

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABDELHAMID IBN BADIS - MOSTAGANEM



Faculté des Sciences Exactes et d'Informatique
Département de Mathématiques et informatique
Filière : Informatique

RAPPORT DE MINI-PROJET

Option : **Ingénierie des Systèmes d'Information**

THEME :

VERS UN MODELE DE PREDICTION D'EPIDEMIE

Etudiant(e): « **Ilyes Benarroum** »

« **Sahinine Mohammed Chems Eddine** »

Encadrant(e) : « **YOUNSI Fatima Zohra** »

Année Universitaire 2020-2021

Résumé

La prédiction des tendances nouvelles et urgentes des données épidémiologiques est un problème important pour la santé publique. Ce problème a attiré une attention croissante dans les communautés d'exploration de données et d'apprentissage automatique. L'intelligence artificielle permet d'extraire des informations pertinentes d'un ensemble de données accessibles de plus en plus conséquent qu'il serait difficile de parcourir de manière non automatisée.

Dans ce travail de projet de fin d'étude, notre objectif est de proposer un nouveau système de prédiction de la COVID-19 en Algérie. Ce système est basé essentiellement les méthodes de l'IA, à savoir : ARIMA, LSTM et Prophet. Le but de notre système est d'aider les responsables et les planificateurs de la santé publique à gérer les services et organiser l'infrastructure médicale ainsi qu'évaluer les plans d'actions pour lutter l'épidémie

Mots-clés :

Système de Prédiction de la COVID-19 (SYSPC), L'intelligence artificielle, Apprentissage automatique, Prédiction, l'épidémie de la COVID-19, autoregressive integrated moving average (ARIMA), Long short-term memory (LSTM), Prophet.

Abstract:

Predicting new and urgent trends in epidemiological data is an important public health problem. This problem has gained increasing attention in the data mining and machine learning communities. Artificial intelligence can extract relevant information from an increasingly accessible dataset that would be difficult to navigate in a non-automated fashion.

In this mini project work, our objective is to describe the tools of artificial intelligence and to present the objectives and prediction systems in epidemiology. In the literature, there are several methods and approaches allowing prediction. In this work, we are interested in Artificial Neural Networks (ARNs) which allows to extract relationships in order to predict future value.

Keywords:

Système de Prédiction de la COVID-19 (SYSPC), Artificial intelligence, machine learning prediction, COVID-19 epidemic, autoregressive integrated moving average (ARIMA), Long short-term memory (LSTM), Prophet.

Liste des figures

Figure N°	Titre de la figure	Page
Figure 1	Présentons la chronologie de l'IA depuis sa création	7
Figure 2	Le processus général de l'apprentissage automatique	9
Figure 3	Les étapes de résolution d'un problème de clustering	12
Figure 4	Les différentes étapes du processus d'apprentissage profond	14
Figure 5	Réseau de neurones récurrents	16
Figure 6	Architecture standard d'un réseau de neurone convolutif	17
Figure 7	Présente les différentes branches de l'épidémiologie	20
Figure 8	Les actions essentielles en épidémiologie	22
Figure 9	Un exemple d'un arbre de décision médical	24
Figure 10	Un exemple de la régression linéaire Multiple	24
Figure 11	Organigramme de la méthode LSTM	32
Figure 12	Organigramme de la méthode ARIMA	34
Figure 13	Organigramme de la méthode Prophet	36
Figure 14	Démarche globale de SYSPC	38
Figure 15	Diagramme de séquence de SYSPC	40
Figure 16	Représentation tabulaire des données de la COVID-19 en Algérie	43
Figure 17	La comparaison entre les cas positifs COVID-19 et les cas prédits en Algérie avec les 3 méthodes : a. ARIMA, b. LSTM, c. Prophet	45
Figure 18	La comparaison entre les cas positifs COVID-19 et les cas prédits à Alger avec les 3 méthodes : a. ARIMA, b. LSTM, c. Prophet	46
Figure 19	La comparaison entre les cas positifs COVID-19 et les cas prédits à Blida avec les 3 méthodes : a. ARIMA, b. LSTM, c. Prophet	47
Figure 20	La comparaison entre les cas positifs COVID-19 et les cas prédits à Oran avec les 3 méthodes : a. ARIMA, b. LSTM, c. Prophet	48
Figure 21	La propagation du COVID-19 en Algérie	49
Figure 22	La propagation du COVID-19 en Mostaganem	49
Figure 23	Comparaison entre les cas positifs Covid-19 prévus et réels des 15 derniers jours par le modèle ARIMA pour l'Algérie	50
Figure 24	Comparaison entre les cas positifs Covid-19 prévus et réels des 15 derniers jours par le modèle Prophet pour l'Algérie	51

Liste des tableaux

Tableau N°	Titre du tableau	Page
Tableau 1	Les domaines qui utilisent l'IA	8
Tableau 2	Les deux types de RNA	15
Tableau 3	Récapitulatif des caractéristiques de : RNA, RNC et RNR	17
Tableau 4	Travaux de recherche qui utilisant les arbres de décisions	25
Tableau 5	Travaux de recherche qui utilisant la régression	26
Tableau 6	Travaux de recherches utilisant les techniques citées auparavant	28
Tableau 7	Représente les outils que nous avons utilisées.	42

Liste des abréviations

Abréviation	Expression Complète	Page
IA	Intelligence Artificielle	1
RNA	Réseaux de Neurones Artificiels	12
RNR	Réseau de Neurones Récurrents	13
RNC	Réseau neuronal convolutif	14
ARIMA	Auto Regressive Integrated Moving Average	25
SARIMA	Seasonal Auto Regressive Integrated Moving Average	25
LSTM	Long short-term memory	25
SYSPC	<u>S</u> ystème de <u>P</u> rédiction de la <u>C</u> COVID-19	29

Table of Contents

Introduction Générale	4
Intelligence Artificielle	6
1.1 Introduction	6
1.2 Définition de l'Intelligence Artificielle (IA)	6
1.3 Histoire de l'Intelligence Artificielle (IA)	6
1.4 Avantages et inconvénients de l'Intelligence Artificielle	7
1.5 L'importance de l'intelligence artificielle	8
1.6 Applications de l'Intelligence Artificielle (IA)	8
1.7 L'apprentissage automatique (Machine Learning)	9
1.7.1 Définition	9
1.7.2 Processus de l'apprentissage automatique	9
1.7.3 Différents types d'apprentissage automatique	9
1.8 Apprentissage profond (Deep Learning)	13
1.8.1 Définition	13
1.8.2 Le schéma général du processus d'apprentissage profond	14
1.8.3 Les différents types d'apprentissage profond	14
1.9 Conclusion :	18
Chapitre 2 Prédiction en épidémiologie	19
2.1 Introduction	19
2.2 Qu'est-ce que L'épidémiologie	19
2.2.1 Stratégies dans la recherche épidémiologique	20
2.3 Objectifs de l'épidémiologie	20
2.4 Santé publique	21
2.4.1 Surveillance en épidémiologie	21
2.4.2 Actions en épidémiologie	21
2.5 Prédiction en épidémiologie	22

2.6 Modèles de prédiction basé sur l'apprentissage.....	23
2.6.1 Arbre de décision	23
2.6.2 Régression linéaire.....	24
2.7 Modèle de prédiction basé sur les RNA(s)	27
2.7.1 Auto Regressive Integrated Moving Average (ARIMA).....	27
2.7.2 LSTM.....	27
2.7.3 Prophet.....	27
2.8 Conclusion	29
Chapitre 3.....	30
Conception et Modélisation	30
3.1 Introduction	30
3.2 Objectif de notre système SYSPC.....	31
3.3 Description de notre système	31
3.3.1 Système LSTM :	31
3.3.2 Système ARIMA.....	33
3.3.3 Système Prophet.....	35
3.4 Description de Base de données.....	37
3.5 Démarche globale de SYSPC.....	38
3.6 Langage UML	40
3.7 Diagramme de séquence.....	40
3.8 Description du diagramme	42
3.9 Conclusion.....	42
Chapitre 4 Mise en oeuvre	43
4.1 Introduction.....	43
4.2 Outils de développement.....	43
4.2.1 Langage de programmation PYTHON	43
4.3 Etapes de développements	44

4.3.1 Base de données utilisées.....	44
4.4 Résultat de système SYSPC.....	45
4.4.1 Représentation graphique des résultats.....	45
4.4.2 Représentation cartographique des résultats.....	50
4.5 Validation des méthodes utilisées.....	50
4.6 Conclusion :	52
Conclusion Générale.....	53
Bibliographie.....	54

Introduction Générale

L'épidémiologie est le domaine d'étude de la fréquence à laquelle les maladies surviennent dans différents groupes de personnes ainsi que l'étude de la dynamique des différents états de santé. Les techniques utilisées dans ce domaine permettent de trouver les causes des problèmes de santé et des maladies dans les populations et il traite de l'incidence, de la distribution et du contrôle éventuel des maladies et d'autres facteurs liés à la santé. En effet, l'élaboration des mesures d'urgence, de prévention des infections et des plans de protection nécessite la mise en place des outils de prédiction future de la maladie. Dans ce contexte, la prédiction peut se faire à partir de la collecte des données et l'utilisation des méthodes d'IA.

Dans ce travail de projet de fin d'étude, nous nous intéressons à l'élaboration d'un nouveau système de prédiction dédié à l'épidémiologie de la COVID-19. Ce dernier est basé principalement sur trois méthodes de l'IA, à savoir : Ce système est basé essentiellement les méthodes de l'IA, à savoir : ARIMA, LSTM et Prophet. Le but de notre système est d'aider les responsables et les planificateurs de la santé publique à gérer les services et organiser l'infrastructure médicale ainsi qu'évaluer les plans d'actions pour lutter l'épidémie.

Organisation du mémoire

Ce mémoire est composé de quatre chapitres :

Le premier chapitre est dédié à la présentation de l'IA. Nous allons, d'abord, donner des définitions sur les différents concepts de l'IA. Puis, nous allons, expliquer la méthode RNA. Enfin, nous allons exposer les modèles RNA les plus utilisés en prédiction épidémiologique.

Le deuxième chapitre présente le contexte général de la prédiction en épidémiologie. Dans ce chapitre, nous allons donner des définitions sur l'épidémiologie et ses actions de base. Nous allons, également, présenter un panorama les modèles de prédiction basé sur l'apprentissage les plus connus.

Le troisième chapitre est consacré à la description succincte de notre système pour la surveillance épidémiologique, les différents outils d'investigations qui ont participé à

l'élaboration d'un tel système, notre démarche adoptée ainsi que les détails des méthodes utilisées.

Le quatrième chapitre comporte la description des outils utilisés dans la mise en œuvre de notre prototype ainsi que les différents résultats d'expérimentation obtenus.

Enfin, nous terminons ce manuscrite par une conclusion générale et quelques perspectives.

Intelligence Artificielle

1.1 Introduction

Dans ce chapitre, nous allons décrire les notions de base sur IA, nous allons également, détailler les Réseaux de Neurones Artificiels (RNA)

1.2 Définition de l'Intelligence Artificielle (IA)

L'intelligence artificielle (IA) est un ensemble d'algorithmes qui permet de résoudre des problèmes complexes qu'on aurait cru réservés à l'intelligence humaine dans les domaines de la compréhension, la communication (dialogue entre machines et avec l'humain), la structuration de la mémoire, le raisonnement critique, l'adaptation, l'apprentissage en autonomie (Deep Learning).

Selon E. Feigenbaum: «AI is the part of computer science concerned with designing intelligent computer systems » [1]

Dans la langue courante, l'IA désigne également les dispositifs informatiques ou robotiques mettant en œuvre ces fonctions. Deux écoles se distinguent dans l'approche des objectifs de l'intelligence artificielle : La première considère que l'IA doit fonctionner sur le même modèle que l'intelligence humaine, la deuxième préconise un modèle idéal, dont la perfection rationnelle serait dissociée des critères et processus de l'intelligence biologique [2].

1.3 Histoire de l'Intelligence Artificielle (IA)

L'histoire de l'IA se compose de travaux et de recherches originaux non seulement des mathématiciens et des informaticiens, mais des études de psychologues, de physiciens et d'économistes ont également été très utilisées. Dans la figure 1, nous présentons la chronologie de l'IA depuis sa création.

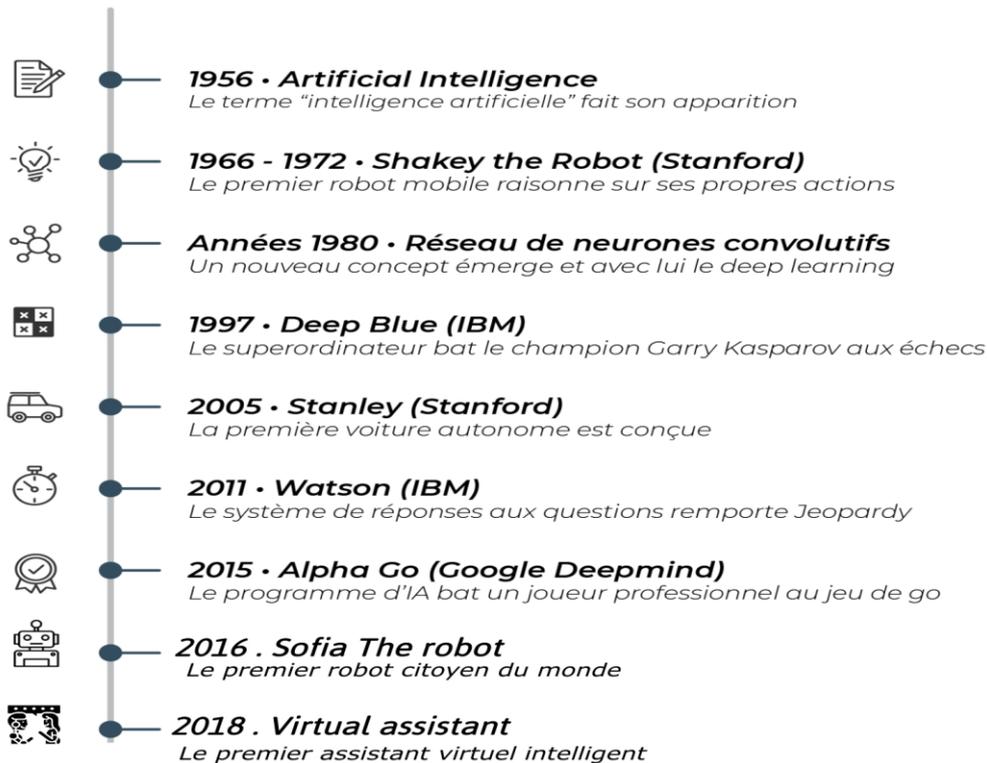


Figure 1 - présentons la chronologie de l'IA depuis sa création

1.4 Avantages et inconvénients de l'Intelligence Artificielle

Dans la littérature, l'IA dispose des avantages et des inconvénients. Dans la suite, nous présentons quelqu'un :

👍 Les avantages de l'intelligence artificielle

- Efficacité et efficacité et la réduction des erreurs
- L'application quotidienne
- Les assistants numériques
- Les travaux répétitifs
- Les applications médicales
- Aucune pause

👎 Les inconvénients de l'intelligence artificielle

- Un coût élevé
- Le chômage
- Pas de créativité
- Aucune amélioration avec l'expérience
- Aucune initiative[3].

