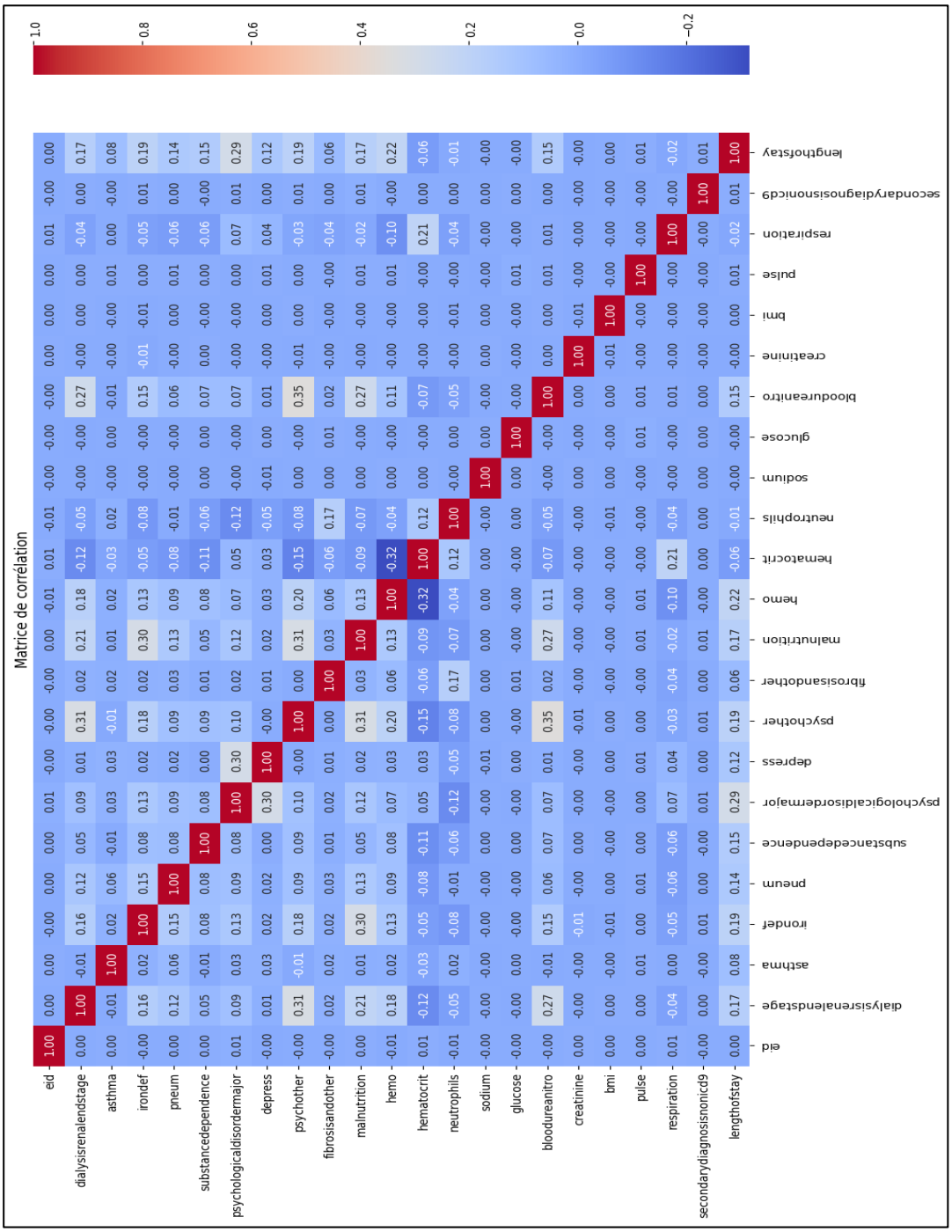


Figure IV.7. Corrélation entre données de DataSet

8.3 Prétraitement des données :

Les données sont prétraitées en effectuant des transformations telles que la normalisation, la discrétisation ou l'encodage des variables catégorielles, afin de les rendre adaptées aux algorithmes de prédiction.

- Gérer des colonnes.

- L'encodage des colonnes catégorielles spécifiées dans DataSet entrées, pour créer nouvelle colonne binaire et représente chaque valeur unique
- Séparation les variables cibles des variables explicatives de DataSet.
- Normalisation des variables numérique à l'aide de la classe « **standardscaler** » pour mettre toutes les variables à la même échelle commune et faciliter apprentissage des modèles.
- Pour améliorer le modèle prédictif de la durée de séjour, nous supprimons la date de sortie et la date de visite car elles reflètent directement la durée de séjour, et nous voulons que le modèle apprenne à partir d'autres facteurs plus subtils. En excluant ces caractéristiques, le modèle est encouragé à découvrir des variables supplémentaires qui influencent la durée de séjour, dans le but d'une prédiction plus complète au-delà des dates de visite et de sortie.
- Une fonctionnalité créée appelée "total number of issues" est obtenue en additionnant les colonnes binaires telles que : dialysisrenalendstage, asthma, irondef, pneum, substancedependence, psychologicaldisordermajor, depress, psychother, fibrosisandother, malnutrition et hemo. Ces colonnes représentent des attributs "oui" ou "non", avec 1 indiquant la présence et 0 indiquant l'absence. La fonctionnalité du "total number of issues" condense ces informations en une seule mesure, capturant le compte global de problèmes pour chaque point de données, simplifiant le modèle et fournissant des informations sur l'ampleur des problèmes présents, tout en réduisant la dimensionnalité des données.
- **Entraînement du modèle :** Les modèles sont entraînés sur un sous-ensemble des données, en utilisant une partie pour l'entrainement et une autre pour le test
- 80% pour l'entrainement(train).
- 20% pour le test.
- Random-state=42 : garantit que chaque fois que nous exécutons le code, la division de « train et test » sera le même.
- Tracer l'histogramme.

