Ecole Nationale Supérieure d'Informatique
et d'Analyse des Systèmes

E-PREDICTIONS
DESIGNED BY IWIM TEAM

APPROCHES AXÉES SUR LES DONNÉES POUR LA PRÉDICTION DES CAS DE CONTAMINATIONS PAR COVID-19 AU MAROC

Rapport de Projet Intégré du Semestre 4

E-PREDICTIONS

Présenté par:

Arnaud YARGA
Yassine ATMANI
Zakarya ZAITAR
Chadi LAOULAOU

Supervisé Par:

Pr. Abdellatif KOBBANE, (PhD)

Jurés:
Pr. XXXXXXXXXXXXX, (PhD)
Pr. XXXXXXXXXXXXX, (PhD)

DEDICACE

Ce fruit de notre travail est humblement dédié à tous nos précieux trésors dans la vie .

À nos parents qui nous ont inspiré des idées de vie plus élevées

Pour leurs prières,

Pour leurs amours,

Et pour leur patience sans fin, ils

sont le paradis sur terre pour nous.

À nos sœurs et frères, leur soutien, leur confiance

et leurs amours.

À nos chères familles,

à toute personne qui nous soutenu durant notre

parcours A nos amis de ...

Yassine,Arnaud,Zakarya,Chadi

REMERCIEMENTS

Au terme de ce projet intégré du Semestre 4, on remercie Allah pour avoir fait en sorte que nous puissions mener à bien notre projet. Nous tenons à exprimer nos sincères gratitude au Dr Abdellatif KOBBANE, notre encadrant, pour son soutien et ses conseils continus au cours de cette période. Son expertise précieuse, ses commentaires opportuns et sa grande patience nous ont permis de terminer ce projet Intégré du Semestre 4.

Nous sommes également profondément reconnaissant envers notre école L'ENSIAS, et notre département IWIM et ses professeurs, pour tout le travail acharné qu'ils ont accompli au cours de ces deux années.

Nous remercions également toute personne qui a contribué de près ou de loin à l'accomplissement de notre travail.

Dernier point mais non le moindre, nous voudrons remercier les membres du jurés d'avoir accepté d'évaluer notre travail.

RÉSUMÉ

La gravité de la pandémie du coronavirus a poussé les gouvernements de plus de 200 pays à travers le monde à prendre des mesures préventives drastiques, au détriment de leurs économies. Par conséquent, la crise du COVID-19 a le potentiel de générer un impact social et économique durable et dommageable. Au Maroc, certains secteurs ont montré des signes précoces de vulnérabilité tels que le tourisme, les transports et la logistique dans les chaînes d'approvisionnement, mais aussi – et plus difficile à mesurer – le secteur informel transversal. Afin d'éradiquer la pandémie de COVID-19, toutes les armes à disposition de l'humanité doivent être exploitées. Partout dans le monde, les technologies telles que le Big Data,l'IA,Machine learning ou Deep learning sont déjà largement utilisées pour tenter d'enrayer la propagation du coronavirus. À présent, ce sont les données ouvertes ou Open Data qui pourraient jouer un rôle majeur dans ce combat.

Le but de ce projet est de tirer parti de ces données et de développer une approche de prédiction à base d'un algorithme de Machine Learning du nombre futur de contaminations par jour de COVID-19 au Maroc pour aider les autorités à avoir une vision approximative de la propagation du virus et les aider à prendre des décisions cruciales afin de prendre plus de mesures pour stopper cette pandémie. Nous avons étudié plusieurs approches de prévision, notamment la régression linéaire multiple, le modèle SVN, le Perceptron multicouche (MLP), et le modèle ARIMA. Nous avons appliqué ces approches pour prédire le nombre de contamination prévu pour une journée au Maroc en travaillant sur une dataset. Nous avons comparé les performances de chaque modèle à l'aide de deux statistiques: l'erreur quadratique moyenne racine (RMSE) et l'erreur absolue moyenne(MAE) pour enfin choisir l'algorithme qui s'avèrera le plus bon pour notre dataset.

ABSTRACT

The severity of the coronavirus pandemic has prompted governments in over 200 countries around the world to take drastic preventive measures, to the detriment of their economies. Consequently, the COVID-19 crisis has the potential to generate a lasting and damaging social and economic impact. In Morocco, certain sectors have shown early signs of vulnerability such as tourism, transport and logistics in supply chains, but also - and more difficult to measure - the transversal informal sector. In order to eradicate the COVID-19 pandemic, all the weapons available to humanity must be exploited. Around the world, technologies such as Big Data, AI, Machine learning or Deep learning are already widely used in an attempt to stem the spread of the coronavirus. Now, it is open data or Open Data that could play a major role in this fight.

The aim of this project is to take advantage of this data and to develop a predictive approach based on a Machine Learning algorithm of the future number of COVID-19 contaminations per day in Morocco to help the authorities to have an approximate vision. of the spread of the virus and help them make crucial decisions in order to take more action to stop this pandemic. We studied several forecasting approaches, including multiple linear regression, the SVN model, the multilayer Perceptron (MLP), and the ARIMA model. We applied these approaches to predict the number of contamination expected for a day in Morocco by working on a dataset. We compared the performance of each model using two statistics: the root mean square error (RMSE) and the mean absolute error (MAE) to finally choose the algorithm that will prove the best for our dataset.