1.5 L'importance de l'intelligence artificielle

De nos jours, les êtres humains et les machines génèrent des données plus vite qu'il n'est humainement possible de les absorber et de les interpréter pour prendre des décisions complexes. L'IA est la base de tout apprentissage par un ordinateur et représente l'avenir des processus décisionnels complexes. En effet, les ordinateurs sont capables de calculer ces combinaisons et les meilleures permutations possibles très efficacement, afin de prendre la bonne décision. L'IA avec son évolution logique, machine learning et deep learning représentent l'avenir de la prise de décisions [4].

1.6 Applications de l'Intelligence Artificielle (IA)

Nous bénéficions de l'IA au quotidien dans notre vie car elle est partout et parmi les domaines qui utilisent l'IA se trouvent [5] :

Tableau 1- les domaines qui utilisent l'IA

Recherche Internet	Les moteurs de recherche apprennent à partir des immenses flux de données entrants, fournis par leurs utilisateurs pour fournir des résultats de recherche pertinents.
Traduction	L'IA est arrivée au niveau de la compréhension du sens de la phrase entière et de la traduire dans n'importe quelle langue.
Reconnaissance faciale	Cette technologie est principalement utilisée dans les téléphones pour les déverrouiller et les aéroports pour reconnaître les criminels et dans les banques pour protéger le compte de leurs clients, etc.
Voiture autonome	Bien que les véhicules autonomes ne soient pas encore de série, ils sont l'avenir en raison du confort et de la sécurité qu'ils offrent.
Santé	Nous l'utilisons dans la prévision des épidémies et découvrons des modèles qui pourraient conduire à de nouvelles découvertes en médecine et des moyens d'améliorer les diagnostics individuels, etc.
Assistant virtuel	Les smartphones utilisent l'IA pour fournir des services aussi pertinents et personnalisés que possible. Les assistants virtuels répondant aux questions, fournissant des recommandations et

	aidant à organiser les routines quotidiennes sont devenus omniprésents.
Maison intelligente et Ville intelligente	C'est le coup d'œil de l'IA dans le futur. L'IA sera partout, de nos petits appareils à notre maison en passant par nos villes, tout sera connecté !

1.7 L'apprentissage automatique (Machine Learning)

1.7.1 Définition

L'apprentissage automatique est une application d'IA qui offre aux systèmes la possibilité d'apprendre et de s'améliorer automatiquement à partir de l'expérience sans être explicitement programmés. Il s'agit d'algorithmes qui analysent un ensemble de données afin de déduire des règles constituant des connaissances permettant d'analyser de nouvelles situations [6].

Cette technologie est en effet capable d'extraire les données de valeur parmi d'immenses sources d'informations complexes.

1.7.2 Processus de l'apprentissage automatique

Le processus de l'apprentissage automatique se déroule comme suit :

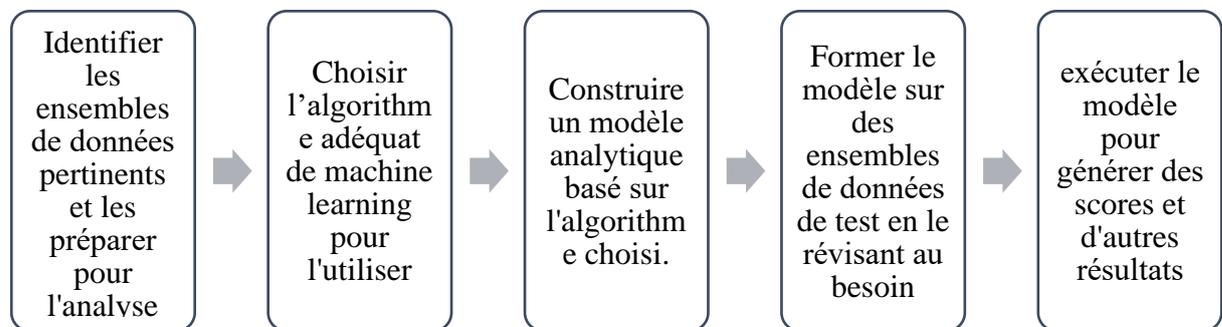


Figure 2 - Le processus général de l'apprentissage automatique

1.7.3 Différents types d'apprentissage automatique

Dans la littérature de l'apprentissage automatique, nous distinguons, usuellement, les types suivants :

1.7.3.1 Apprentissage supervisé

L'apprentissage supervisé est une forme d'apprentissage automatique dans laquelle les entrées et les sorties de notre modèle d'apprentissage automatique nous sont toutes deux disponibles, c'est-à-dire que nous savons à quoi ressemblera la sortie en regardant simplement l'ensemble de données. Le nom « supervisé » signifie qu'il existe une relation entre les caractéristiques d'entrée et leur sortie respective dans les données. Le but de tout algorithme d'apprentissage automatique que nous implémentons est de prédire une sortie nouvelle mais similaire pour des données jamais vues auparavant en estimant cette relation [7].

L'apprentissage supervisé se découpe en deux parties :

- La première correspond à déterminer un modèle de données étiquetées.
- La deuxième consiste à prédire l'étiquette d'une nouvelle donnée connaissant le modèle préalablement appris.

➤ Les algorithmes d'apprentissage supervisé

Classification

La *classification supervisée* consiste à attribuer automatiquement une catégorie (ou une classe) à des données dont on ne connaît pas la catégorie. Pour ce faire, un classifieur (algorithme de machine learning) est entraîné sur des données similaires ou très proches des données que l'on souhaite classer [8].

C'est d'ailleurs de là que vient le supervisée du mot classification supervisée, car l'humain a d'ores et déjà trié et classé les données sur lesquelles va s'entraîner le classifieur. Dans le cas où les données d'entraînement ne sont ni triées ni classées, l'on parle alors de classification non supervisée.

Exemples d'algorithmes de classification : Naïve Bayésienne, Classification, naïve bayésienne, Régression logistique, Arbre de décision, Machine à vecteurs de support

Régression

Un algorithme de régression permet de trouver un modèle en fonction des données d'entraînement. Le modèle calculé permettrait de donner une estimation sur une nouvelle donnée non encore vue par l'algorithme. Les algorithmes de régression

peuvent prendre plusieurs formes en fonction du modèle qu'on souhaite construire [9]. La régression linéaire est le modèle le plus simple, il consiste à trouver la meilleure droite qui s'approche le plus des données d'apprentissage. La fonction de prédiction sera donc une droite et il existe d'autres modèles de régression par exemple : régression simple, multiple et logistique.

1.7.3.2 **Apprentissage non supervisé**

L'apprentissage non supervisé est un système d'apprentissage automatique qui apprend et crée des modèles à partir de jeu de données d'entraînement ni étiquetées ni catégorisées et les algorithmes du système agissent sur les données sans formation préalable. On demande alors à la machine de créer ses propres réponses. Elle propose ainsi des réponses à partir d'analyses et de groupement de données.

Une application centrale de l'apprentissage non supervisé est dans le domaine de l'estimation de la densité dans les statistiques, mais englobe de nombreux autres domaines d'apprentissage sans supervision impliquant la synthèse et expliquer les caractéristiques de données [10].

➤ **Algorithmes d'apprentissage non supervisé**

Partitionnement de données (Clustering)

Le clustering est une méthode d'analyse statistique utilisée pour organiser des données brutes dans de nombreux domaines. L'algorithme de clustering est pratique lorsque nous voulons obtenir des informations détaillées sur nos données. Le déroulement classique de la résolution d'un problème de clustering est illustré dans le schéma suivant [11] :

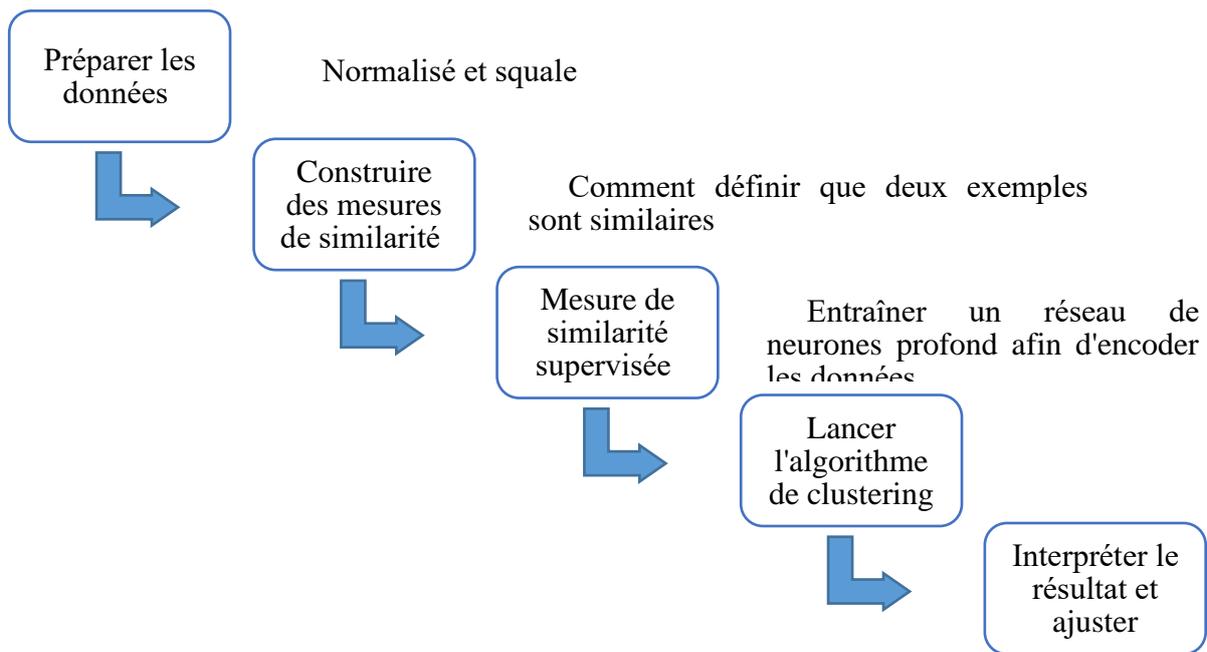


Figure 3 - Les étapes de résolution d'un problème de clustering

Exemples algorithms de Clustering: K-means clustering, K-nearest neighbor

Réduction de la dimensionnalité

La Réduction de la dimensionnalité est une méthode qui consiste à prendre des données dans un espace de grande dimension, et à les remplacer par des données dans un espace de plus petite dimension.

La réduction de dimensionnalité peut être considérée comme la compression d'un fichier. Cela signifie retirer les informations qui ne sont pas pertinentes. Il réduit la complexité des données et tente de conserver les données significatives. Par exemple, dans la compression d'image, nous réduisons la dimensionnalité de l'espace dans lequel l'image reste telle qu'elle est sans détruire trop de contenu significatif de l'image [12].

1.7.3.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est une approche de l'apprentissage automatique qui combine une petite quantité de données étiquetées avec une grande quantité de données non étiquetées pendant la formation. En général, les algorithmes d'apprentissage semi-supervisé fonctionnent comme suit [13] :

- Un algorithme d'apprentissage automatique semi-supervisé utilise un ensemble limité de données d'échantillons étiquetées pour s'entraîner, ce qui donne un modèle « partiellement entraîné ».
- Le modèle partiellement formé des données étiquette et des données non étiquetées. Étant donné que l'ensemble de données étiqueté de l'échantillon présente de nombreuses limites sévères (par exemple, un biais de sélection dans les données du monde réel), les résultats de l'étiquetage sont considérés comme des données « pseudo-étiquetées ».
- Les ensembles de données étiquetés et pseudo-étiquetés sont combinés, créant un algorithme unique qui combine à la fois les aspects descriptifs et prédictifs de l'apprentissage supervisé et non supervisé.

L'apprentissage semi-supervisé utilise le processus de classification pour identifier les actifs de données et le processus de regroupement pour les regrouper en parties distinctes.

1.7.3.4 Apprentissage par renforcement

L'apprentissage par renforcement consiste à apprendre par interaction avec l'environnement et en observant le résultat de certaines actions. Il permet à des machines de déterminer automatiquement le comportement idéal dans un contexte spécifique, afin de maximiser ses performances. Pour cela, un simple retour des résultats est nécessaire pour apprendre comment les machines doivent agir. Ceci est appelé le signal de renforcement [14].

1.8 Apprentissage profond (Deep Learning)

1.8.1 Définition

L'apprentissage automatique est une autre forme d'intelligence artificielle (IA), dérivée du Machine Learning (apprentissage automatique) qui dispose de réseaux capables d'apprendre sans surveillance à partir de données non structurées ou non étiquetées. Également connu sous le nom d'apprentissage neuronal profond (deep neural learning) ou réseau de neurones profond (deep neural network) [15].

1.8.2 Le schéma général du processus d'apprentissage profond

La Figure illustre les différentes étapes du processus d'apprentissage profond :

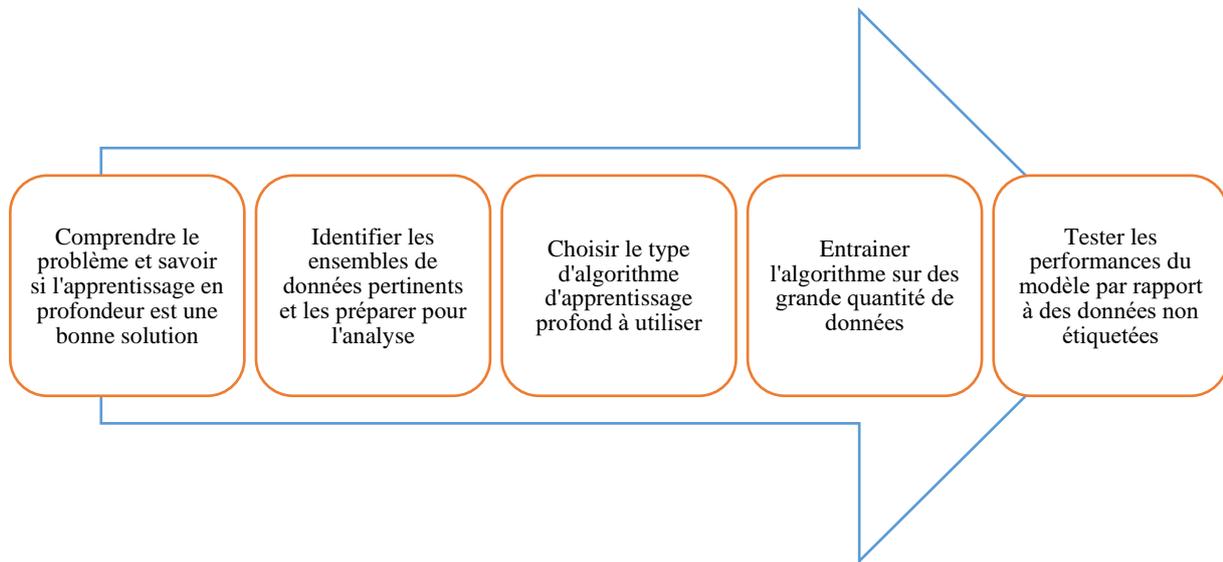


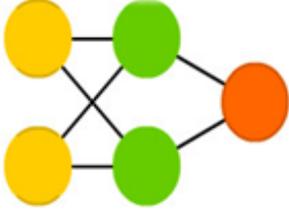
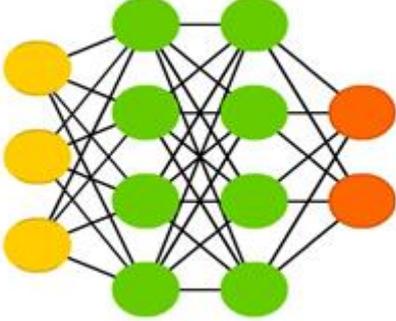
Figure 4 - Les différentes étapes du processus d'apprentissage profond

1.8.3 Les différents types d'apprentissage profond

1.8.3.1 Réseau de Neurones Artificiels (RNA)

Réseau de Neurones Artificiels est un groupe de plusieurs perceptrons ou neurones à chaque couche. Le RNA est également connu sous le nom de réseau neuronal Feed-Forward car les entrées sont traitées uniquement dans le sens direct. Ce type de réseaux de neurones est l'une des variantes les plus simples des réseaux de neurones. Ils transmettent les informations dans une direction, via divers nœuds d'entrée, jusqu'à ce qu'elles parviennent au nœud de sortie[16]. Le réseau peut avoir plusieurs ou non couches de nœuds cachés (Voir Table 2), ce qui rend leur fonctionnement plus interprétable.