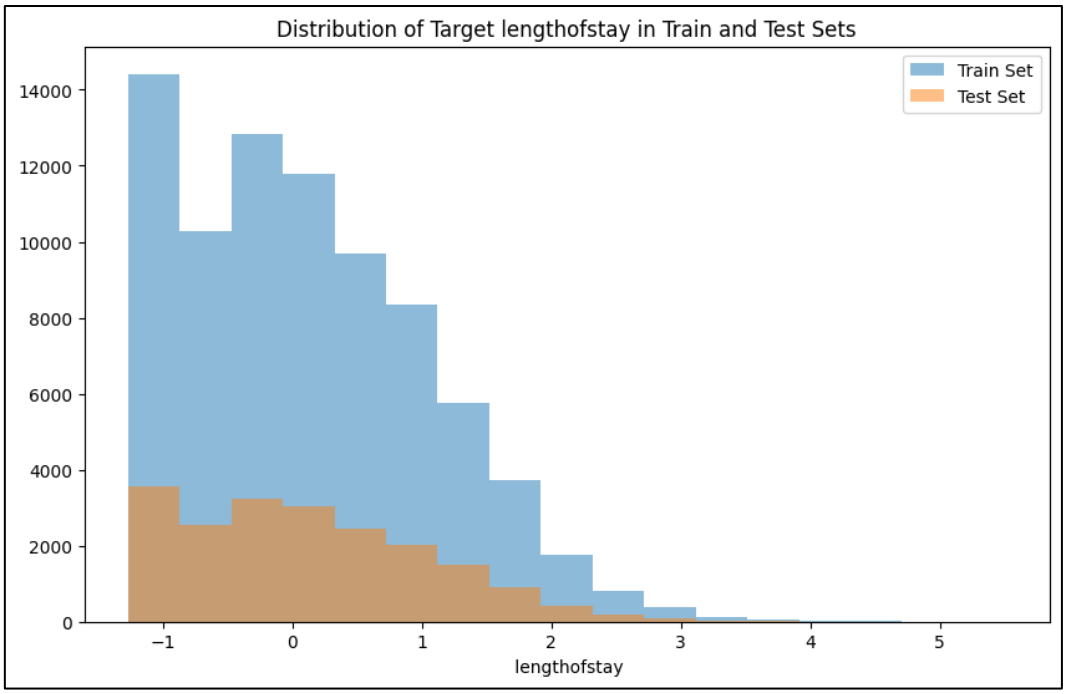


Figure IV.8. Histogramme Test-Train

8.4 Modélisation :

Différents algorithmes de prédiction, tels que les Régression Linéaire (MLPR), les Machines à Vecteurs de Support (SVM-SVR-), les Arbres de Décision ou les Forêts Aléatoires, sont utilisés pour construire des modèles de prédiction de la DDSH.

- Entraine les modèles sur les données de l'entrainement.
- Faites des prédictions sur l'ensemble test.
- affichage les performances de chaque modèle :

8.5 Évaluation du modèle :

Les modèles sont évalués en utilisant des mesures de performance telles que l'erreur moyenne, le coefficient de détermination (R^2) ou la RMSE. Cela permet de sélectionner le modèle le plus performant.

```
RMSE: 0.24104568820838113
R2 score: 0.9410094988417209
RMSE: 0.34805893453859293
R2 score: 0.8770045844850834
RMSE: 0.3013873310207522
R2 score: 0.9077782548474896
RMSE: 0.26146128343914615
R2 score: 0.9305938221857224
```

Figure IV.9. Performances des modèles

Meilleur résultat avec l'algorithme **foret aléatoire** de la performance :

- ✓ **RMSE : 0.24**
- ✓ **R^2 : 0.94**

Le tableau suivant affiche les résultats de performance obtenue pour chaque modèle :

Tableau.IV.3. Performance des modèles.

Modèle	RMSE	R^2
Forêt Aléatoire	**0.24***	**0.94***
Arbre Décision	0.34	0.87
SVR	0.30	0.90
MLPR	0.26	0.93

La figure.IV.7. affiche les résultats de chaque modèle avec son RMSE. Un RMSE plus faible indique une meilleure performance, car cela signifie que les prédictions du modèle sont plus proches des vraies valeurs.