LISTE DES TABLES

LISTE DES FIGURES

ABRÉVIATIONS

Table des matières

INTRODUCTION

La crise COVID-19 a un impact considérable sur tous les aspects de notre vie. La priorité immédiate et permanente est inévitablement, et à juste titre, la santé publique, et il est probable qu'elle le restera au cours des semaines et des mois à venir. Dans le monde de la statistique, l'accent est également mis sur l'information en temps utile sur la propagation et l'impact du virus. En premier lieu, il s'agit bien sûr des statistiques sur le nombre de cas et leur issue.

Toutefois, les nombreux autres impacts de COVID-19 suscitent un grand intérêt, notamment les nombreux impacts économiques et sur le marché du travail, qui ont été immédiats et très importants, et qui devraient se poursuivre dans un avenir proche ou potentiellement au-delà. Dans le cas du marché du travail, plusieurs millions de travailleurs dans un grand nombre de pays ont été directement touchés par les lock-out. Certains sont en mesure de poursuivre leur travail grâce au télétravail ou à des accords de travail à distance. Beaucoup d'autres ont vu leurs moyens de subsistance réduits ou complètement perdus. D'autres encore, par exemple les travailleurs de la santé ou de la sécurité publique, connaîtront un autre type de changement, à savoir une augmentation considérable de la charge de travail face à la crise.

Alors que le monde est confronté à la pandémie COVID-19, diverses initiatives voient le jour pour exploiter les talents des analystes, des développeurs en IA et des data engineers. Ces initiatives peuvent donner aux individus et aux équipes l’occasion de travailler sur un projet qui a du sens avec d’autres corps de métier et d’acquérir de nouvelles compétences.

Dans le cadre du projet intégré du semestre 3, nous avons pensé à mettre en place une application Web basée sur la technologie JEE, nommée E-COLOCATION pour nos clients, pour leur permettre de bénéficier des différents services offerts par celle-ci et de les gérer proprement et efficacement.

Le présent rapport synthétise le travail réalisé. Nous ‘y présentons notre projet à travers cinq chapitres :

- Le premier chapitre définit l’étude préliminaire en présentant une étude sur l’existence du projet, le projet et la démarche suivie pour sa conduite;

- Le deuxième chapitre est consacré aux travaux connexes qui ont relation avec notre sujet;

- Le troisième chapitre traite la méthodologie. Ce chapitre présente la méthodologie suivie pour réaliser le projet en citant les principaux algorithmes et outils qui nous ont permis de réaliser notre expérimentation de prédiction ainsi que notre projet.

- Le quatrième chapitre est consacré à la description de notre Dataset et le prétraitement des données;

- Enfin, le cinquième chapitre présente les résultats de notre expérimentation et la mise en œuvre du projet.

Le rapport est clôturé par une conclusion dressant les perspectives de notre travail.

CHAPITRE 1.

ÉTUDE PRÉLIMINAIRE

Dans la première partie de ce chapitre, nous présentons une description générale du projet, qui consiste en la prédiction des cas de contaminations par Covid-19 au Maroc dans les jours à venir. Il explique également le contexte et les objectifs du projet. La dernière section du premier chapitre décrit la planification du projet qui aide à planifier le temps et les objectifs pour chaque étape du projet.

1.1 Covid-19

Les coronavirus (CoV) sont une grande famille de virus qui provoquent des maladies qui vont du simple rhume à des maladies plus graves telles que le syndrome respiratoire du Moyen-Orient (MERS-CoV) et le syndrome respiratoire aigu sévère (SRAS-CoV). Un nouveau coronavirus (nCoV) correspond à une nouvelle souche qui n'a pas été identifiée chez l'homme précédemment.

Les coronavirus sont de type zoonotique, c'est-à-dire qu'ils sont transmis de l'animal à l'homme. Des investigations détaillées ont révélé que le SRAS-CoV et et le MERS-CoV étaient transmis à l'homme par les chats civette et les dromadaires respectivement. Plusieurs coronavirus connus circulent chez des animaux qui n'ont pas encore infecté l'homme.

Les signes courants de l'infection sont les symptômes respiratoires, la fièvre, la toux, l'essoufflement et les difficultés respiratoires. Dans les cas les plus graves, l'infection peut provoquer une pneumonie, un syndrome respiratoire aigu sévère, une insuffisance rénale et même la mort. [W4]

Les recommandations standard pour prévenir la propagation de l'infection comprennent le lavage régulier des mains, le fait de se couvrir la bouche et le nez lorsqu'on tousse et éternue. Éviter tout contact étroit avec toute personne présentant des symptômes de maladie respiratoire tels que la toux et les éternuements.