Tableau 2 - Les deux types de RNA.

RNA Basique	RNA Perceptron Multicouches
	
<p>Un réseau neuronal artificiel (RNA) est un</p>	<p>Un réseau neuronal profond (RND) est un réseau neuronal artificiel (RNA) avec plusieurs couches entre les couches d'entrée et de sortie.</p>

Ce type de réseau neuronal est le réseau neuronal très basique où le contrôle de flux se produit à partir de la couche d'entrée et va vers la couche de sortie.

- Étant donné que les données ne se déplacent que dans une direction, il n'y a pas de technique de rétropropagation dans ce réseau
- Dans ce réseau, la somme des poids présents dans l'entrée est introduite dans la couche d'entrée
- Ces types de réseaux sont utilisés dans l'algorithme de reconnaissance faciale utilisant la vision par ordinateur.

1.8.3.2 Réseau de Neurones Récurrents (RNR)

Les réseaux de neurones récurrents sont des modèles un peu plus complexes et ils prennent en compte un contexte dans leur fonction de décision. Ils enregistrent la sortie des nœuds de traitement et réintroduisent le résultat dans le modèle (ils ne transmettent pas les informations dans une seule direction). C'est ainsi que le modèle apprend à prédire le résultat d'une couche. Chaque nœud du modèle RNR agit comme une cellule mémoire, poursuivant le calcul et la mise en œuvre des opérations. Si la prédiction du réseau est incorrecte, le système s'auto-apprend et continue de travailler à la bonne prédiction pendant la rétropropagation [17].

Réseau de neurones récurrents

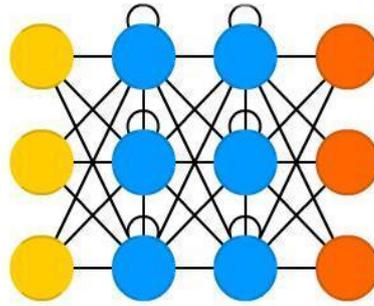


Figure 5 – Réseau de neurones récurrents

Parmi les avantages de ce type de réseau, nous citons :

- Le RNR est très efficace pour la reconnaissance d'image et l'identification de différents modèles d'image.
- Il est utile pour maintenir un petit état de mémoire ce qui est très utile pour développer le chatbot.
- Il est utilisé dans le développement de chatbot et les technologies de synthèse vocale.
- Cette méthode aide le réseau à prédire la sortie.

1.8.3.3 Réseau neuronal convolutif (RNC)

Réseau Neuronal Convolutif c'est l'un des modèles les plus populaires utilisés aujourd'hui. Ce modèle de calcul de réseau de neurones utilise une variante de perceptrons multicouches et contient une ou plusieurs couches convolutives qui peuvent être entièrement connectées ou regroupées. Ces couches convolutives créent des cartes de caractéristiques qui enregistrent une région d'image qui est finalement divisée en rectangles et envoyée pour un traitement non linéaire[18], la Figure 6 présente l'architecture d'un RNC.

Réseau Neuronal Convolutif

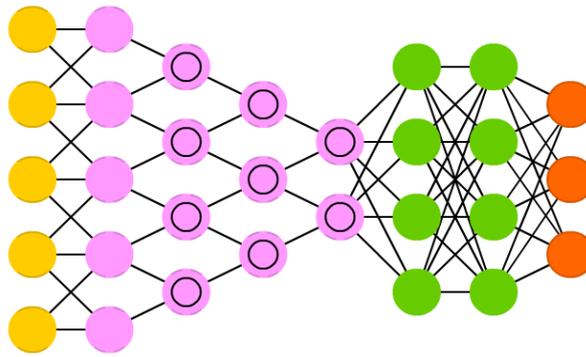


Figure 6 - Architecture standard d'un réseau de neurone convolutif

Parmi les avantages de RNC, nous citons :

- Le RNC est l'une des variantes du perceptron multicouche.
- Il peut contenir plus d'une couche de convolution et comme il contient une couche de convolution, le réseau est très profond avec moins de paramètres.
- Le RNC est très efficace pour la reconnaissance d'image et l'identification de différents modèles d'image.

Le tableau suivant récapitule les caractéristiques qui différencient chacune des méthodes :

Tableau 3 - récapitulatif des caractéristiques de : RNA, RNC et RNR

	RNA	RNC	RNR
Type de données	Données tabulaires, données textuelles	Données d'image	Données de séquence
Partage de paramètres	Non	Oui	Oui
Entrée de longueur fixe	Oui	Oui	Non
Connexions récurrentes	Non	Non	Oui

Dégradé de disparition et d'explosion	Oui	Oui	Oui
Relation spatiale	Non	Oui	Non
Performance	RNA est considéré comme moins puissant que RNC, RNR.	RNR est considéré comme plus puissant que RNR, RNA.	RNR inclut moins de compatibilité des fonctionnalités par rapport à RNC.
Application	Reconnaissance faciale et vision par ordinateur.	Reconnaissance faciale, numérisation de texte et traitement du langage naturel.	Conversions de synthèse vocale.
Principaux avantages	Ayant une tolérance aux pannes, Capacité à travailler avec des connaissances incomplètes.	Haute précision dans les problèmes de reconnaissance d'image, partage de poids.	Se souvient de chaque information, prédiction de séries chronologiques.
Désavantages	Dépendance matérielle, comportement inexplicable du réseau.	De grandes données d'entraînement sont nécessaires, ne codez pas la position et l'orientation de l'objet.	Dégradé disparaissant, dégradé qui explose.

1.9 Conclusion :

Dans le présent chapitre, nous avons d'abord introduit l'IA et son importance dans la société humaine. Puis, nous avons expliqué l'apprentissage automatique ainsi que ces différents types. Finalement, nous avons brièvement introduit le concept d'apprentissage profond.

Dans le chapitre suivant, nous allons aborder le concept de prédiction en épidémiologie.

Chapitre 2

Prédiction en épidémiologie

2.1 Introduction

Les systèmes d'information actuels dédiés à la santé publique ne permettent pas souvent de détecter les maladies infectieuses émergentes ni l'évolution future de la maladie. En intégrant des données cliniques et les outils d'IA pertinents, nous pouvons combler les lacunes critiques en matière de couverture et accélérer la prédiction et la compréhension des épidémies.

Dans ce présent chapitre, nous allons décrire, d'une façon détaillée, le concept de la prédiction en épidémiologie.

2.2 Qu'est-ce que L'épidémiologie

L'épidémiologie peut être définie comme étant : “ *Une discipline scientifique dont l'objet est l'étude de distribution des maladies et des états de santé dans la population humaine, ainsi que les facteurs qui influent cette distribution il existe trois types L'épidémiologie descriptive, Epidémiologie analytique et L'épidémiologie évaluative*” [19]

La Figure. 7 présente les différentes branches de l'épidémiologie :

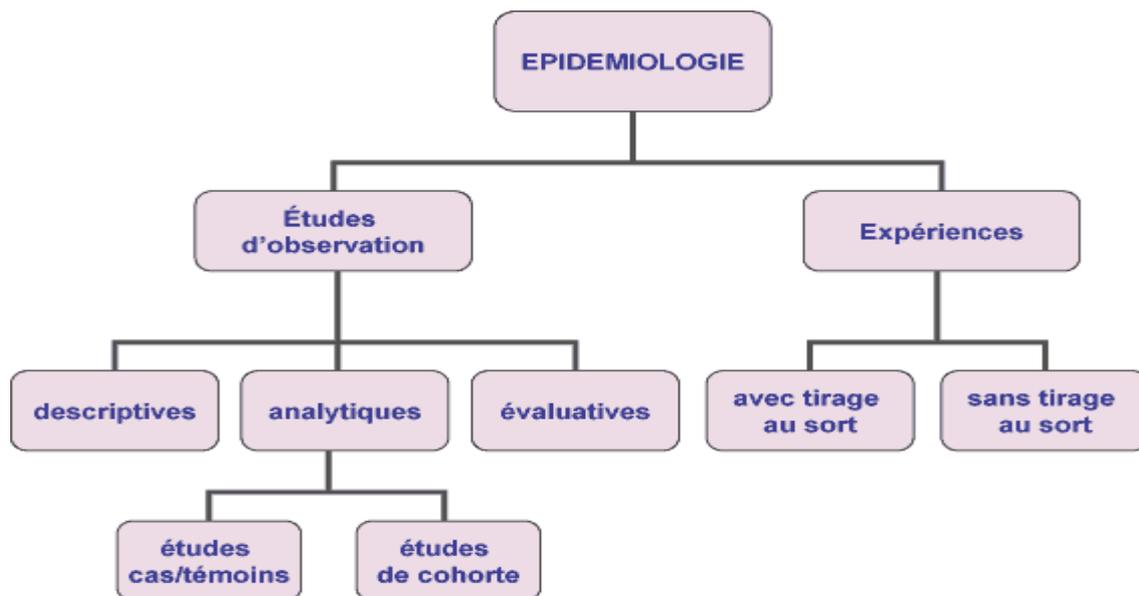


Figure 7 - présente les différentes branches de l'épidémiologie

2.2.1 Stratégies dans la recherche épidémiologique

2.2.1.1 Épidémiologie descriptive

La finalité de l'épidémiologie descriptive est la description d'un phénomène de santé dans une population (fréquence, variations) en fonction des caractéristiques de cette population et de paramètres tels que le temps et l'espace. Elle est nécessaire pour mesurer l'importance d'un problème de santé et soulever des hypothèses étiologiques (facteurs de risque de survenue de cette maladie) [20].

2.2.1.2 Épidémiologie analytique

L'épidémiologie explicative (ou analytique, ou étiologique) se donne pour but de chercher les causes des problèmes de santé. Elle étudie le rôle de l'exposition à des facteurs susceptibles d'intervenir dans l'apparition d'affections. Elle procède par des comparaisons de la fréquence de l'affection dans des groupes de sujets diversement exposés à ces facteurs, comparaison de la fréquence et de l'intensité de l'exposition entre des sujets malades et d'autres qui ne le sont pas [21].

2.2.1.3 Épidémiologie évaluative

Les enquêtes évaluatives étudient les actions de prévention ou de traitement [22].

2.3 Objectifs de l'épidémiologie

Parmi les objectifs de l'épidémiologie, nous citons les suivants :

- Expliquer les principes causaux des maladies.
- L'application de l'épidémiologie dans la prévention et la promotion de la sante.
- Introduction et application de concept d'épidémiologie clinique, pharmaco-épidémiologie.
- La formation de spécialistes des professions connexes.
- La stimulation du l'intérêt sur l'épidémiologie.

2.4 Santé publique

D'après Bourdillon et al., le terme santé publique peut être défini comme suit : « *La santé publique est un ensemble de savoirs, de savoir-faire, de pratiques et de règles juridiques qui visent à connaître, à expliquer, à préserver, à protéger et à promouvoir l'état de santé des personnes* [23]. »

2.4.1 Surveillance en épidémiologie

Le dictionnaire Larousse définit littéralement la surveillance comme étant : « *Une action de veiller avec attention, autorité et souvent avec défiance, de contrôler* ». Cette définition académique du terme surveiller renvoie à une notion ancienne de la surveillance épidémiologique au temps où surveiller était synonyme de Se **protéger** contre les maladies contagieuses (peste, Choléra, et les cas suspects détectés étaient alors mis en quarantaine pour éviter la propagation des épidémies dans la population).

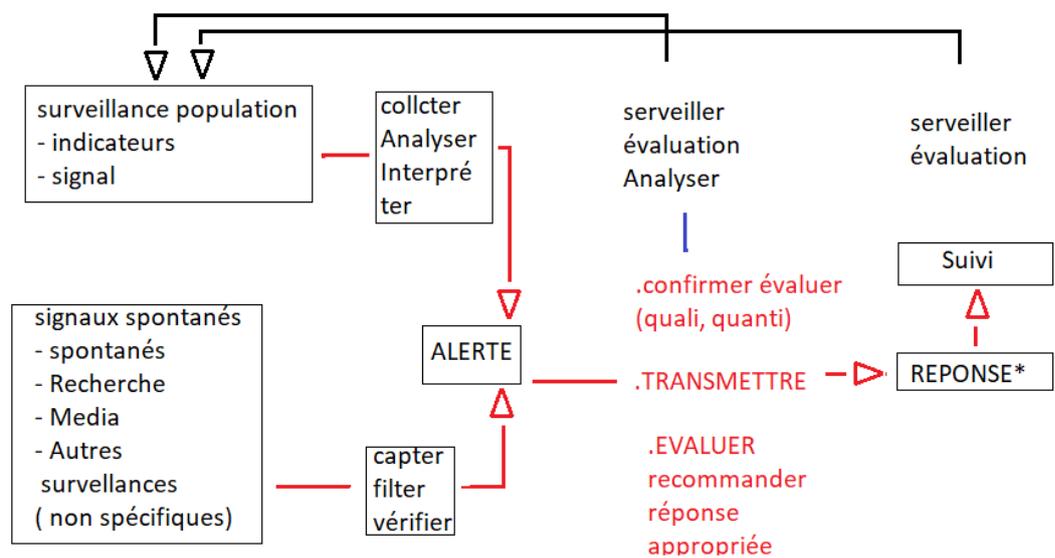
Dans une société de l'image et de l'information, la cartographie et l'analyse spatiale des événements sanitaires sont devenues indispensables pour la compréhension, la prise de décision et la communication. L'apparition des Systèmes D'information Géographique (SIG) a marqué un tournant dans la prise en compte de l'aspect spatial de l'information épidémiologique.