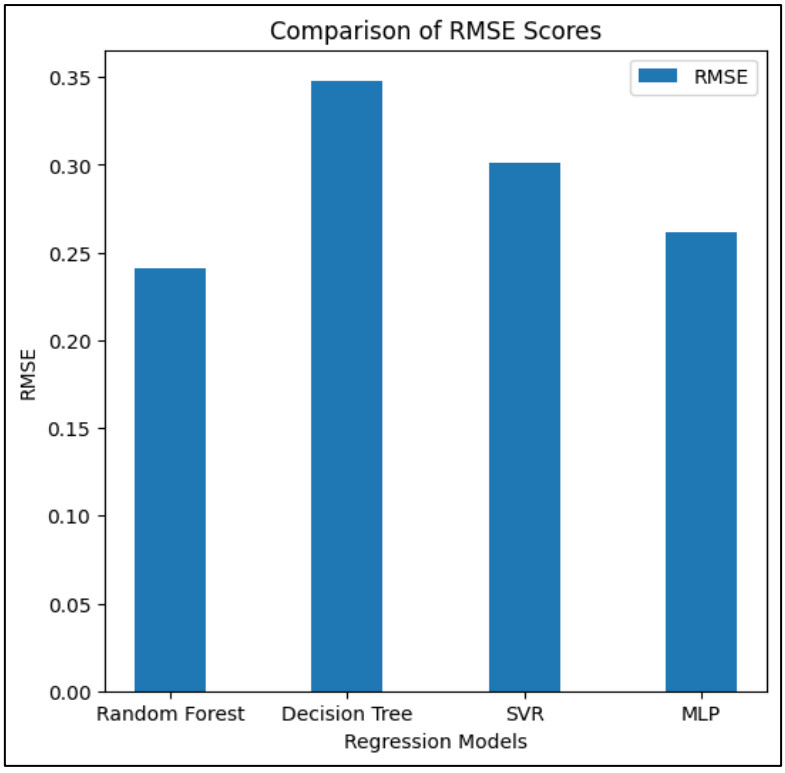


Figure IV.10. Comparaison « RMSE » des modèles

La figure.IV.8. affiche les résultats de chaque modèle avec son R^2. Un R^2 élevé indique que le modèle est capable d'expliquer une grande partie de la variance de la variable cible, ce qui suggère une bonne adéquation du modèle aux données. Cela peut être interprété comme une performance solide du modèle dans la capture des variations de la variable cible.

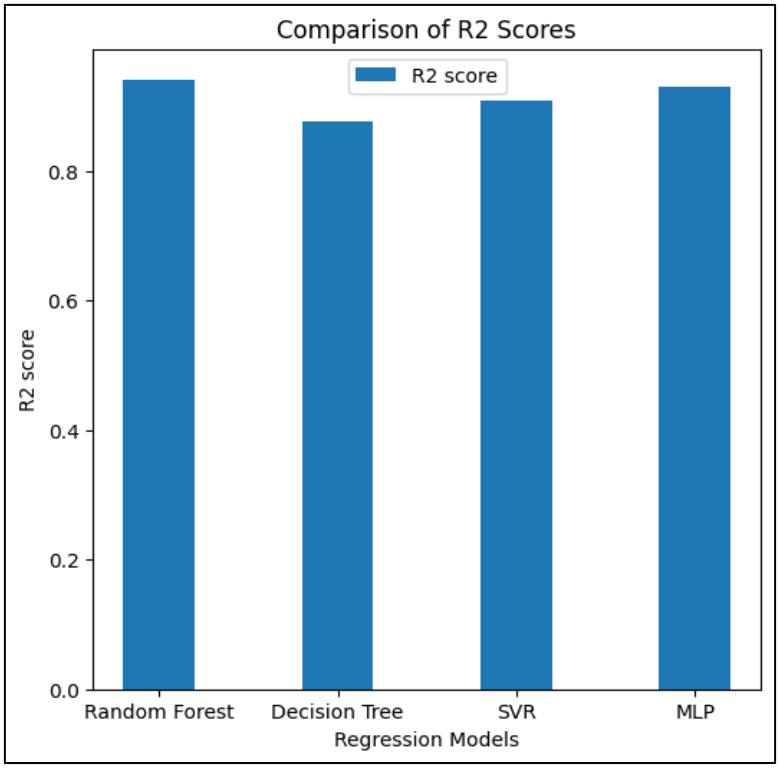


Figure IV.11. Comparaison « R^2 Score » des modèles

9. Discussion et comparaison :

9.1 Comparaison les résultat avec des travaux utilisent même DataSet avec notre travail

Dans la discussion finale de notre travail, il est nécessaire de parler des travaux connexes à notre étude, en particulier ceux qui utilisent le même **ensemble de données**. Nous pouvons discuter des résultats obtenus dans ces travaux, ainsi que des résultats que nous avons obtenus dans notre propre étude en utilisant différents modèles.

Dans notre étude, nous avons utilisé quatre modèles différents : **MLPR**, **AD**, **FA** et le **SVR**. Parmi ces modèles, la **forêt aléatoire** a donné les meilleurs résultats, avec une **performance** de **RMSE** de **0,24** et un **R2** Score de **0,94**. Cela indique que notre modèle de forêt aléatoire a réussi à prédire les valeurs cibles de manière très précise.

Comparativement, le travail **[125]** utilisait deux modèles, la **forêt aléatoire** et le **gradient Boosting**, pour analyser le **même ensemble de données**. Ils ont obtenu un **R2 Score** de **0,92** et une erreur quadratique moyenne (**MAE**) de **0,45**. Bien que ces résultats soient également prometteurs, ils sont légèrement inférieurs à ceux de notre étude.

En conclusion, notre étude a démontré que le modèle de forêt aléatoire était le plus performant parmi les quatre modèles que nous avons utilisés. Ces résultats sont comparables à ceux d'autres travaux utilisant des modèles similaires. Cependant, il est important de poursuivre les recherches pour affiner davantage les performances des modèles et comprendre les facteurs qui influent sur leurs résultats.