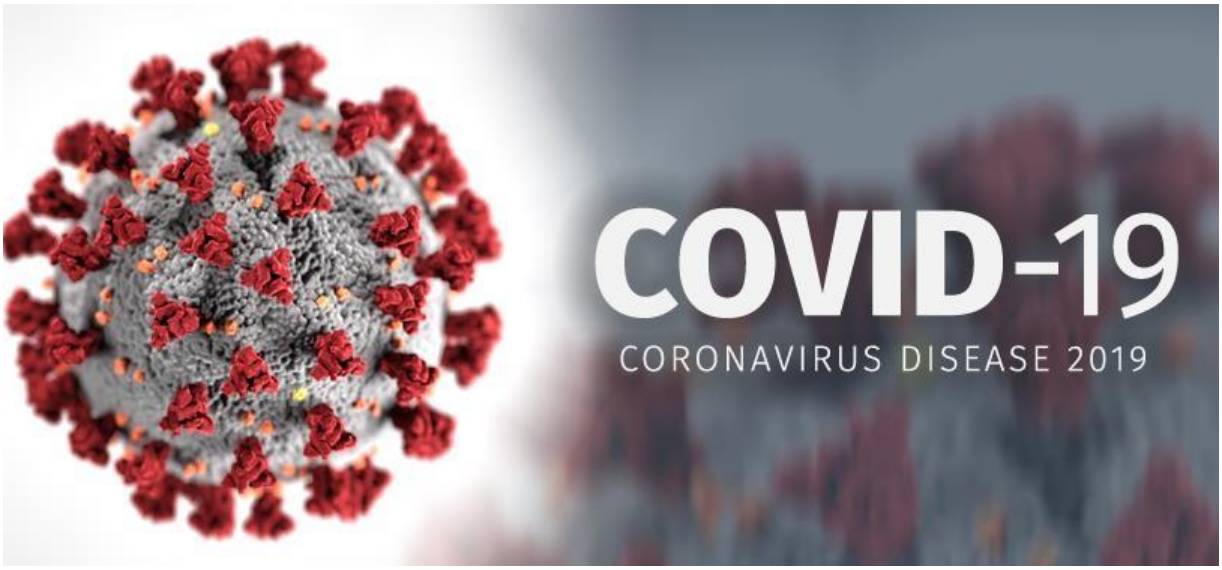

Figure 1: Covid-19.

1.2 Description du projet

1.2.1 Périmètre du projet

Notre projet consiste en la création d'une application Web responsive nommée E-PREDICTIONS à base d'un algorithme Machine learning pour prédire le nombre de contaminations par Covid-19 au Maroc pour les journées à venir. L'objectif de celle-ci est de permettre aux visiteurs(scientifiques, chercheurs, responsables…) d'avoir une idée sur la situation épidémiologique pour agir correctement et prendre de bonnes mesures afin de stopper cette pandémie.

Le site offre bien d'autres services tels un Chatbot pour interagir avec les visiteurs, une Map dynamique mondiale sur la situation de Covid-19, des pages de sensibilisations(vidéos, documents…) et bien d'autres services pour satisfaire nos visiteurs, tout ceci à distance au moyen des nouvelles technologies.

1.2.2 Problématique

Alors que la pandémie du Covid-19 fait des ravages dans le monde, mathématiciens, ingénieurs, datascientists ne cessent de fournir d'efforts pour étudier ce phénomène en analysant toutes les données disponibles pour essayer de trouver des solutions aux effets néfastes que cette pandémie laisse sur le plan économique, sociale et sanitaire et aider à prendre les bonnes décisions.

1.2.3 Objectifs du projet

En prenons compte des problématiques précitées, le projet ayant pour objectif principal la prédiction le nombre de contaminations possible par covid-19 au Maroc en se basant sur un algorithme machine learning.

Cependant notre objectif est de réaliser une application Web responsive ayant pour objectif principal la prédiction du nombre de contaminations possible par covid-19 au Maroc pour une journée donnée, ceci à l'aide d'un algorithme machine learning et des données d'une dataset, pour donner une vision sur l'évolution possible de la pandémie afin de permettre aux experts de prendre les bonnes mesures de préventions et de lutte contre ce fléau.

1.3 Planification du projet

Il est important de planifier notre projet avec souplesse. Une planification bien articulée peut être référencée encore et encore, nous gardant sur la bonne voie et concentrés tout au long du projet.

Le tableau 1 suivant présente toutes les phases de notre projet intégré du Semestre 4:

Table 1: Les phases de réalisation du projet.

Phases	
Phase1 : Définir le problème	
Identifier la problématique	Rencontre avec l'encadrant pour identifier la question de recherche: Comment utiliser le Machine learning dans la prévision des contaminations par Covid-19 ?
Développer la problématique	-Identifier le problème sur lequel nous enquêtons -Identifier l'argument ou la thèse -Les limites de la recherche
Phase2 : Analyse	
Analyse du projet	Lire les livres, documents, informations liés au sujet du projet

Phase3 : Méthodologie	
Choix de l'algorithme Machine learning pour la prédiction	Choisir l'algorithme qui a donné les meilleurs résultats parmi plusieurs pour la prédiction des contaminations par Covid-19 au Maroc.
Phase4: Collecte et analyse des données	
Explorer un ensemble de données antérieur	Vérification de l'exactitude de l'ensemble de données.
L'acquisition des données	Collecte de données à partir de différents emplacements et fusion des fichiers de données.
Analyse exploratoire	Résumer les données à l'aide de méthodes visuelles.
Phase5 : Réaliser l'expérience	
Formation et ajustement du modèle	Formation et adaptation de notre modèle de prévision
Phase6 : Résultats & Déploiement	
Créer un site Web pour afficher les résultats	Compare the performance of statistical and AI methods.

Figure 2: Gantt Chart.

1.4 Conclusion

À travers ce chapitre, le contexte général du projet est identifié. Ensuite, il y a une description du projet et de ses objectifs qui peut être résumée en étudiant les techniques de prévision du niveau de l'eau. Enfin, nous avons décrit l'approche adoptée pour la planification de projet afin de mener à bien nos travaux. Le chapitre suivant présente les travaux connexes réalisés dans le domaine de la prévision par Covid-19.

CHAPITRE 2 Travaux connexes

Dans ce chapitre, nous présenterons quelques travaux autour de la pandémie Covid-19 publiés par des Data Scientists et Data Engineers et qui sont similaires au notre, afin que nous puissions bénéficier de l'expertise existante sur ce sujet pour bien travailler le nôtre. Par la suite nous sortirons par une conclusion qui nous aidera à résoudre notre problématique et bien abordé notre projet.