Cependant, la représentation géographique des données de surveillance doit d'abord correspondre à une réalité épidémiologique celle d'une cartographie du risque ou des facteurs influant sur les maladies. Malheureusement, la représentation spatiale de ces données traduit encore trop souvent la cartographie du réseau de recueil de l'information épidémiologique. La mise en place de système de surveillance pour l'alerte précoce permet de suivre en temps quasi réel la survenue des cas dans l'espace [24].

2.4.2 Actions en épidémiologie

Dans cette section, nous décrivons les actions essentielles en épidémiologie d'après :

- **Signal** : Le signal peut être défini comme « un événement de santé ou une situation d'exposition à un danger environnemental pouvant révéler une menace pour la santé publique » quant à elle correspond à « un signal validé pour lequel,
- **Alerte** : après évaluation du risque, il a été considéré qu'il représente une menace pour la santé des populations et qui nécessite une réponse adaptée.
- **Vigilance** : elle correspond à un système de surveillance orienté vers les effets indésirables ou inattendus des produits ou des activités humaines et en particulier médicales (médicaments, produits dérivés du sang, cosmétiques, composés radioactifs, alimentation, etc.)



* Réponse = décision de gestion

Figure 8 - les actions essentielles en épidémiologie

2.5 Prédiction en épidémiologie

Dans la littérature, il existe de nombreux modèles et approches permettant la prédiction en épidémiologie. La plupart de ces modèles utilisent une méthodologie courante en épidémiologie des maladies infectieuses (modèles à compartiments, modèles à base d'agents, etc.) intégrant des paramètres liés à la maladie. Ces modèles peuvent être plus ou moins complexifiés c.-à-d. adoptent d'autres paramètres tels que : l'âge des personnes, l'immigration, la mortalité naturelle, etc.). Certains modèles représentent la diffusion géographique de la maladie et utilise l'analyse spatiale des données.

En effet, l'objectif de la prédiction en épidémiologie permet de recueillir et analyser divers gros ensembles de données de terrain (médicales, géographiques, socio-économiques, etc.) et de mettre en place de nouveaux algorithmes et approches pour limiter la propagation de l'épidémie afin de sauver autant de vies humaines que possible ou d'arrêter complètement la propagation.

2.6 Modèles de prédiction basé sur l'apprentissage

Dans la suite, nous présentons les modèles de prédiction basé sur l'apprentissage les plus connus dans la littérature :

2.6.1 Arbre de décision

L'arbre de décision est l'outil le plus puissant pour la classification et la prédiction. Un arbre de décision est un organigramme similaire à une arborescence, où chaque nœud interne désigne un test sur un attribut, chaque branche représente un résultat du test et chaque nœud feuille (nœud terminal) détient une étiquette de classe. Il inclut la situation sous forme de branches avec les mêmes options graphiques différentes, avec des décisions potentielles aux deux extrémités des branches.

Les arbres de décision existent dans de nombreux domaines, tels que la médecine, l'intelligence d'affaires, l'exploration de données, la sécurité, etc. Dans la plupart des cas, l'arbre de décision peut affecter une population attribuée à un groupe homogène pour atteindre un objectif spécifique sans ignorer les caractéristiques discriminatoires du groupe. Les arbres de décision facilitent la lecture des situations complexes Cela vous permet d'exécuter rapidement les décisions prises.

Un arbre peut être « appris » en divisant l'assemblage source basé sur un test précieux attribué dans les sous-ensembles. Ce processus est répété sur chaque sous-ensemble dérivé de la partition récursive. La récursivité est terminée si le sous-ensemble à un nœud a la valeur totale des variables cible ou lorsque la séparation ajoute la valeur à la prédiction. La construction d'une classification d'arbres décisive n'exige pas de connaissance du domaine ou du paramètre et est donc reconnaissable pour la détection des connaissances. Les arbres décidés peuvent gérer de grandes données. Généralement, le classificateur d'arbres décisif à une bonne précision. L'induction de la prise de décision est un accès inductif typique à l'acquisition de connaissances pour la classification [25]. La Figure 9 représente un exemple d'un arbre de décision.

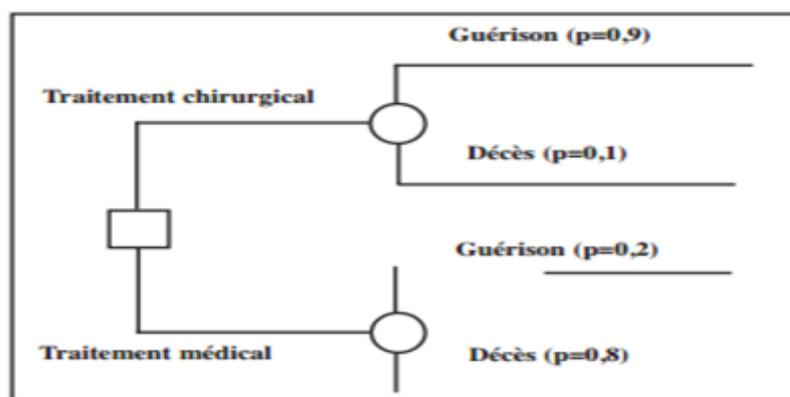


Figure 9 - Un exemple d'un arbre de décision médical

2.6.2 Régression linéaire

La régression linéaire est un algorithme d'apprentissage automatique basé sur l'apprentissage supervisé et elle est une sorte de technique de régression qui peut estimer le futur sur la base d'informations passées. La régression linéaire permet la prise en compte de variables dit « explicatives ou indépendantes ».

La régression modélise une valeur de prédiction cible basée sur des variables indépendantes. Il est principalement utilisé pour découvrir la relation entre les variables et les prévisions. Différents modèles de régression diffèrent selon le type de relation entre les variables dépendantes et indépendantes qu'ils envisagent et le nombre de variables indépendantes utilisées. En effet, la régression linéaire effectue la tâche de prédire une valeur de variable dépendante (y) basée sur une variable indépendante donnée (x). Ainsi, cette technique de régression découvre une relation linéaire entre x (entrée) et y (sortie).

La figure N illustre un exemple de la régression multiple comparant le nombre d'infectés observé par le nombre d'infectés prédit de la grippe saisonnière [26].

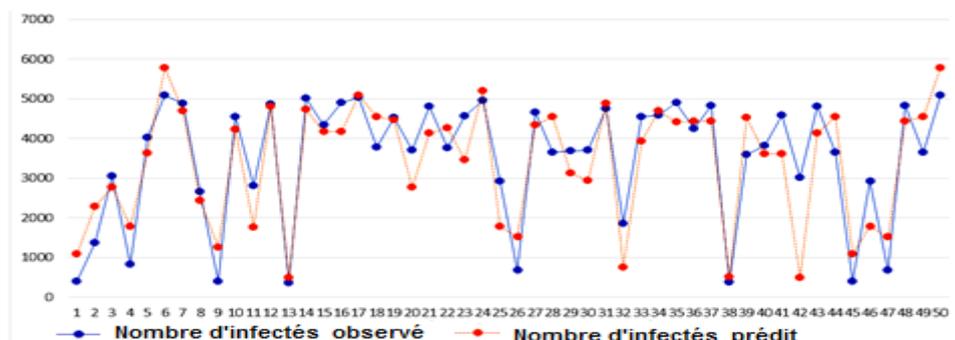


Figure 10 - Un exemple de la régression linéaire Multiple

Dans le tableau suivant, nous présentons quelques travaux de recherche utilisant les arbres de décisions et la régression :

Tableau 4 - travaux de recherche qui utilisent les arbres de décisions

Référence	Objectifs	Méthodes
[Venkatasubramaniam et al., 2017][25]	L'objectif principal de cet article est d'introduire et de décrire la famille des méthodes statistiques appelées arbres de décision, une famille particulièrement bien adaptée pour explorer les relations potentiellement non linéaires entre les variables et identifier des sous-groupes de population homogènes en ce qui concerne les résultats.	Arbre de décision
[Podgorelec,2002][26]	le but est de construire un arbre de décision de telle manière que le La précision de la classification pour les décisions les plus importantes est maximisée. Encore une fois, ce Problème n'est pas résolu de manière adéquate dans les méthodes classiques d'induction d'arbre de décision.	Arbre de décision

Tableau 5 - travaux de recherche qui utilisant la régression

Référence	Objectifs	Méthodes
[Varaksin,2002][27]	<p>Le document discute des différences dans l'interprétation de $Y = Y(x_i \{x_0, j\})$ fonctions entre des cas impliquant des variables prédictives corrélées et non corrélées.</p> <p>La construction et l'analyse de modèles de régression pour les données épidémiologiques et environnementales sont illustrés d'exemples.</p>	Régression Linéaire
[Kikuchi,2018][28]	<p>Le but de cet article est d'identifier le modèle linéaire à quantifier la cause de décès la plus importante à partir d'un ensemble de données distribué lié au patient et aux informations sur la maladie</p>	Régression Linéaire

2.7 Modèle de prédiction basé sur les RNA(s)

Les RNA sont très utilisés en prédiction des épidémies. Dans ce contexte, nous présentons dans la section suivantes les modèles les plus célèbres :

2.7.1 Auto Regressive Integrated Moving Average (ARIMA)

ARIMA est un outil d'analyse est essentiellement utilisé à des fins de prévision des valeurs futures, de détermination des valeurs manquantes dans une série de points ou d'identification de la structure de la série temporelle. Un modèle ARIMA s'exprime en fonction de la notation ARIMA (p, d, q) où p, d, q renvoient au nombre (ou l'ordre) de termes autorégressifs, de différenciation et de moyenne mobile dans le modèle ARIMA final.

(Par exemple : un modèle ARIMA (1,0,0) est un modèle qui comprend un terme autorégressif et aucun terme de différenciation ni de moyenne mobile).

Si une série chronologique a des modèles saisonniers, vous devez ajouter des termes saisonniers et elle devient SARIMA, abréviation de « Seasonal ARIMA ».

2.7.2 LSTM

La mémoire à long court terme (LSTM) est une architecture de réseau neuronal récurrent (RNN) spécifique qui a été conçue pour modéliser les séquences temporelles et leurs dépendances à longue portée avec plus de précision que les RNN conventionnels [29].

Les LSTM sont un type spécial de RNN, capable d'apprendre les dépendances à long terme et de mémoriser des informations pendant des périodes prolongées par défaut. Selon Olah, le modèle LSTM est organisé sous la forme d'une structure en chaîne. Cependant, le module répétitif a une structure différente. Au lieu d'un seul réseau neuronal comme un RNN standard, il comporte quatre couches d'interaction avec une méthode de communication unique [30].

2.7.3 Prophet

Prophet est une procédure de prévision des données de séries chronologiques basée sur un modèle additif où les tendances non linéaires sont adaptées à la saisonnalité annuelle, hebdomadaire et quotidienne, ainsi qu'aux effets des vacances. Cela fonctionne mieux avec des séries chronologiques qui ont de forts effets saisonniers et plusieurs saisons de données historiques. Prophet est robuste aux données manquantes et aux changements de tendance, et gère généralement bien les valeurs aberrantes [32].

Dans ce tableau, nous présentons quelques travaux de recherches utilisant les techniques citées auparavant :

Tableau 6 - travaux de recherches utilisant les techniques citées auparavant

Référence	Objectifs	Méthodes
[Wang, L,2020] [33].	<p>Une contribution clé de l'article est d'utiliser des données synthétiques générées par la théorie pour entraîner le neural réseau. Cela est rendu nécessaire par le fait que les données de surveillance des maladies sont rares. De plus, les données sont bruyantes et incomplètes. Nous surmontons les limitations en entraînant TDEFSI en utilisant les données générées par des simulations basées sur le calcul haute performance de processus causaux bien acceptés qui capturent dynamique épidémique.</p> <p>Epidemic Forecasting with Synthetic Information =(TDEFSI).</p>	LSTM RNN
[Yahia, N.B,2020][34].	<p>le but ici est de réutiliser les modèles déjà construits et formés sur les séries de porcelaine fois l'ensemble de données pour prédire l'épidémie de COVID-19 en Tunisie. Notre motivation est d'augmenter la taille des données d'entraînement car DNN nécessite souvent un grand nombre de données d'entraînement</p>	LSTM DNN CNN
[Pal, R., Sekh,2020] [35].	<p>Dans cet article, nous avons proposé une méthode guidée par l'IA pour prédire le risque de coronavirus à long terme spécifique au pays. Les principaux défis de ce problème sont les suivants : petit ensemble de données, données incertaines,</p>	LSTM

	données incertaines. Cela introduit une nouvelle façon de prédire une épidémie et d'établir une corrélation avec le risque d'un pays.	
--	--	--

2.8 Conclusion

Dans ce chapitre, nous avons d'abord introduit l'épidémiologie ainsi que ces différentes stratégies dans la recherche épidémiologique. Puis, nous avons expliqué la Prédiction en épidémiologie et ces modèles. Finalement, nous avons brièvement défini les modèles de prédiction basé sur les RNA.

Chapitre 3

Conception et Modélisation

3.1 Introduction

Prédire l'évolution future d'une maladie dans l'espace et dans le temps pour la population humaine est nécessaire pour améliorer les fonctionnalités du système de santé. En effet, la prédiction par l'utilisation des outils d'IA permettra de fournir des alertes précises afin de réagir rapidement face à une épidémie. Elle permet, également, aux épidémiologistes et aux responsables de santé publique de mettre en place des plans stratégiques pour empêcher la hausse des cas de transmission et de maintenir un niveau de transmission durablement faible, voire nul.

Dans ce chapitre, nous allons décrire notre système de prédiction, nous allons, également, présenter des organigrammes des méthodes utilisées et la démarche adaptée.

3.2 Objectif de notre système SYSPC

Notre objectif est d'élaborer un nouveau Système de Prédiction de la COVID-19 nommé SYSPC. Le but de ce dernier est de prédire la propagation du COVID-19 en Algérie et dans ses villes en utilisant trois méthodes : méthode classique ARIMA, apprentissage automatique : PROPHET, une autre méthode moderne en utilisant l'apprentissage profond et RNN : LSTM.

3.3 Description de notre système

Nous allons utiliser trois méthodes d'IA pour élaborer notre système. En effet, notre système est composé de 3 sous-systèmes, à savoir :

3.3.1 Système LSTM :

L'objectif de ce sous-système est de prédire l'évolution de la COVID-19 en Algérie en utilisant le RNA et l'apprentissage en profondeur. Nous avons choisi la méthode LSTM car cette méthode est bien adaptée au traitement et à la réalisation de prédictions basées sur des données de séries chronologiques. La méthode LSTM est basée sur des connexions de rétroaction. Elle permet non seulement de traiter des points de données uniques, mais également des séquences entières de données comme dans notre cas. Les modèles LSTM sont, également, capables de stocker des informations sur une période de temps, ce qui est utile lorsque nous traitons des séries temporelles ou des données séquentielles. Les modèles LSTM sont capables de stocker des informations sur une période de temps, ce qui est utile lorsque nous traitons des séries temporelles ou des données séquentielles.