Voici un tableau comparatif entre notre travail et l'autre travail, en se basant sur le DataSet utilisé, les modèles utilisés, les mesures évaluatives utilisées et les résultats obtenus :

Tableau.IV.4. Comparaison les résultat des travaux utilisent même DataSet avec notre travail

	DataSet utilisé	**Modèles utilisés**	**Mesures évaluatives**	**Résultats**
Notre travail	LengthOfStay Microsoft	MLPR Arbre Décision Foret Aléatoire SVR	RMSE R2 Score	Meilleur résultat avec Foret Aléatoire RMSE : 0.24 **R2 Score : 0.94***
[125]	LengthOfStay Microsoft	Foret Aléatoire Gradient Boosting	R2 Score MAE	R2 Score : 0.92 MAE : 0.45

9.2. Comparaison avec des autres travaux qui utilisent des différentes DataSets

Le tableau ci-dessous présente une comparaison entre les travaux connexes et mon travail, en termes de précision des prédictions de la durée de séjour hospitalier. Les résultats sont résumés dans ce tableau pour permettre une analyse comparative des performances des différents modèles et approches utilisés.

Tableau.IV.5. Comparaison résultats des travaux utilisent différentes DataSets

	DataSet utilisé	Modèles utilisés	Résultats
Notre travail	LengthOfStay Microsoft	Foret Aléatoire Arbre Décision SVR MLPR	Foret Aléatoire (RMSE= 0.24, R2 Score=0.94) Arbre décision (RMSE= 0.34, R2 Score=0.87) SVR (RMSE= 0.30, R2 Score= 0.90) MLPR (RMSE=0.26 , R2 Score=0.93)
[127]	MIMIC III	Arbre Décision SVM RNA	SVM Précision : 0.964 ANN Précision : 0.845 DT précision : 0.835
[129]	Collecte de 913 dossiers complets de patients Taiwan	Foret Aléatoire SVM Arbre Décision	FA est plus précis et le plus stable avec 0.877 SVM = 0.816 AD = 0.758
[130]	N'est pas disponible	FA. SVM. RNA. AD. RG model (GBM.)	Foret Aléatoire avec précision de 0,75
[134]	Données collecter a partir (USI) de l'hôpital spécialisé de l'Université du canal de Suez en Égypte.	RB SVM KNN	SVM précision : 0.69 KNN précision : 0.57

10. Conclusion :

En conclusion du ce chapitre, nous avons utilisé les outils puissants du langage Python et de Google Colab pour mener notre étude. Nous avons travaillé avec un ensemble de données spécifique et avons suivi plusieurs étapes clés pour prédire la DDSH grâce à l'intelligence artificielle. Notre architecture de contribution, illustrée par un schéma, a démontré notre approche méthodique et notre compréhension approfondie du problème. Les captures d'écran du code ont permis d'apporter une vision concrète de notre travail. Grâce à cette combinaison d'outils et de méthodes, nous sommes en mesure d'apporter des prédictions précises sur la durée de séjour hospitalier, ouvrant ainsi la voie à de futures améliorations dans le domaine de la santé.

Conclusion générale

Conclusion générale et perspectives :

Ce projet a exploré l'application de l'intelligence artificielle et de l'apprentissage automatique dans la prédiction de la durée de séjour hospitalier. En utilisant le DataSet "LenghtOfStay.CSV" fourni par Microsoft, nous avons pu étudier les possibilités offertes par ces techniques avancées pour améliorer la gestion hospitalière.

L'objectif principal de ce projet était de développer des modèles prédictifs précis et fiables pour estimer la durée de séjour des patients hospitalisés. À travers les différentes étapes de collecte des données, d'exploration, de préparation et de modélisation, nous avons pu analyser les caractéristiques des patients et identifier les facteurs qui influencent la durée de séjour.

Grâce à l'utilisation de l'intelligence artificielle et de l'apprentissage automatique, nous avons pu développer des modèles prédictifs performants. Ces modèles ont démontré leur capacité à prédire avec précision la durée de séjour des patients, ce qui peut être d'une grande utilité pour la planification des ressources hospitalières, l'optimisation des soins et la gestion des coûts.

Ce projet souligne l'importance croissante de l'intelligence artificielle dans le domaine de la santé, en particulier pour la prédiction et la gestion des durées de séjour hospitalier. Les résultats obtenus mettent en évidence le potentiel de ces techniques pour améliorer l'efficacité des soins, réduire les coûts et améliorer l'expérience des patients.

Cependant, il est important de souligner que ces modèles prédictifs doivent être utilisés avec précaution et en complément de l'expertise médicale. Ils ne doivent pas remplacer le jugement clinique des professionnels de la santé, mais plutôt servir d'outil d'aide à la décision.

En conclusion, ce projet de mémoire a contribué à l'avancement de la prédiction de la durée de séjour hospitalier en exploitant les techniques de l'intelligence artificielle et de l'apprentissage automatique. Il ouvre la voie à de futures recherches et applications dans le domaine de la gestion hospitalière et de l'amélioration des soins de santé.

Perspectives :

Développement de modèles spécifiques à certaines pathologies : La prédiction de la durée de séjour peut varier selon les pathologies spécifiques. Les futures études pourraient se concentrer sur le développement de modèles prédictifs spécifiques à certaines pathologies, telles que les maladies cardiaques, les affections respiratoires, les cancers, etc. Cela permettrait d'obtenir des prédictions plus précises et adaptées à chaque situation médicale.

Exploration de nouvelles variables : Bien que nous ayons utilisé un ensemble de données complet, il existe toujours des variables potentielles qui pourraient influencer la durée de séjour des patients et n'ont pas été incluses dans notre modèle. Les futures recherches pourraient se

concentrer sur l'exploration de nouvelles variables, telles que les antécédents médicaux détaillés, les résultats de tests spécifiques, les comorbidités, etc., afin d'améliorer la précision des modèles prédictifs.