2.1 Les recherches liées au Covid-19

2.1.1 Institut AMADEUS

Le 08 Avril 2020, l'institut AMADEUS a publié une recherche sur l'analyse statistique et prévision de la propagation de Covid-19 au Maroc par région. Sur la base des paramètres de base qui ont un impact direct sur l'évolution de la pandémie, ils ont tenté de simuler une prédiction du résultat final.

Le but de cette étude n'est pas de donner un bilan exact car il n y'a pas un contrôle total sur les paramètres d'entrée. Mais ils ont pu définir des scénarios de simulation.

Par conséquent, nous pouvons dire que pour diminuer le nombre d'infections par jour, nous devons gérer l'un des trois paramètres ou tous :

- Le nombre de contacts qu'un infecté peut avoir par jour et ne serait évidemment possible que si nous respections tous l'isolement et restions autant que possible à la maison.

- Le nombre d'individus en bonne santé qui étaient déjà infectés et guéris, cette deuxième option, telle que choisie par les États-Unis au départ et la Suède jusqu'à aujourd'hui, cette stratégie est principalement basée sur le sacrifice de personnes âgées et faibles pour le bien de l'économie.

- Nombre de tests par jour pour le coronavirus au cas où il serait partiel. De nombreux pays utilisent désormais différents laboratoires pour développer des tests moins chers et plus rapides que les tests conventionnels.

Nous avons commencé par calculer ce nombre sur la base de l'équation qui permet de calculer le nombre de nouveaux cas enregistrés un certain jour sous forme de multiplication du nombre d'infections enregistrées la veille, le nombre de contacts par jour (que nous devons trouver) et la probabilité que l'autre contact ou individu ne soit pas infecté et ne l'ait jamais été. (Égal à la moyenne de la population non infectée à l'état initial). De là, nous avons rassemblé des valeurs sur ce nombre, selon chaque jour à partir du 2 mars ; le jour où le premier cas officiel a été enregistré dans le pays. Et en utilisant l'équation de ce nombre par rapport à la journée (en comptant du 2 mars à 1 ... et du 4 avril à 33), nous avons réalisé que selon la situation actuelle de la quarantaine, le nombre de cas continuera d'augmenter pour atteindre un pic de 326 cas par jour le 19 avril et le dernier cas confirmé sera le 3 mai avec un total de 7275 cas confirmés.

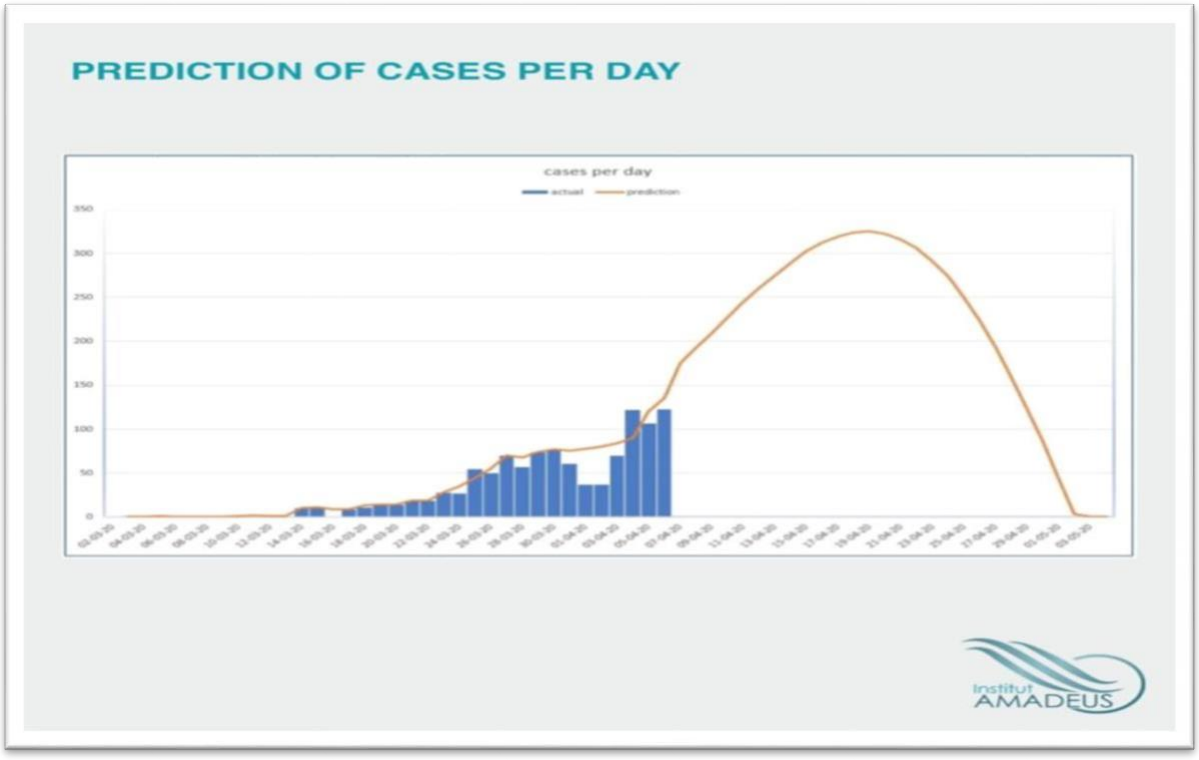

Figure 3: Prédiction des cas par jour pour l'institut AMADEUS.