L'organigramme présenté dans la figure 11 illustre le fonctionnement général de la LSTM

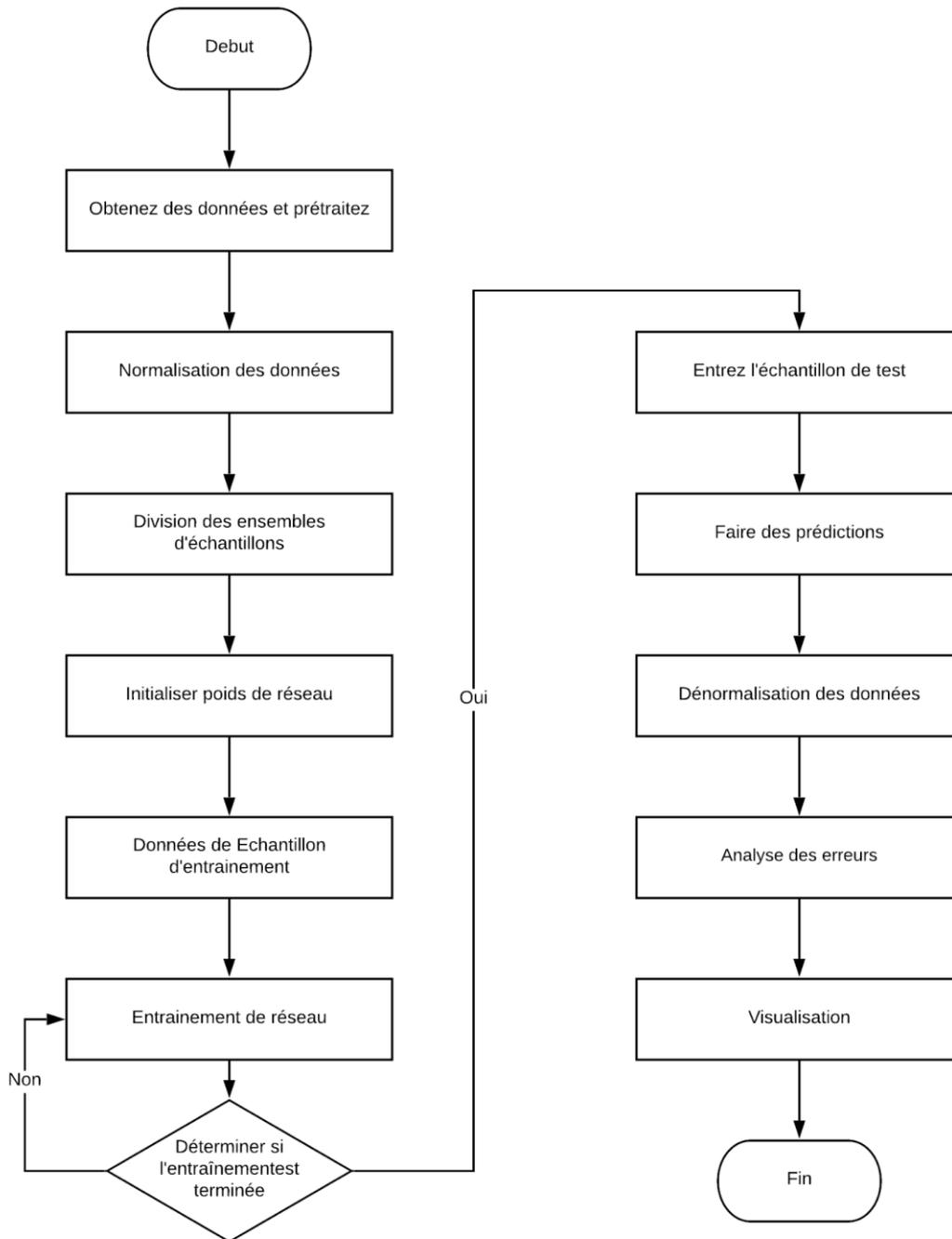


Figure 11 : Organigramme de la méthode LSTM.

3.3.2 Système ARIMA

L'objectif de ce sous-système est de prédire l'évolution de la COVID-19 en Algérie en utilisant de séries chronologiques classiques. Nous allons utiliser les méthodes classiques telles que ARIMA car ces dernières surpassent les méthodes d'apprentissage automatique et d'apprentissage en profondeur pour la prévision en une étape sur des ensembles de données univariés et sur la prévision en plusieurs étapes sur des ensembles de données univariés. Les méthodes d'apprentissage automatique et d'apprentissage en profondeur ne tiennent pas encore leur promesse de prévision de séries chronologiques univariées et il y a beaucoup de travail à faire.

L'organigramme présenté dans la figure 12 illustre le fonctionnement général de la méthode ARIMA :

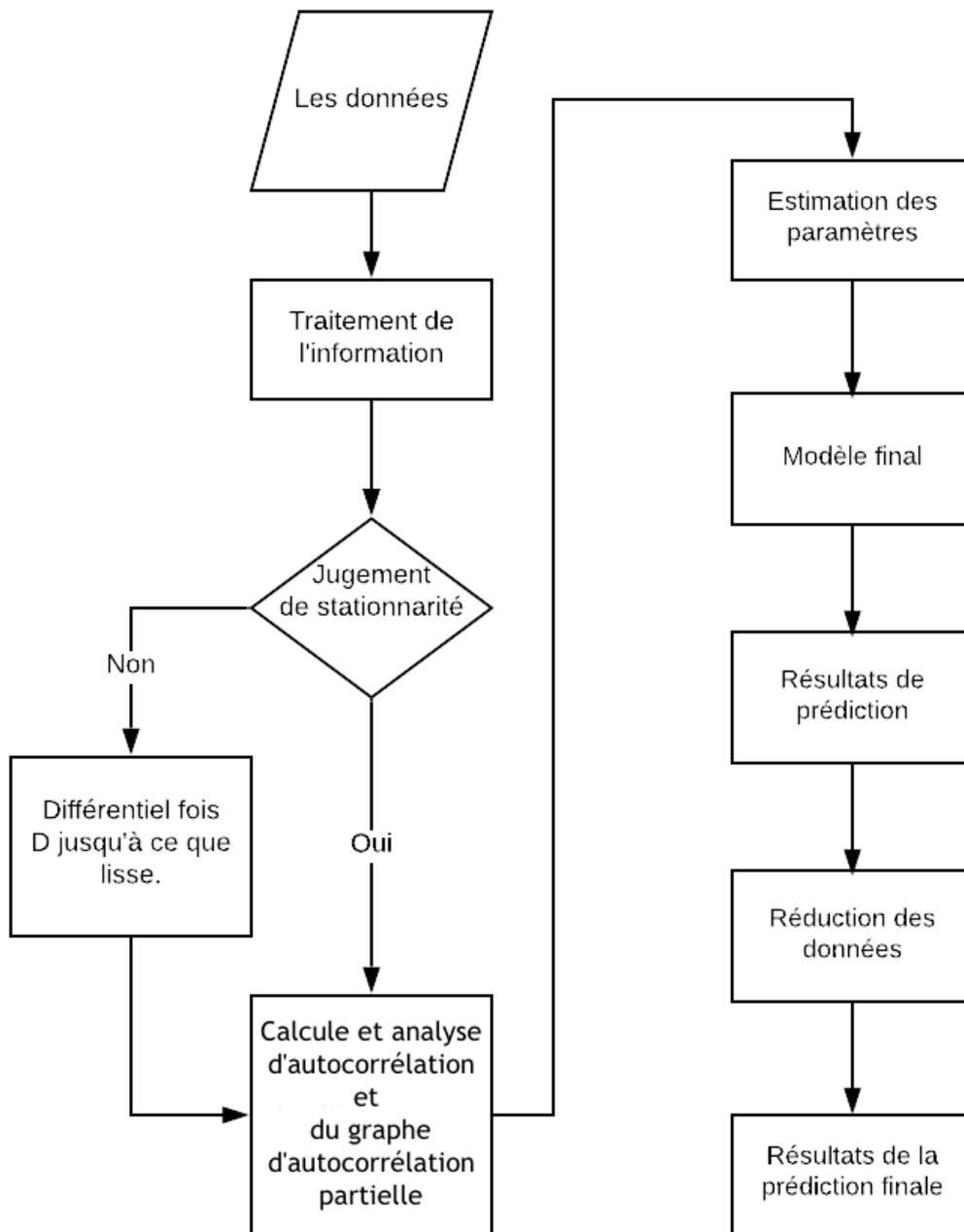


Figure 12 : Organigramme de la méthode ARIMA

3.3.3 Système Prophet

Prophet est une procédure de prévision des données de séries chronologiques basée sur un modèle additif où les tendances non linéaires sont adaptées à la saisonnalité annuelle, hebdomadaire et quotidienne, ainsi qu'aux effets de vacances. Cela fonctionne mieux avec des séries chronologiques qui ont de forts effets saisonniers et plusieurs saisons de données historiques [36]. Prophet est robuste aux données manquantes et aux changements de tendance, et gère généralement bien les valeurs aberrantes. Nous l'avons utilisé car il s'agissait de prévisions précises et rapides, entièrement automatiques et ajustables.

L'organigramme présenté dans la figure 13 illustre le fonctionnement général de Prophet :

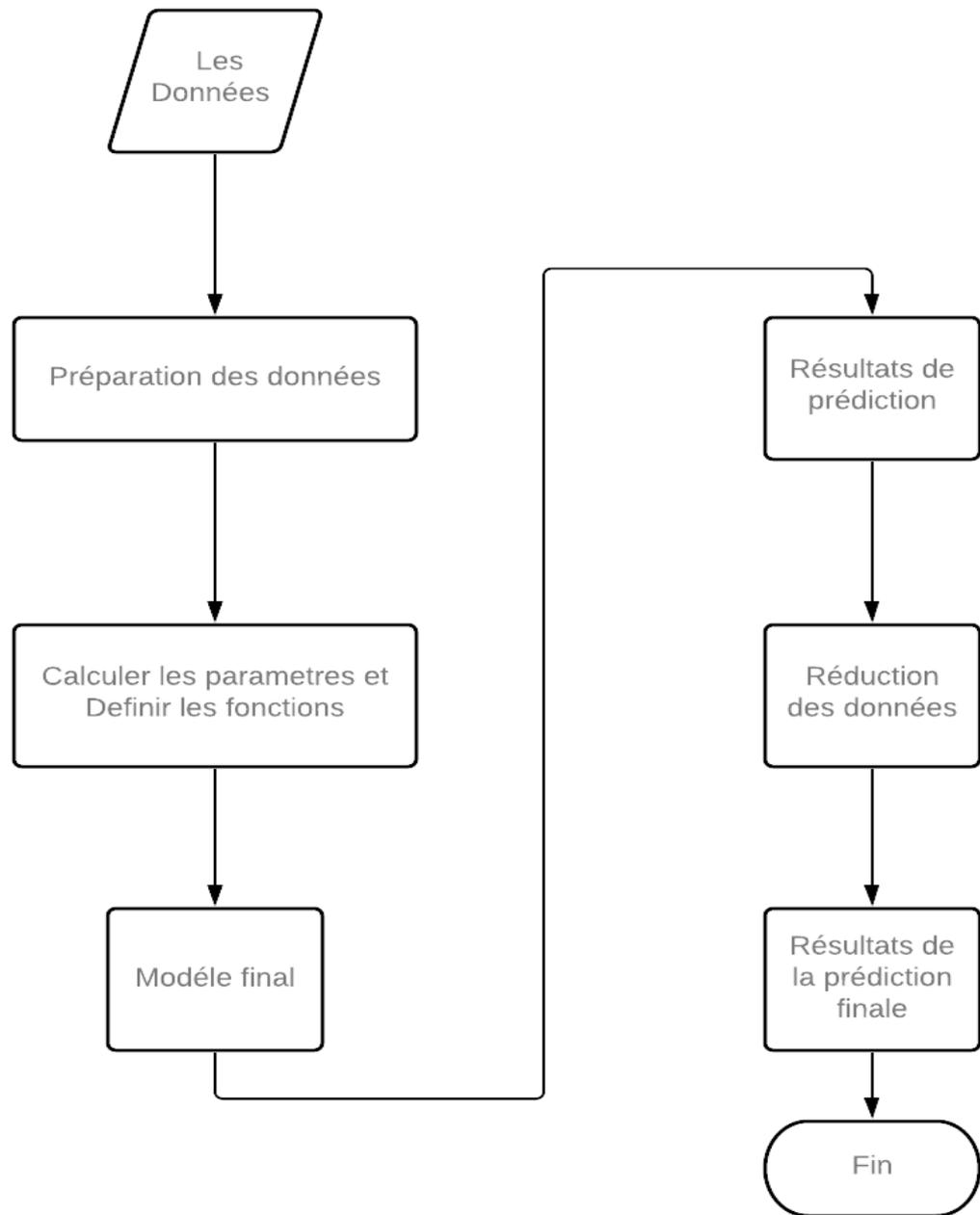


Figure 13 : Organigramme de la méthode Prophet.

3.4 Description de Base de données

La COVID-19 est une infection respiratoire virale apparue en 2019 en Chine. C'est l'un des virus qui se propagent rapidement avec des risques de mortalité élevés chez certaines personnes, il est très contagieux et cause de différents symptômes. Dans ce contexte, plusieurs études et recherches ont été effectuées afin de contrôler et de suivre la maladie et de prédire son évolution future.

La base de données utilisée par la présente étude est composée des cas confirmés de la COVID-19 en Algérie entre 01/05/2020 et 25/04/2021. Elle a été subdivisée en deux ensembles : le premier ensemble (base d'apprentissage) est de 70% pour l'apprentissage du modèle, le deuxième ensemble (base de test) est de (30%) ont été utilisés pour la validation du modèle.