Bibliographie

BIBLIOGRAPHIE

[1] https://apps.who.int/gb/bd/PDF/bd47/FR/constitution-fr.pdf

[2]https://www.choleraalliance.org/sites/g/files/dvc3616/files/styles/default/public/image/2017/07/OMS_logo_0.jpg?itok=cL7pMqiU

[3] Art. 34, Prévention en santé, Chapitre 2, TITRE II PROTECTION ET PREVENTION EN SANTE, Loi n° 18-11 du 18 Chaoual 1439 correspondant au 2 juillet 2018 relative à la santé, JOURNAL OFFICIEL DE LA REPUBLIQUE ALGERIENNE N° 46

[4] OMS. Charte d'Ottawa du 21 novembre 1986. https://www.has-sante.fr/

[5] Art. 29, Protection en santé, Chapitre 1er, TITRE II PROTECTION ET PREVENTION EN SANTE, Loi n° 18-11 du 18 Chaoual 1439 correspondant au 2 juillet 2018 relative à la santé, JOURNAL OFFICIEL DE LA REPUBLIQUE ALGERIENNE N° 46.

[6] https://www.cairn.info/revue-sante-publique-2001-3-page-287.htm

[7] Art. 120. Chapitre 5, Éducation pour la santé, TITRE II PROTECTION ET PREVENTION EN SANTE, Loi n° 18-11 du 18 Chaoual 1439 correspondant au 2 juillet 2018 relative à la santé, JOURNAL OFFICIEL DE LA REPUBLIQUE ALGERIENNE N° 46.

[8] https://www.britannica.com/science/hospital

[9] https://www.who.int/health-topics/hospitals.

[10]https://lvdneng.rosselcdn.net/sites/default/files/dpistyles_v2/vdn_864w/2021/04/11/node_979846/51041481/public/2021/04/11/B9726698396Z.1_20210411094239_000%2BGGFHUADSL.1-0.jpg?itok=7UXoG0GJ1618126999

[11] https://www.vocabulaire-medical.fr/ & https://www.vocabulaire-medical.fr/

[12]https://www.google.com/url?sa=i&url=https%3A%2F%2Facteurdemasante.lu%2Ffr%2Forthopedie%2Fle-genou-le-sejour-a-lhopital%2F&psig=AOvVaw2Kve6qagEiCjd1bN31UVRb&ust=1684749412196000&source=images&cd=vfe&ved=0CBEQjRxqFwoTCNiY7uWShv8CFQAAAAAdAAAAABAE

[13]https://sante.gouv.fr/IMG/pdf/Foire_aux_questions__reforme_des_soins_psychiatriques_-_janvier_2013.pdf

[14] https://www.etp29.fr/wp-content/uploads/2019/11/2010-09-adsp-n%C2%B072-maladies-chroniques-et-ETP.pdf

[15] https://www.cairn.info/revue-reflets-et-perspectives-de-la-vie-economique-2014-4-page-83.htm

[16]https://www.oec-ilibrary.org/sites/265429e7fr/index.html?itemId=/content/component/265429e7-fr#

[17] La gestion des lits dans les hôpitaux et cliniques bonnes pratiques organisationnelles et retours d'expériences https://documentation.ehesp.fr/doc_num.php?explnum_id=3056

[18] https://blog.minitab.com/fr/analyse-predictive-et-determination-de-la-duree-de-sejour-dun-patient-au-moment-de-ladmission

[19] https://blog.minitab.com/fr/analyse-predictive-et-determination-de-la-duree-de-sejour-dun-patient-au-moment-de-ladmission#:~:text=L'%C3%Age

[20] https://publications.msss.gouv.qc.ca/msss/fichiers/2000/00-705-01.pdf

[21] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2679566/

[22] https://aurore.unilim.fr/theses/nxfile/default/049e3737-fa6e-42a4-9a4f-b35d362f5f82/blobholder:0/M20143167.pdf

[23] https://www.aspros-sante.fr/les-avantages-et-les-inconvenients-du-sejour-hospitalier-un-regard-approfondi/

[24] https://www.asjp.cerist.dz/en/downArticle/331/1/1/40386

[25] https://www.caducee.net/DossierSpecialises/systeme-information-sante/pmsi.asp

[26] https://www.cairn.info/revue-sante-publique-2013-5-page-617.htm

[27] Chapitre 22, Le programme de médicalisation du système d'information (PMSI), Robert Holcman, Dans Management hospitalier (2015), pages 553 à 568 https://www.cairn.info/management-hospitalier--9782100724291-page-553.htm

[28] https://sante.gouv.fr/IMG/pdf/guide_pmsi_mco.pdf

[29] BELACHOUI, Ahmed. Le système d'information à l'épreuve de l'organisation Mémoire de Magister, Science de Gestion, TLEMCEN : Université Abou Bekr Belkaid de Tlemcen, FSECSG, 2014, p. 37.