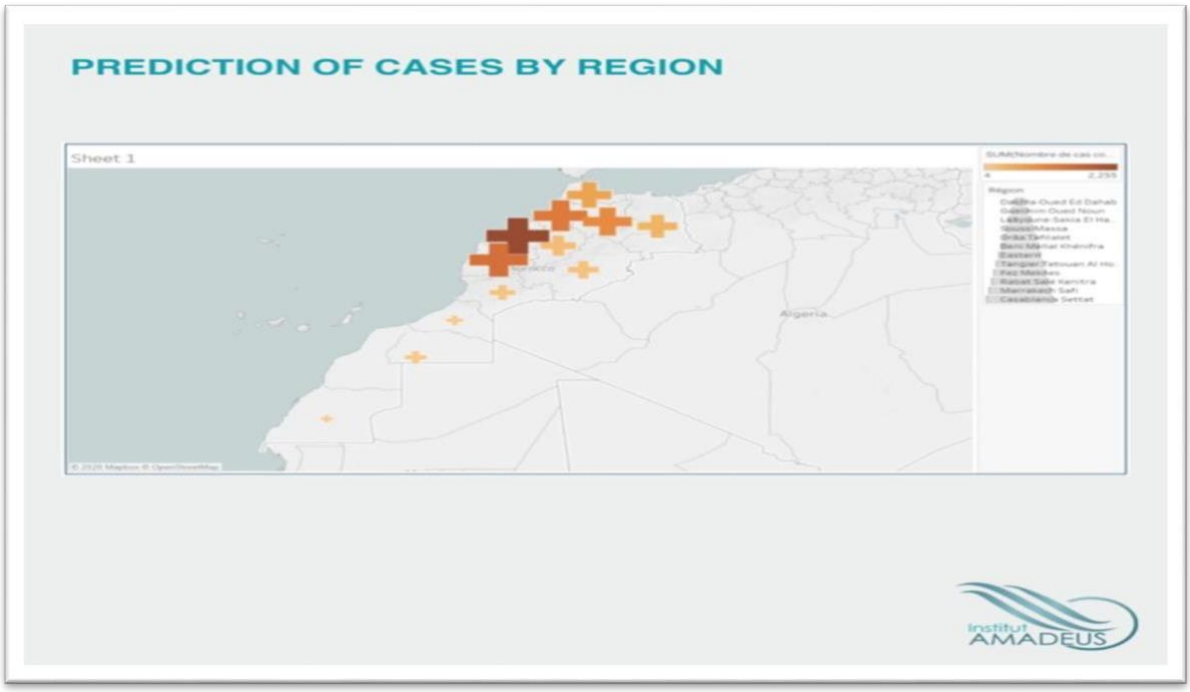

Figure 4: Prédiction des cas par région pour l'institut AMADEUS.

2.1.2 Travaux publiés sur le Web

Kaggle est, une plateforme web très puissant, la plus grande communauté de science des données au monde avec des outils et des ressources puissants pour vous aider à atteindre vos objectifs de science des données. »

Un travail publié sur kaggle intitulé « Prediction of death and confirmed cases (COVID-19) » pour prédire le nombre de cas confirmés et morts en utilisant les différents modèles en l'occurrence la machine à vecteurs de support (SVM), la régression linéaire et la régression de crête. Pour la prédiction des cas confirmés, on observe les graphes représentés ci-dessous pour chaque algorithme :

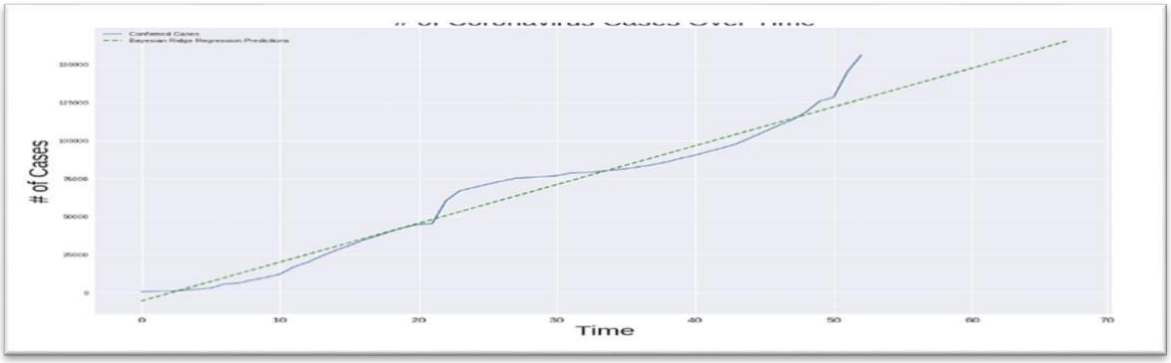

Figure 5: Prédiction des cas confirmés par Ridge Régression.