3.5 Démarche globale de SYSPC

Dans la suite, nous présentons la démarche globale adoptée par notre système SYSPC

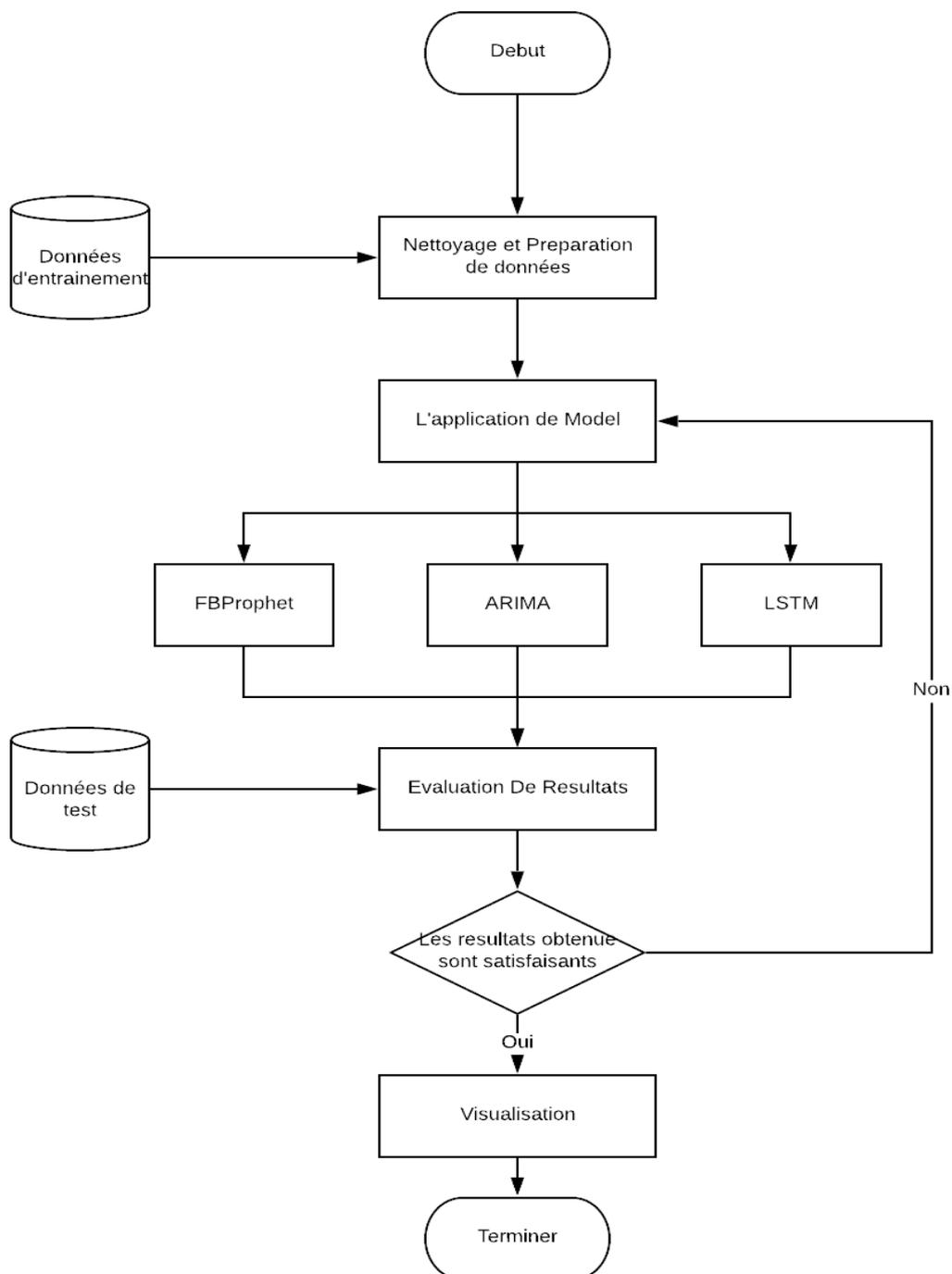


Figure 14 : Démarche globale de SYSPC.

3.6 Langage UML

UML, ou Unified Modeling Language, est un langage de modélisation standard composé d'un ensemble intégré de diagrammes qui a été développé pour faciliter la spécification, la visualisation, la construction et la preuve des artefacts du système logiciel, ainsi que la modélisation commerciale et d ' autres systèmes non logiciels.

3.7 Diagramme de séquence

Le diagramme de séquence, également appelé diagramme d'événements, représente le flux des communications dans le système. Il décrit l'interaction entre deux lignes de vie comme une série d'événements ordonnés dans le temps, comme si ces lignes de vie étaient présentes en même temps.

La figure suivante représente le diagramme de séquence de SYSPC :

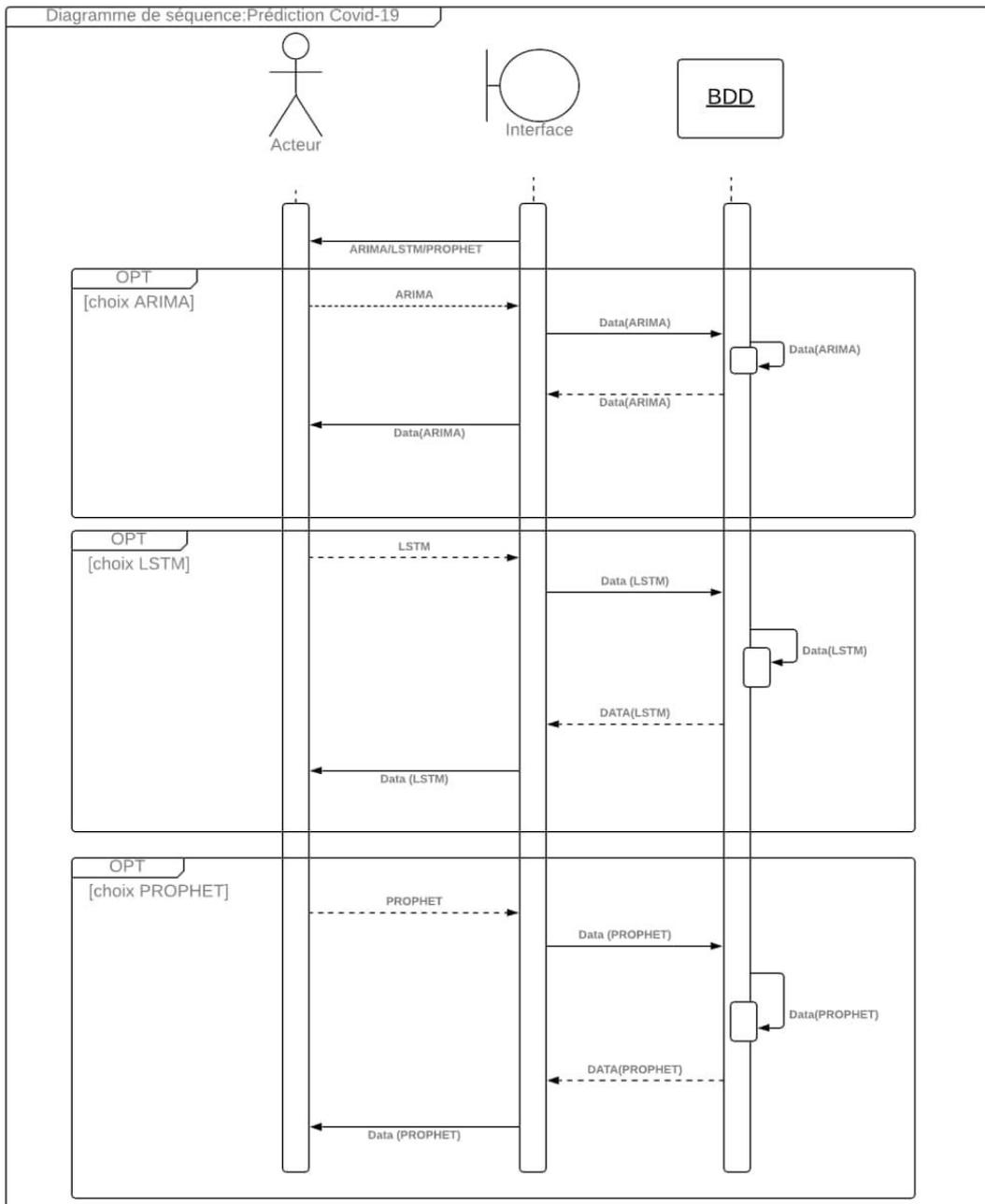


Figure 15 : Diagramme de séquence de SYSPC.

3.8 Description du diagramme

Tout d'abord, le client accède à l'interface API (Application Programming Interfaces). Il demande une prédiction COVID-19 dans l'Algérie. Ensuite, il choisit entre trois méthodes : Prophet, LSTM et ARIMA selon son besoin. Puis, il effectue une requête qui sera envoyée à la base de données pour récupérer les données et les afficher sur l'interface, sur une carte ou une table.

3.9 Conclusion

Dans le présent chapitre, nous avons d'abord introduit le système suggéré SYSPC. Puis, nous avons expliqué d'une façon détaillée chaque sous système de SYSPC. Finalement, nous avons présenté la démarche globale et le diagramme de séquence de notre système.

Dans le chapitre suivant, nous allons mettre en œuvre le système SYSPC.

Chapitre 4

Mise en oeuvre

4.1 Introduction

Dans ce chapitre, nous allons présenter les outils de développement que nous avons utilisés pour développer notre système. Nous allons, également, présenter les différentes étapes et les résultats obtenus par le système SYSPC.

4.2 Outils de développement

Au cours de ce projet, nous avons fait recours à plusieurs outils de développement :

4.2.1 Langage de programmation PYTHON

Le langage de programmation Python a été créé en 1989 par Guido van Rossum, aux Pays-Bas. Le nom Python vient d'un hommage à la série télévisée Monty Python's Flying Circus dont G. van Rossum est fan. La première version publique de ce langage a été publiée en 1991.

Ce langage de programmation présente de nombreuses caractéristiques intéressantes, il est : multiplateforme, gratuit, un langage de haut niveau, orienté objet, simple et très utilisé en bio-informatique et plus généralement en analyse de données. Il demande relativement peu de connaissance sur le fonctionnement d'un ordinateur pour être utilisé, c'est un langage interprété.

Le tableau suivant récapitule les outils utilisés

Tableau 7 - représente les outils que nous avons utilisés.

Lagunage	Bibliothèques et packages	Additionnel bibliothèques et packages	Éditeur de texte
Python	TensorFlow Statsmodels Fbprophet	Datetime Numpy Pandas Matplotlib Math Sklearn	Jupyter notebook

4.3 Etapes de développements

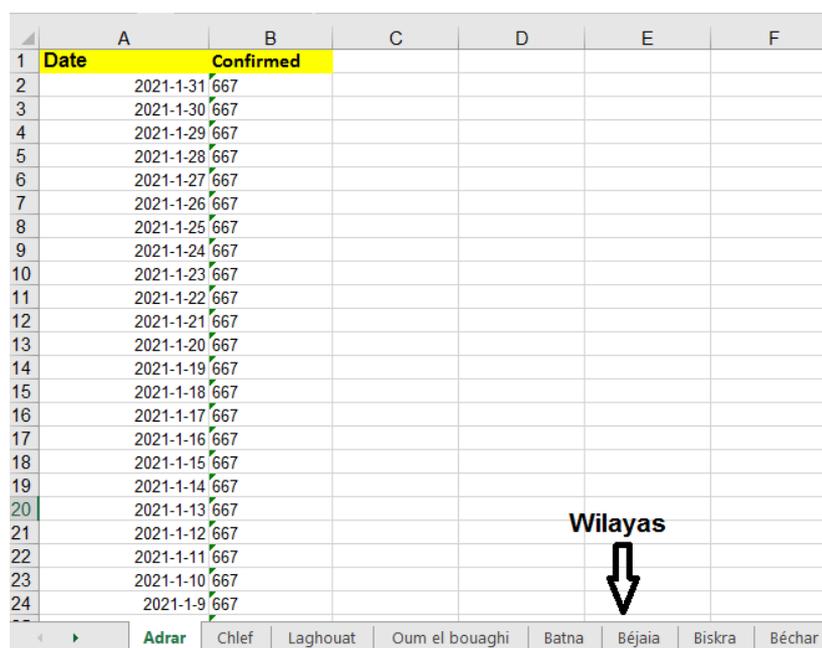
Notre système vise à aider les responsables de la santé publique à mener un processus de surveillance épidémiologique en essayant de prévoir l'évolution de la propagation des épidémies en vue de les contrôler et donc pouvoir maîtriser une situation épidémiologique.

En effet, nous avons développé un système de prédiction basé sur les méthodes d'Intelligence Artificielle à savoir : ARIMA, LSTM et Prophet pour prédire l'évolution future de la COVID-19 en Algérie. Pour assurer une bonne analyse par les personnels de santé publique, nous avons présenté les résultats de prédiction sous forme de carte géographique de prédiction.

4.3.1 Base de données utilisées

Les données utilisées dans notre étude ont été obtenues à partir d'un fichier CSV fourni par Algeria Coronavirus Tracker API pour les données du COVID 19 en Algérie.

Ces données représentent les données en entrée (input) de notre système SYSPC, elles sont illustrées sur la figure 17 qui représente un aperçu des données de COVID-19 en Algérie pour les différentes wilayas. Une description détaillée de la base de données utilisée est dans le chapitre précédent (Voir section 3.4).



	A	B	C	D	E	F
1	Date	Confirmed				
2	2021-1-31	667				
3	2021-1-30	667				
4	2021-1-29	667				
5	2021-1-28	667				
6	2021-1-27	667				
7	2021-1-26	667				
8	2021-1-25	667				
9	2021-1-24	667				
10	2021-1-23	667				
11	2021-1-22	667				
12	2021-1-21	667				
13	2021-1-20	667				
14	2021-1-19	667				
15	2021-1-18	667				
16	2021-1-17	667				
17	2021-1-16	667				
18	2021-1-15	667				
19	2021-1-14	667				
20	2021-1-13	667				
21	2021-1-12	667				
22	2021-1-11	667				
23	2021-1-10	667				
24	2021-1-9	667				

Wilayas

Adrar Chlef Laghouat Oum el bouaghi Batna Bénéjaia Biskra Béchar

Figure 16 : Représentation tabulaire des données de la COVID-19 en Algérie

4.4 Résultat de système SYSPC

Dans cette section, nous présentons les résultats obtenus par notre système SYSPC. Nous illustrons les graphes de comparaison entre les données réelles et les données prédites de l'Algérie et des wilayas ayant le nombre de cas COVID-19 confirmé élevés. Enfin, nous présentons la carte géographique de prédiction de la COVID-19.

4.4.1 Représentation graphique des résultats

La figure 17 présente les deux courbes (cas confirmés COVID-19 et les cas prédits) par les trois méthodes ARIMA, LSTM et Prophet de toute l'Algérie. Nous remarquons que l'écart entre les deux courbes est petit.