[30] « Degoulet P, Fieschi M (1998) Informatique médicale, 3e éd. Paris, Masson » P. Degoulet () – Université Paris 5 – E-mail : patrice.degoulet@egp.aphp.fr Sous la direction de Alain Venot, avec la collaboration de Anita Burgun et Catherine Quantin, Informatique médicale, e-

Santé, ISBN : 978-2-8178-0337-1, © Springer-Verlag Paris 2013. http://www.himss.org/ASP/index.asp [Site de l'Association HIMMS (Healthcare Information and Management Systems Society)

[31] https://www.capterra.fr/glossary/644/his-hospital-information-system-or-healthcare-information-system

[32]https://www.ummto.dz/dspace/bitstream/handle/ummto/17668/M%C3%A9moire%20de%20fin%20d%27%C3%A9tudes.pdf?sequence=1&isAllowed=y

[33]http://lertim.timone.univmrs.fr/Ecoles/infoSante/2001/SupportsEcole/Patrice.SIH.corte2001.pdf

[34]https://www.has-sante.fr/upload/docs/application/pdf/2009-08/dossier_du_patient_fascicule_1_reglementation_et_recommandations_-_2003.pdf

[35] https://www.alptis.org/lexique-assurance/professionnel-de-sante/

[36] https://www.has-sante.fr/upload/docs/application/pdf/2009-08/dossier_du_patient_-_fascicule_1_reglementation_et_recommandations_-_2003.pdf

[37] https://www.google.com/url?sa=i&url=https%3A%2F%2Fluquet-duranton.fr%2Farticle-de-papeterie-secteur-medical%2F5-dossier-medical-standard.html&psig=AOvVaw3NAPJWLHSHrPtGYCYvvUon&ust=1684749526194000&source=images&cd=vfe&ved=0CBEQjRxqFwoTCNjN4ZuThv8CFQAAAAAdAAAAABAE

[38] réglementation et recommandations – Juin 2003, OUTIL D'AMÉLIORATION DES PRATIQUES PROFESSIONNELLES - Mis en ligne le 26 nov. 2008. https://www.has-sante.fr/jcms/c_438115/fr/dossier-du-patient

[39]https://www.has-sante.fr/upload/docs/application/pdf/2009-08/dossier_du_patient_-_fascicule_1_reglementation_et_recommandations_-_2003.pdf

[40]https://www.has-sante.fr/jcms/c_438115/fr/dossier-du-patient#:~:text=Le%20dossier%20du%20patient%20assure,le%20parcours%20hospitalier%20du%20patient.

[41]https://www.has-sante.fr/upload/docs/application/pdf/2009-08/dossier_du_patient_ _fascicule_1_reglementation_et_recommandations_-_2003.pdf .

[42] https://www.cloudlyyours.com/2020/08/15/dossier-patient-informatise-dmi/

[43]https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRBRAZapi8tG8P6hwLeAQgjhEkEUH3UUAl0gA&usqp=CAU

[44] http://www.medialogis.com/produit/medecinliberal/dossierpatient/

[45] https://masante.oiis.re/portal/thematiques/droits-et-sante/droits-sante-tous-nos-articles/

[46] https://www.cnil.fr/fr/quest-ce-ce-quune-donnee-de-sante

[47] https://locarchives.fr/faq/quest-ce-quune-donnee-de-sante-a-caractere-personnel/

[48] https://mhemo.fr/la-recherche/la-recherche-clinique/les-bases-de-donnees/

[49] https://www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=ad1121416.pdf

[50] Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données. J.O. des Communautés Européennes, n° L 281, 23 nov. 1995 : 0031-0050 https://www.cairn.info/revue-laennec-2007-1-page-12.htm

[51] https://www.avf-biomedical.com/blog/conseils/comment-proteger-les-donnees-de-sante-du-patient/

[52] Rachda, N. M. (2021). Conception et développement des méthodes de prédiction de la durée de séjour hospitalier centrées sur des techniques de "machine learning". Thèse de doctorat, Université de Lorraine. HAL archives-ouvertes. https://theses.hal.science/tel-03609133/document

[53] https://www.cairn.info/revue-laennec-2007-1-page-12.htm

[54]https://www.desmarais-avocats.fr/ce-qui-compte-nest-ni-la-confidentialite-ni-la-disponibilite-des-donnees-de-sante/

[55]https://www.desmarais-avocats.fr/ce-qui-compte-nest-ni-la-confidentialite-ni-la-disponibilite-des-donnees-de-sante/

[56] https://www.avf-biomedical.com/blog/conseils/comment-proteger-les-donnees-de-sante-du-patient/

[57] https://journals.openedition.org/activites/4941

[58] https://hal.science/hal-02327501v3/document consulté le 11/04/2023 a 00 :15

[59] https://www.europeanlawinstitute.eu/

[60] Ouverage Artificial Intelligence A Modern Approach Third Edition, Stuart J. Russell and Peter Norvig]
https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf consulté le 11/04/2023 à 22:42

[61]https://www.widoobiz.com/2019/05/22/camilo-rodriguez-lhomme-qui-voulait-demystifier-lia/consulté 11/04/2023 a 01 :00

[62] https://www.ibm.com/topics/machine-learning consulté le 11/04/2023 à 02:00

[63] https://www.javatpoint.com/applications-of-machine-learning

[64] https://machinelearnia.com/apprentissage-supervise-4-etapes/ consulté le 12/04/2023 a 23 :15

[65] https://machinelearnia.com/apprentissage-supervise-4-etapes/ consulté le 12/04/2023 a 23 :42

[66] https://blent.ai/blog/a/apprentissage-supervise-definition

[67] https://fr.linedata.com/apprentissage-supervise-et-classification consulté le 12/04/2023 à 00:10

[68] https://www.jedha.co/formation-ia/matrice-confusion consulté le 12/04/2023 à 00:45

[69] https://fr.mathworks.com/matlabcentral/fileexchange/105825-multiclass-metrics-of-a-confusion-matrix?s_tid=FX_rc1_behav

[70] Introduction à l'apprentissage automatique, Chapitre V : Régression, https://projeduc.github.io/intro_apprentissage_automatique/regression.html consulté le 13/04/2023 à 01:15