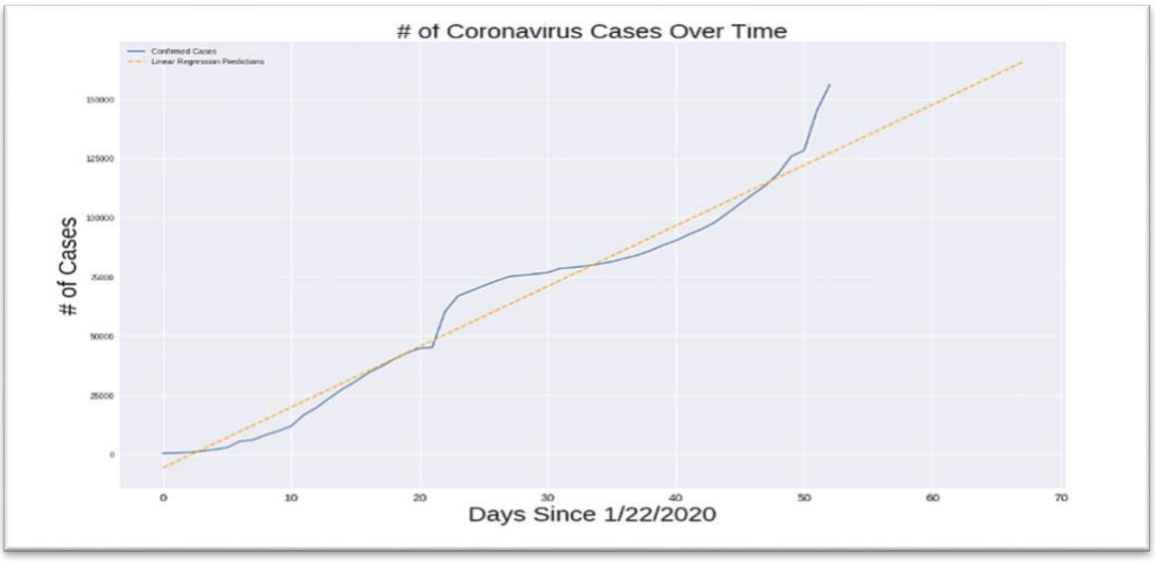

Figure 6: Prédiction des cas confirmés par la régression linéaire.

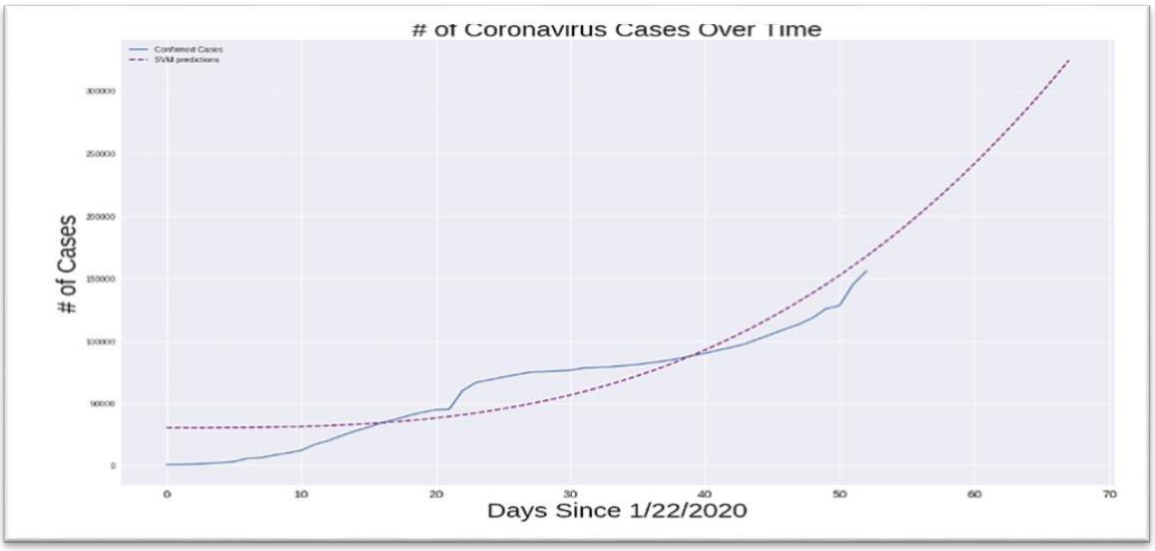

Figure 7: Prédiction des cas confirmés par SVM.

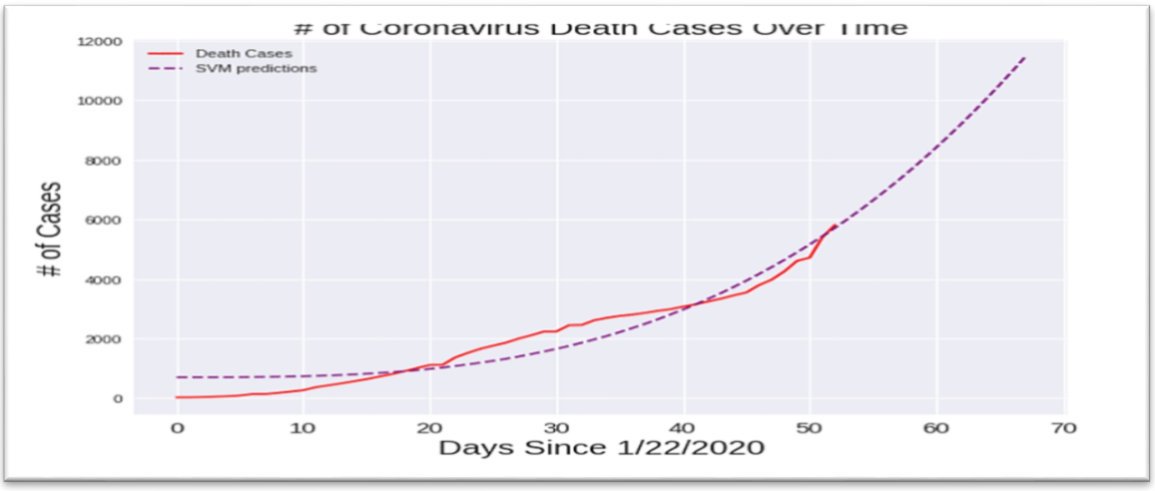

Figure 8: Prédiction des cas morts par SVM.