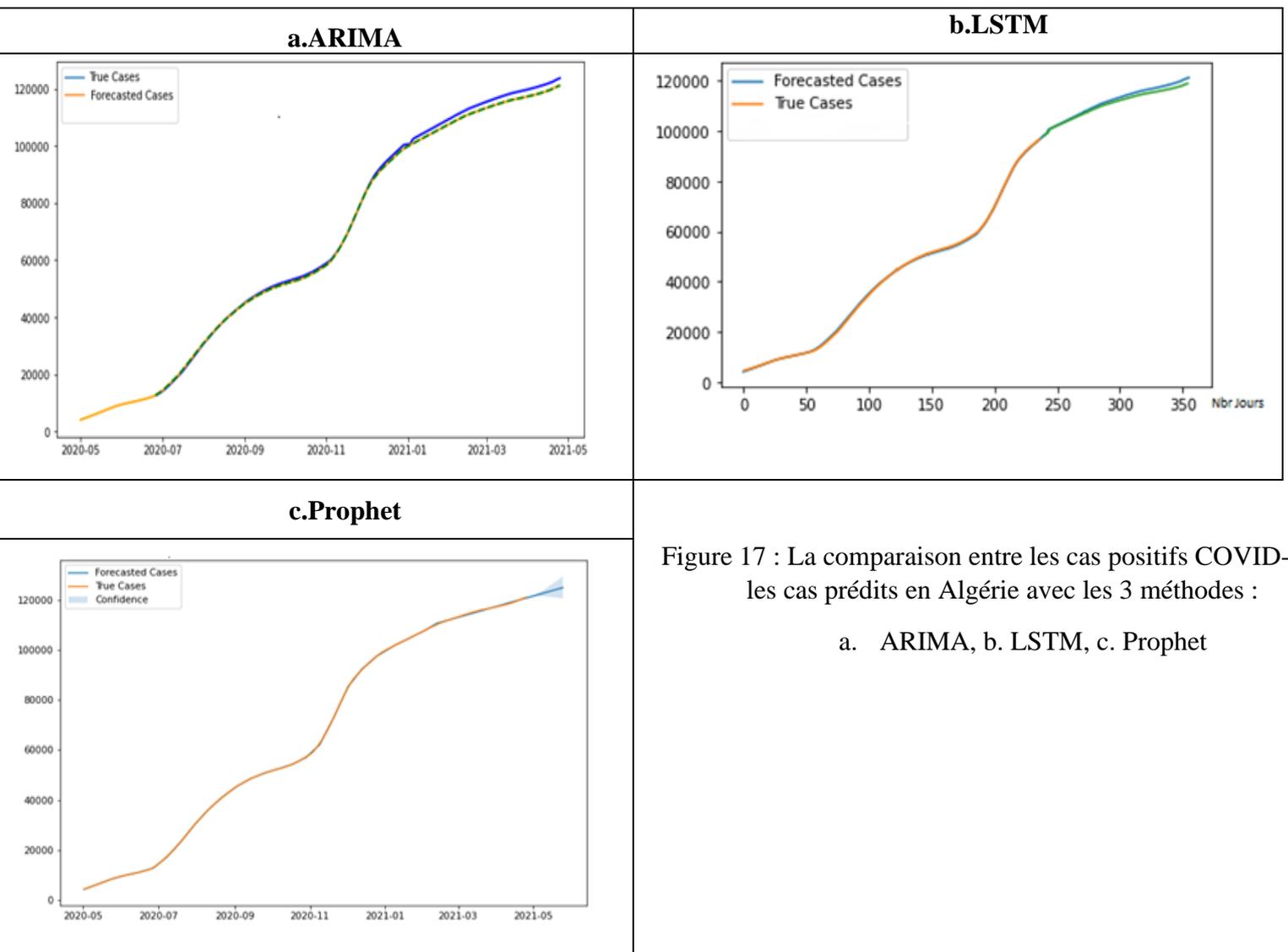


Figure 17 : La comparaison entre les cas positifs COVID-19 et les cas prédits en Algérie avec les 3 méthodes :

a. ARIMA, b. LSTM, c. Prophet

La figure 18 présente les deux courbes (cas confirmés COVID-19 et les cas prédits) par les trois méthodes pour la wilaya d'Alger. Il est à noter que la wilaya d'Alger a enregistré un nombre important des cas COVID-19. Nous remarquons que l'écart entre les deux courbes est petit.

Le premier cas positif a été déclaré dans la wilaya de Blida en mois de Mars 2020. Cette ville a recensé plus de 130 morts par ce virus, et plus de 4.400 cas d'atteintes, selon les chiffres du Comité scientifique de suivi de l'évolution de la pandémie du Coronavirus.

La figure 19 illustre les deux courbes (cas confirmés COVID-19 et les cas prédits) par les trois méthodes pour la wilaya de Blida. L'écart entre les deux courbes est moins important.

La wilaya d'Oran est la deuxième ville d'Algérie, la population de la wilaya d'Oran en 2018 est de 2 053 204 habitants. Durant la pandémie du COVID-19. Cette ville a enregistré un nombre important de cas confirmés. En effet, l'étude de la prédiction de l'évolution future de la COVID-19 semble importante. Nous montrons dans la figure 20 la comparaison entre les courbes (cas confirmés COVID-19 et les cas prédits) par les trois méthodes pour la wilaya d'Oran. Il est très clair que l'écart entre les deux courbes est petit.

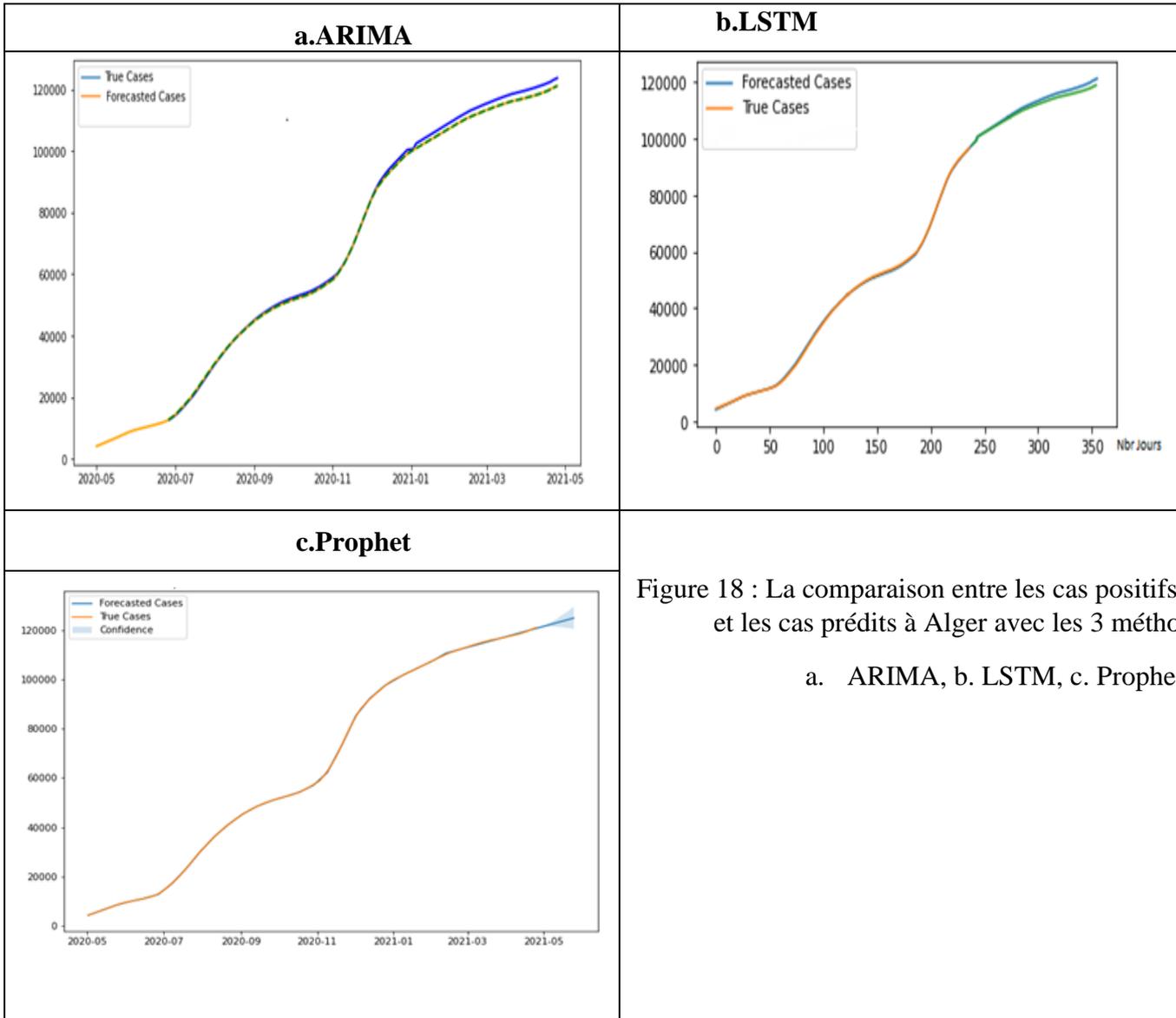


Figure 18 : La comparaison entre les cas positifs COVID-19 et les cas prédits à Oran avec les 3 méthodes :

a. ARIMA, b. LSTM, c. Prophet

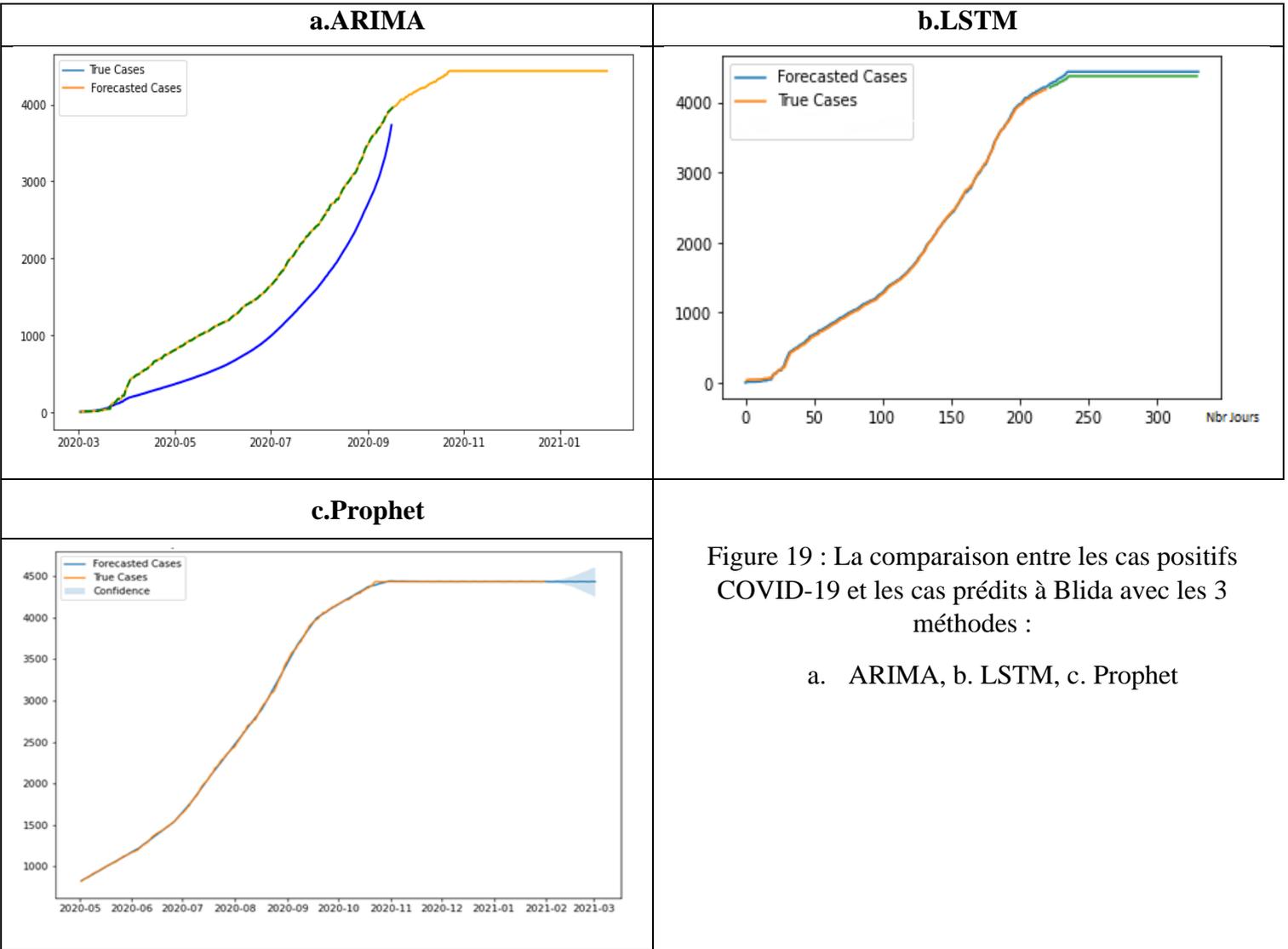
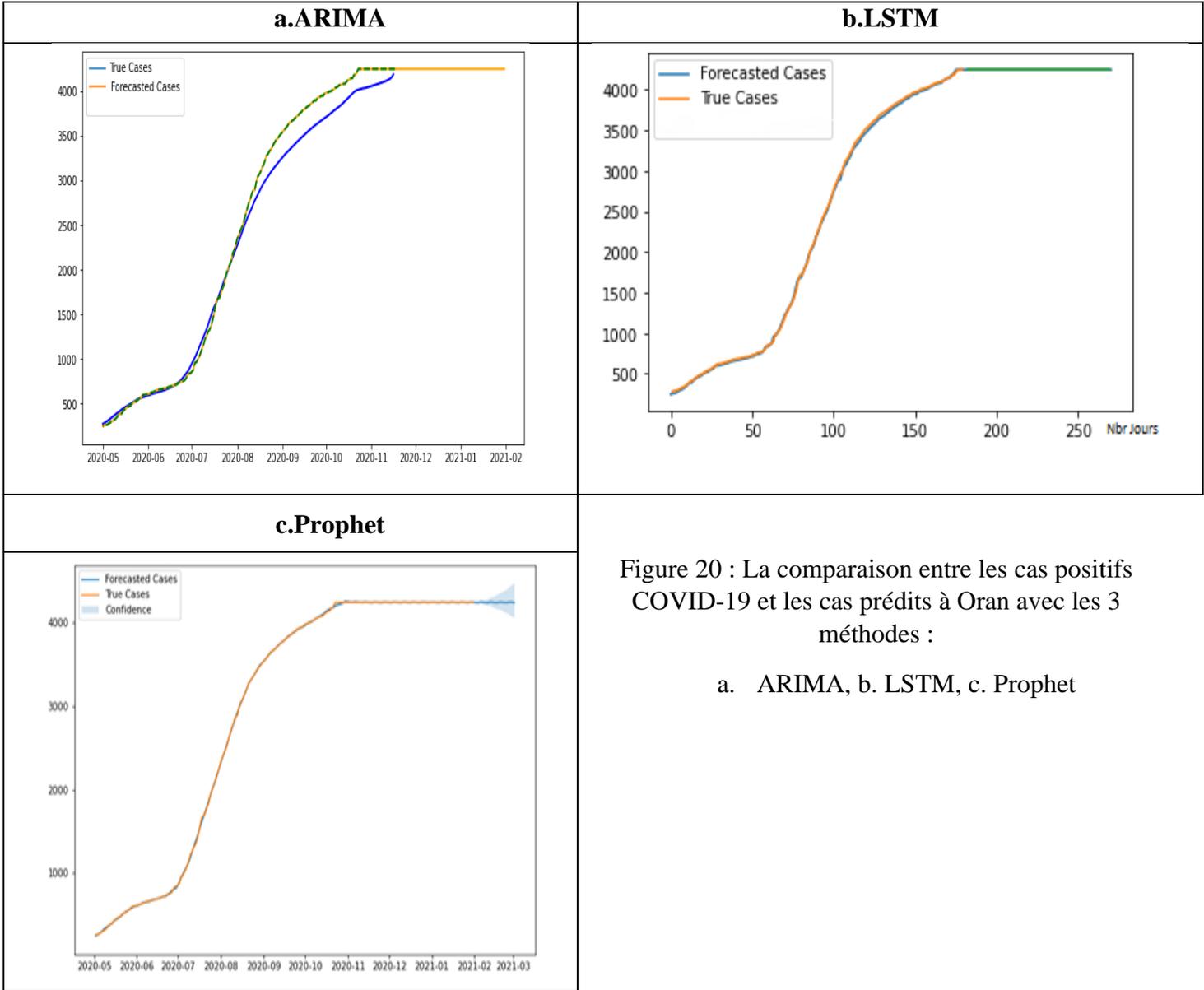


Figure 19 : La comparaison entre les cas positifs COVID-19 et les cas prédits à Blida avec les 3 méthodes :

a. ARIMA, b. LSTM, c. Prophet



4.4.2 Représentation cartographique des résultats

La représentation cartographique des données réelles de la COVID-19 en Algérie est donnée par la figure 21