[71] Regression metrics, https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics consulté le 13/04/2023 a 01 :32

[72]https://projeduc.github.io/intro_apprentissage_automatique/regression consulté le 14/04/2023 à 22:30

[73] https://www.lucidchart.com/pages/fr/arbre-de-decision consulté le 14/04/203 à 22:48

[74] https://cedric.cnam.fr/vertigo/cours/ml2/coursArbresDecision.html 14/04/2023 et https://up2.fr/M1/td/TD10_2.html 14/04/2023 à 23:30

[75] https://up2.fr/M1/td/TD10_2.html

[76] https://www.tibco.com/fr/reference-center/what-is-a-random-forest consulté le 14/04/2023 à 00:06

[77] https://www.tibco.com/reference-center/what-is-a-random-forest

[78] https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-rn.pdf consulté le 14/04/2023 à 00:30

[79] https://ml4a.github.io/ml4a/neural_networks/

[80] https://www.jedha.co/formation-ia/algorithme-svm consulté le 14/04/2023 a 01 :02

[81] https://camus.espaceweb.usherbrooke.ca/revue/revue1/article2.pdf consulté le 14/04/2023 à 01:20

[82] https://www.ibm.com/docs/fr/spss-modeler/saas?topic=models-how-svm-works consulté le 14/04/2023 a 01 :35

[83] https://www.ibm.com/ca-fr/topics/knn consulté le 15/04/2023 à 23:23

[84] https://datascientest.com/knn consulté le 15/04/2023 à 23:25

[85] Amandine Allmagn AI Analyst , https://fr.linedata.com/quest-ce-que-lapprentissage-non-supervise consulté le 15/04/2023 à 00:00

[86]https://parlonssciences.ca/ressources-pedagogiques/documents-dinformation/introduction-a-lapprentissage-machine

[87] https://datascientest.com/apprentissage-non-supervise consulté le 15/04/2023 00:32

[88] https://datascientest.com/apprentissage-non-supervise consulté le 16/04/2023 à 22:47

[89] https://brightcape.co/le-clustering-definition-et-implementations/consulté le 16/04/2023 à 23:10

[90] Machine Learning, Algorithmes de clustering, https://developers.google.com/machine-learning/clustering/clustering-algorithms?hl=fr consulté le 16/04/2023 a 23 :22

[91] Dendrogramme : tout sur le diagramme de clustering hiérarchique Le clustering ou regroupement hiérarchique https://datascientest.com/dendrogramme/consulté le 16/04/2023 à 23:57

[92] https://www.50a.fr/0/clustering consulté le 16/04/2023 à 00:18

[93] https://dataanalyticspost.com/Lexique/clustering/consulté le 16/04/2023 à 00:35

[94] https://www.geeksforgeeks.org/ml-hierarchical-clustering-agglomerative-and-divisive-clustering/consulté le 16/04/2023 à 00:45

[95] https://developers.google.com/machine-learning/clustering/clustering-algorithms?hl=fr consulté le 16/04//2023 à 00 :59

[96] http://abdelhamid-djeffal.net/web_documents/coursclustering1819.pdf consulté le 16/04/2023 à 01:30

[97] Journal of machine learning research 2002 Journal of machine learning research 2002, A. Ben-Hur, D. Horn, H. Siegelmann, V. Vapnik, https://www.semanticscholar.org/paper/Support-Vector-Clustering-Ben-Hur-Horn/ consulté le 16/04/2023 a 01 :47

[98]TechTarget Marketing company https://www.techtarget.com/searchbusinessanalytics/definition/association-rules-in-data-mining consulté le 16/04/2023 a 02:10

[99] https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning consulté le 17/04/2023 à 00:53

[100] https://fr.linedata.com/principaux-algorithmes-dapprentissage-non-supervise consulté le 17/04/2023 à 01:13

[101] https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning consulté le 17/04/2023 à 01:39

[102] Machine Learning & Clustering Focus sur l'Algorithme DBSCAN https://datascientest.com/machine-learning-clustering-dbscan consulté le 17/04/2023 à 01:56

[103] https://i0.wp.com/datascientest.com/wp-content/uploads/2020/07/DBSCAN4.png

[104]https://openclassrooms.com/fr/courses/4379436-explorez-vos-donnees-avec-des-algorithmes-non-supervises/4379561-partitionnez-vos-donnees-avec-un-algorithme-de-clustering-hierarchique

[105]https://medium.com/analytics-vidhya/association-rule-mining-including-apriori-algorithm-8c0f9888e125

[106] https://www.softwaretestinghelp.com/apriori-algorithm/

[107]https://www.alteryx.com/fr/glossary/supervised-vs-unsupervised-learning

[108]https://ged.uphf.fr/nuxeo/nxfile/default/74344994-16f7-4b44-99ec-2f02985bcf8b/file:content/MEKHALDI_Rachda_Naila2.pdf

[109] https://www.voxco.com/fr/blog/le-role-du-machine-learning-dans-lanalyse-predictive

[110] https://datascientest.com/machine-learning-tout-savoir

[111] Mekhaldi, R. N. (2022). Conception et développement des méthodes de prédiction de la durée de séjour hospitalier centrées sur des techniques de "machine learning" [Thèse de doctorat, Université Polytechnique Hauts-de-France; Institut national des sciences appliquées Hauts-de-France]. Retrieved from https://theses.hal.science/tel-03609133/document

[112] steyerberg, E. W., Moons, K. G. M., & van der Windt, D. A. W. M. (2019). Prognosis Research in Healthcare: Concepts, Methods, and Impact. Oxford University Press.