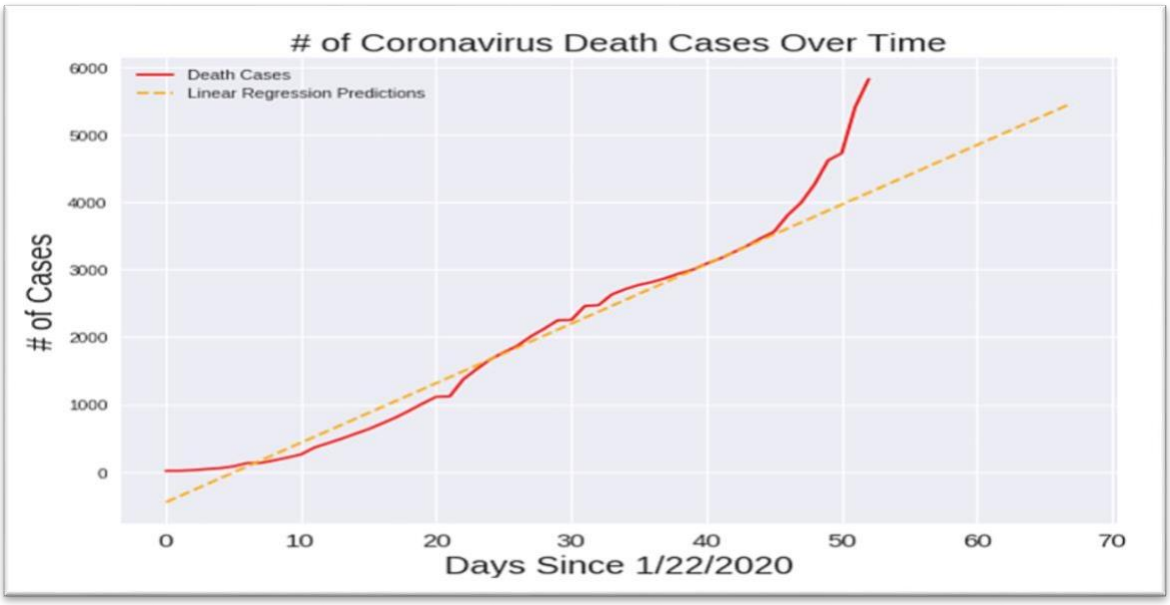

Figure 9: Prédiction des cas morts par LR.

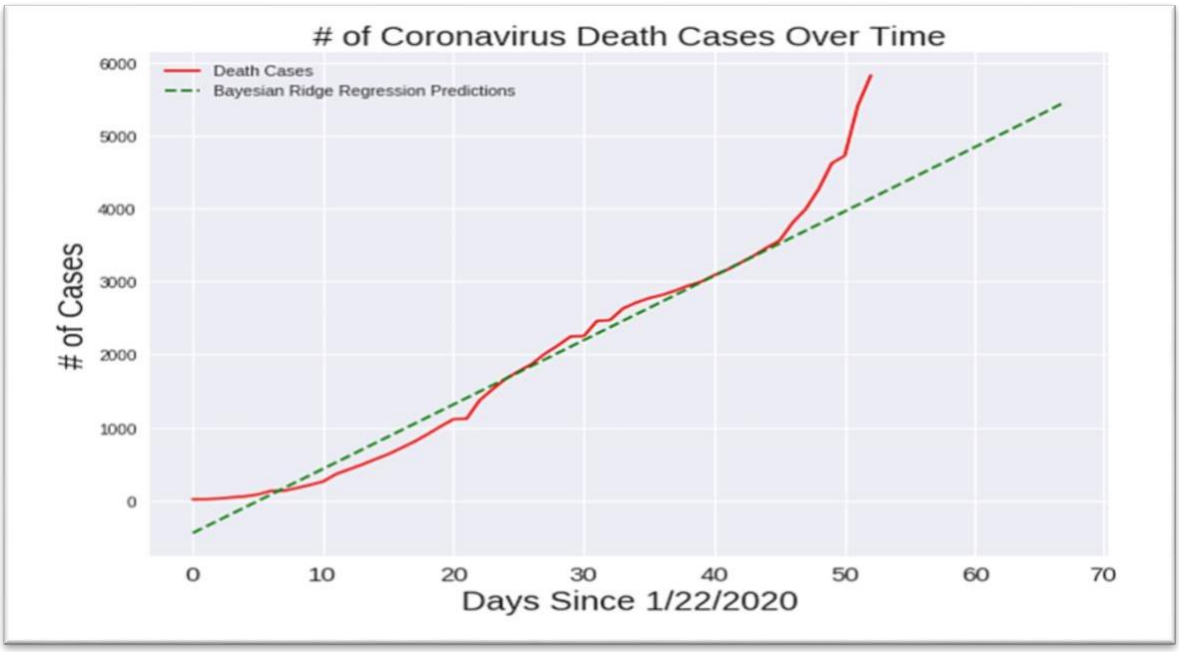

Figure 10: Prédiction des cas morts par Ridge Régression.