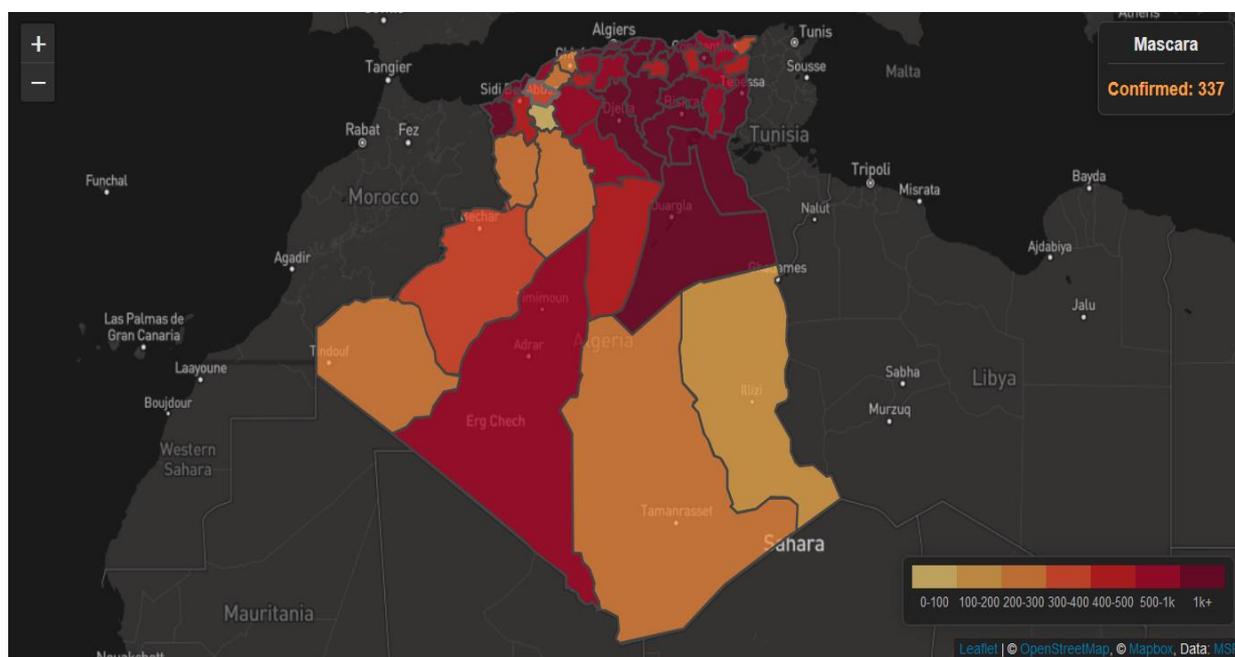


Figure 21 : la propagation du COVID-19 en Algérie entre 05/01/2020 et 04/25/2021

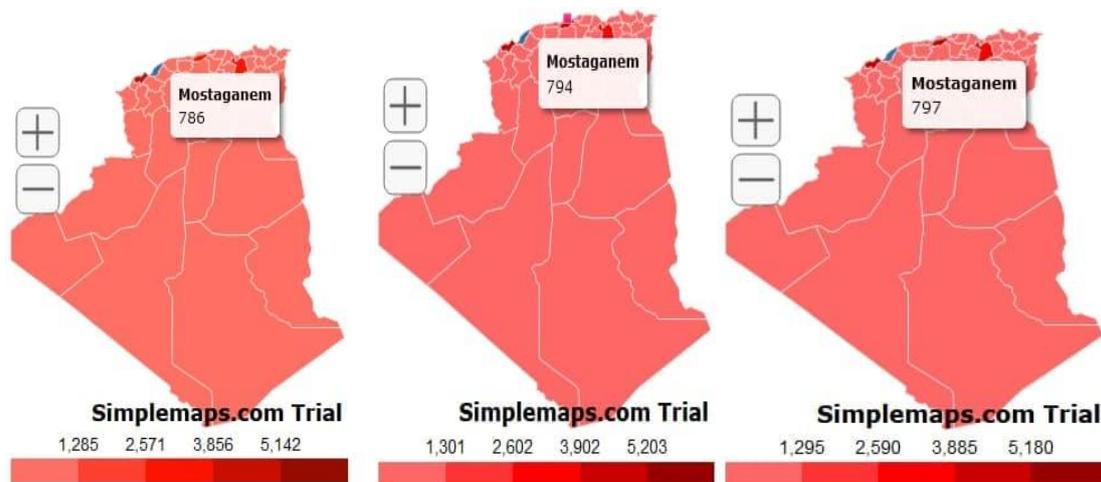


Figure 22 : la propagation du COVID-19 en Mostaganem.

4.5 Validation des méthodes utilisées

Il existe plusieurs approches dans la littérature pour évaluer la capacité de prédiction d'un modèle. Le but de l'étape de validation est de déterminer la capacité de prédiction des modèles développés pour cette utilisation. En effet, nous avons utilisé la méthode MAPE (Mean Absolute Percentage Error) qui signifie erreur moyenne absolue

en pourcentage. Elle est connue comme l'une des métriques les plus utilisées pour mesurer la précision des prévisions d'un modèle. MAPE permet de juger si le système de prévision développé est bon.

La figure 22 et la figure 23 illustrent la comparaison entre les cas positifs COVID-19 prévus et réels des 15 derniers jours par les modèles ARIMA et Prophet pour l'Algérie. Le pourcentage d'erreur (MAPE) est surligné en orange, par exemple dans la figure 22 pour 20 Avril 2021, le modèle ARIMA a une erreur de 1,94%. Dans la figure 23 pour 19 Mai 2021, le modèle Prophet a une erreur de 1,5%. Nous remarquons que les valeurs de MAPE sont plus petites, cela indique un meilleur ajustement.



Figure 23 : Comparaison entre les cas positifs COVID-19 prévus et réels des 15 derniers jours par le modèle ARIMA pour l'Algérie

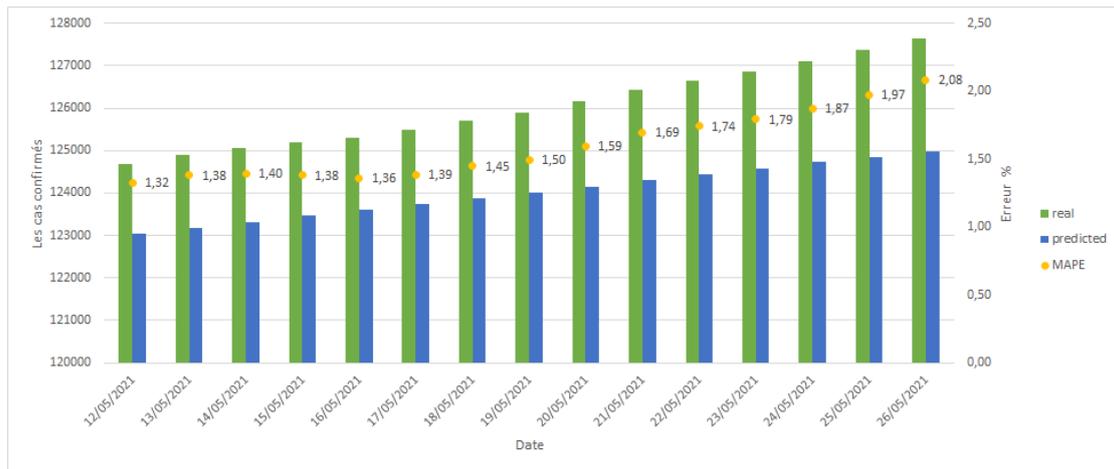


Figure 24 : Comparaison entre les cas positifs COVID-19 prévus et réels des 15 derniers jours par le modèle Prophet pour l'Algérie

4.6 Conclusion :

Dans le présent chapitre, nous avons tout d'abord introduit les outils de développement utilisés pour développer le système SYSPC. Ensuite, nous avons présenté les résultats de prédiction obtenus par le système suggéré (courbes et cartographique). Finalement, nous avons utilisé le modèle MAPE de validation pour valider nos modèles de prédiction.

Conclusion Générale

La propagation de la COVID-19 dans le monde, et en Algérie en particulier, a connu une augmentation significative depuis l'année dernière, et cette augmentation a rendu leur suivi très difficile. Ces difficultés doivent donc être surmontées afin d'aider les agents de santé à surveiller efficacement cette épidémie et à prendre des décisions pour limiter leur propagation.

Dans ce contexte, notre travail de Master a pour objectif de développer un nouveau système de prédiction de la COVID-19 en Algérie basé sur des méthodes d'IA. Le système SYSPC développé est basé sur les méthodes de prévisions suivantes : ARIMA, LSTM et Prophet. Ces méthodes sont le plus souvent utilisées pour prédire l'évolution future des maladies infectieuses. Afin de mettre en place notre système, nous avons défini son entrée qui est une base de données des cas confirmés de COVID-19 de toutes les wilayas d'Algérie. Ensuite, nous avons appliqué chaque méthode sur la base de données pour générer les sorties qui représentent les prédictions de la COVID-19. Chaque méthode suit une procédure différente afin d'obtenir des résultats. Ces résultats sont représentés sous forme de courbes et cartes géographique.

Enfin, les sorties de notre système (les résultats obtenus) sont validées par la méthode MAPE qui est très connu par sa souplesse, simplicité et sa précision. Les résultats de la comparaison des valeurs réelles et prédites par MAPE sont très satisfaisants et nous avons abouti un meilleur ajustement des données réelles et prédites.

Dans des travaux futurs, nous comptons améliorer les fonctionnalités de notre système SYSPC par son intégration avec des résultats de simulation par des modèles d'épidémie pour enrichir la base de données et puis tester les méthodes.

Bibliographie

- [1]. Feigenbaum, E.A., Barr, A. and Cohen, P.R. eds., 1981. The handbook of artificial intelligence.
- [2]. Intelligence artificielle : définition, applications et enjeux.
<https://youmatter.world/fr/definition/intelligence-artificielle-definition-enjeux/>
- [3]. L'intelligence artificielle, ses avantages et inconvénients.
<https://www.axiocode.com/avantages-inconvenients-intelligence-artificielle/>
- [4]. Qu'est-ce que l'intelligence artificielle?
<https://www.netapp.com/fr/artificial-intelligence/what-is-artificial-intelligence/>
- [5]. AI applications: Top 10 Artificial Intelligence Applications.
<https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/artificial-intelligence-applications>
- [6]. What is Machine Learning? A Definition.
<https://www.expert.ai/blog/machine-learning-definition/>
- [7]. Machine Learning: Supervised Learning.
<https://towardsdatascience.com/machine-learning-supervised-learning-8c5331c485f5>
- [8]. C'est quoi la classification supervisée ? | Apprentissage supervisé.
<https://www.kongakura.fr/article?id=Classification%20supervis%C3%A9e>
- [9]. L'apprentissage supervisé – Machine Learning.
<https://mrmint.fr/apprentissage-supervise-machine-learning>
- [10]. Apprentissage non supervisé : Définition, Exemples, Fonctionnement ,
<https://www.24pm.com/117-definitions/531-apprentissage-non-supervise>
- [11]. A quoi sert le clustering des données?
<https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203345-clustering-definition/>

- [12]. Réduction de dimensionnalité: Définition, Exemples, Fonctionnement. <https://www.24pm.com/ia-par-metier/data/117-definitions/338-reduction-de-dimensionnalite>
- [13]. Apprentissage semi-supervisé - Semi-supervised learning
Apprentissage semi-supervisé https://fr.qaz.wiki/wiki/Semi-supervised_learning
- [14]. Machine Learning : L'apprentissage par renforcement. https://www-igm.univ-mlv.fr/~dr/XPOSE2014/Machin_Learning/D_App_Renforcement.html
- [15]. Apprentissage profond [Deep Learning] : Définition, concepts et exemples. <https://superdatacamp.com/machine-learning/apprentissage-profond/>
- [16]. Jordan, M., 1989. Serial Order: A Parallel Distributed Processing Approach Advances in Connectionist Theory. Elman, J.L., 1990. Finding structure in time. *Cognitive science*, 14(2), pp.179-211.
- [17]. Définition Artificial Neuronal Network. <https://actualiteinformatique.fr/intelligence-artificielle/definition-de-artificial-neuronal-network>
- [18]. Stéphanie Boini , Support de cours d'épidémiologie, La Faculté de Médecine de Nancy, UNIVERSITÉ DE LORRAINE, France, 2009
- [19]. **Support** de cours. L'épidémiologie, Université Médicale Virtuelle Francophone, 2012
- [20]. **Support** de cours. L'épidémiologie, Faculté de Médecine Paris-Ile-de-France-Ouest, Université de Versailles Saint-Quentin-en-Yvelines, 2005
- [21]. **Support** de cours. L'épidémiologie, Université Médicale Virtuelle Francophone, 2012
- [22]. Astagneau, P. and Ancelle, T., 2011. *Surveillance épidémiologique: Principes, méthodes et applications en santé publique*. Lavoisier.
- [23]. Hartshorn, S., 2016. Machine learning with random forests and decision trees: A visual guide for beginners.
- [24]. Younsi, F.Z., Bounnekar, A., Hamdadou, D. and Boussaid, O., 2019. Integration of Multiple Regression Model in an Epidemiological Decision Support System. *International Journal of Information Technology & Decision Making*, 18(06), pp.1755-1783.

- [25]. Venkatasubramaniam, A., Wolfson, J., Mitchell, N., Barnes, T., JaKa, M. and French, S., 2017. Decision trees in epidemiological research. *Emerging themes in epidemiology*, 14(1), pp.1-12.
- [26]. Podgorelec, V., Kokol, P., Stiglic, B. and Rozman, I., 2002. Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26(5), pp.445-463.
- [27]. Varaksin, A.N. and Panov, V.G., 2020. Linear Regression Models in Epidemiology. *arXiv preprint arXiv:2002.08113*.
- [28]. Kikuchi, H., Hamanaga, C., Yasunaga, H., Matsui, H., Hashimoto, H. and Fan, C.I., 2018. Privacy-preserving multiple linear regression of vertically partitioned real medical datasets. *Journal of Information Processing*, 26, pp.638-647.
- [29]. Le, X.H., Ho, H.V., Lee, G. and Jung, S., 2019. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water*, 11(7), p.1387.
- [30]. Le, X.H., Ho, H.V., Lee, G. and Jung, S., 2019. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* 11, 1387.
- [31]. Graves, A., Mohamed, A.R. and Hinton, G., 2013, May. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). Ieee.
- [32]. Prophet forecasting at scale. <https://facebook.github.io/prophet/>
- [33]. Wang, L., Chen, J. and Marathe, M., 2020. TDEFISI: theory-guided deep learning-based epidemic forecasting with synthetic information. *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, 6(3), pp.1-39.
- [34]. Yahia, N.B., Kandara, M.D. and Saoud, N.B.B., 2020. Deep Ensemble Learning Method to Forecast COVID-19 Outbreak.
- [35]. Pal, R., Sekh, A.A., Kar, S. and Prasad, D.K., 2020. Neural network-based country wise risk prediction of COVID-19. *Applied Sciences*, 10(18), p.6448.
- [36]. mazon web services 3 mars 2021 _ Amazone forecast manuel du développeur