[113] Conception et développement des méthodes de prédiction de la durée de séjour hospitalier centrées sur des techniques de "machine learning" https://theses.hal.science/tel-03609133/

[114] Reference: Hsieh, M. C., & Lu, C. L. (2018). Predicting hospital length of stay using artificial neural network. Journal of Healthcare Engineering, 2018, 1-8. DOI: 10.1155/2018/8985205.

[115] Zhang, Q., & Zhang, L. (2020). Predicting length of stay for hospital inpatients using decision tree algorithms. Healthcare Informatics Research, 26(4), 329-336. DOI: 10.4258/hir.2020.26.4.329.

[116] Nematollahi, M., Liao, Y., & Huang, Y. (2019). Predicting length of stay in hospitals intensive care unit using random forest algorithm. International Journal of Environmental Research and Public Health, 16(15), 2711. DOI: 10.3390/ijerph16152711.

[117] Alizadeh, S., Mirfazeli, F. S., & Hosseini, M. J. (2020). Predicting length of stay in intensive care units using support vector machine. Journal of Research in Medical and Dental Science, 8(2), 195-201.

[118] https://www.lemagit.fr/conseil/Comment-construire-un-modele-de-Machine-Learning-en-7-etape

[119] https://registry.opendata.aws/mimiciii/

[120] https://physionet.org/content/eicu-crd/2.0/

[121] https://intensivecarenetwork.com/

[122] https://www.kaggle.com/datasets/babyoda/healthcare-investments-and-length-of-hospital-stay

[123] https://microsoft.github.io/r-server-hospital-length-of-stay/

[124] Caetano, N., Laureano, R. M. S., & Cortez, P. (2014). A Data-driven Approach to Predict Hospital Length of Stay: A Portuguese Case Study. In Proceedings of the 11th International Conference on Health Informatics (HEALTHINF 2018) (pp. 269-276). Lisbon, Portugal. Retrieved from https://www.scitepress.org/Papers/2014/48922/48922.pdf

[125] Lazarescu, M., & Vaida, F. M. (2020). Data Mining Algorithms for Length of Stay Prediction in Hospital Admissions. In Advances in Intelligent Systems and Computing (Vol.

1247, pp. 213-222). Springer. Retrieved from https://link.springer.com/chapter/10.1007/978-3-030-45688-7_21

[126] Davari, M., Mirzaei, T., & Poorolajal, J. (2017). Predicting Length of Stay in Intensive Care Unit Using a Data Mining Technique. Health Services Management Research, 30(2), 105-120. DOI: 10.1177/0951484817696212.

[127] Park, J. M., Kim, S. H., & Choi, S. J. (2013). Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients. Healthcare Informatics Research, 19(2), 121-129. Retrieved from https://synapse.koreamed.org/articles/1075642

[128] Gentimis, T., Alnaser, A. J., Durante, A., Cook, K., & Steele, R. (2017). Predicting Hospital Length of Stay using Neural Networks on MIMIC III Data. Dans IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing, et Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech).

[129] Shieh, C.-C., Lin, C.-J., Yang, P.-C., & Liu, C.-H. (2015). The Identification of Prolonged Length of Stay for Surgery Patients. Dans IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 901-906). DOI: 10.1109/SMC.2015.162. https://ieeexplore.ieee.org/abstract/document/7379654

[130] CBH-GRU-2019:081, RITA. (2019). Predicting Patient Length Of Stay at Time of Admission Using Machine Learning. RITA - Rapport, Institutionen för datavetenskap och media, Mittuniversitetet. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1338294&dswid=1104

[131] Harerimana, G., Kim, J. W., & Jang, B. (2021). A deep attention model to forecast the Length Of Stay and the in- hospital mortality right on admission from ICD codes and demographic data. Journal of Biomedical Informatics, 118, 103778.

[132] Huang, C. Y., & Wu, C. H. (2016). Predicting Length of Stay for Cardiology Patients Using Artificial Neural Networks. Journal of Healthcare Engineering, 2016, 7035463, 11 pages. http://dx.doi.org/10.1155/2016/7035463

[133] Almazrou, S., Rehman, M. U., & Alshayban, A. (2021). Comparison of Machine Learning Algorithms in the Prediction of Hospitalized Patients with Schizophrenia. Cureus, 13(2), e13422.DOI:10.7759/cureus.13422.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9003328/

[134] Ghosh, S., Saha, A., & Nag, S. (2021). Predicting length of stay in hospitals intensive care unit using general admission features. Journal of Biomedical Informatics X, 10, 100079. DOI: 10.1016/j.ybiox.2021.100079.
https://www.sciencedirect.com/science/article/pii/S2090447921001349

[135] Li, J. S., Tian, Y., Liu, Y. F., Shu, T., & Liang, M. H. (2010). Applying a BP neural network model to predict the length of hospital stay. Journal of healthcare engineering, 1(4), 509-526. https://link.springer.com/chapter/10.1007/978-3-642-37899-7_2

[136]https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/

[137]https://www.lebigdata.fr/wp-content/uploads/2018/09/python-big-data-machine-learning.jpg

[138] https://research.google.com/colaboratory/faq.html?hl=fr

[139] https://www.kaggle.com/datasets/aayushchou/hospital-length-of-stay-dataset-microsoft

[140] https://microsoft.github.io/r-server-hospital-length-of-stay/input_data.html