2.4 Conclusion

Pour conclure , la prédiction de l'institut AMADEUS tout en comptant sur la variation des paramètres précités était loin un petit peu de la réalité. En revanche, Le travail publié sur Kaggle était très pertinent à travers l'implémentation des différents algorithmes et qui a donné de résultats proches plus ou moins de la réalité. Avec la diversification des algorithmes mis en place pour prédire les cas confirmés, on peut s'interroger sur leur performance. Pour cela, dans le chapitre qui suit, nous allons faire un choix entre les différents algorithmes candidats qui existent, puis nous allons citer les différentes technologies utilisées.

CHAPITRE 3.

MÉTHODOLOGIE

Ce chapitre est consacré à la description des algorithme candidats pour la prédiction. Tout d'abord, nous présenterons les différentes les algorithmes. Puis nous choisirons le meilleur d'entre eux. Et à la fin, nous donnerons un aperçu des outils utilisés pour former et tester notre prédiction ainsi que les outils qu'on a utilisé pour réaliser le site Web.

3.1 Algorithmes candidats

3.1.1 Régression linéaire

En statistiques, en économétrie et en apprentissage automatique, un modèle de régression linéaire est un modèle de régression qui cherche à établir une relation linéaire entre une variable, dite expliquée, et une ou plusieurs variables, dites explicatives. On parle aussi de modèle linéaire ou de modèle de régression linéaire.
Parmi les modèles de régression linéaire, le plus simple est l'ajustement affine. Celui-ci consiste à rechercher la droite permettant d'expliquer le comportement d'une variable statistique y comme étant une fonction affine d'une autre variable statistique X. En général, le modèle de régression linéaire désigne un modèle dans lequel l'espérance conditionnelle de y sachant x est une transformation affine Modèle: Lourd paramètres. Cependant, on peut aussi considérer des modèles dans lesquels c'est la médiane conditionnelle de y sachant x ou n'importe quel quantile de la distribution de y sachant x qui est une transformation affine en les paramètres.
Comme les autres modèles de régression, le modèle de régression linéaire est aussi bien utilisé pour chercher à prédire un phénomène que pour chercher à l'expliquer.

Après avoir estimé un modèle de régression linéaire, on peut prédire quel serait le niveau de y pour des valeurs particulières de x.

Il permet également d'estimer l'effet d'une ou plusieurs variables sur une autre en contrôlant par un ensemble de facteurs. Par exemple, dans le domaine des sciences de l'éducation, on peut évaluer l'effet de la taille des classes sur les performances scolaires des enfants en contrôlant par la catégorie socio-professionnelle des parents ou par l'emplacement géographique de l'établissement. Sous certaines hypothèses restrictives, cet effet peut être considéré comme un effet causal.

En apprentissage statistique, la méthode de régression linéaire est considérée comme une méthode d'apprentissage supervisé utilisée pour prédire une variable quantitatIVE10.

Dans cette perspective, on entraîne généralement le modèle sur un échantillon

d'apprentissage et on teste ensuite les performances prédictives du modèle sur un échantillon de test.
On considère le modèle pour l'individu i. Pour chaque individu, la variable expliquée s'écrit comme une fonction linéaire des variables explicatives.

$$y_i = \beta_0 + \beta_1 x_{i,1} + \ldots + \beta_K x_{i,K} + \varepsilon_i$$

où yi et les xi,j sont fixes et εi représente l'erreur.

3.1.2 ARIMA

Dans les statistiques et l'économétrie, et en particulier dans l'analyse des séries temporelles, un modèle autorégressif à moyenne mobile intégrée (ARIMA) est une généralisation d'un modèle autorégressif à moyenne mobile (ARMA). Ces deux modèles sont ajustés aux données de séries chronologiques soit pour mieux comprendre les données, soit pour prédire les points futurs de la série (prévisions). Les modèles ARIMA sont appliqués dans certains cas où les données montrent des preuves de non-stationnarité, où une étape de différenciation initiale (correspondant à la partie "intégrée" du modèle) peut être appliquée une ou plusieurs fois pour éliminer la non-stationnarité.
un modèle ARIMA est une classe de modèles statistiques pour l'analyse et la prévision de données de séries chronologiques.

Il s'adresse explicitement à une suite de structures standard dans les données de séries temporelles, et en tant que tel, fournit une méthode simple mais puissante pour faire des prévisions de séries temporelles habiles.

ARIMA est un acronyme qui signifie AutoRegressive Integrated Moving Average. C'est une généralisation de la moyenne mobile auto-régressive plus simple et ajoute la notion d'intégration.

Cet acronyme est descriptif, capturant les aspects clés du modèle lui-même. En bref, ce sont:

AR: autorégression. Un modèle qui utilise la relation dépendante entre une

observation et un certain nombre d'observations décalées.
J'ai intégré. L'utilisation de la différenciation des observations brutes (par exemple, soustraction d'une observation d'une observation au pas de temps précédent) afin de rendre les séries chronologiques stationnaires.
MA: Moyenne mobile. Un modèle qui utilise la dépendance entre une observation et une erreur résiduelle d'un modèle de moyenne mobile appliqué aux observations décalées.
Chacun de ces composants est explicitement spécifié dans le modèle en tant que paramètre. Une notation standard est utilisée pour ARIMA (p, d, q) où les paramètres sont remplacés par des valeurs entières pour indiquer rapidement le modèle ARIMA spécifique utilisé.

Les paramètres du modèle ARIMA sont définis comme suit:

p: Le nombre d'observations de décalage inclus dans le modèle, également appelé ordre de décalage.
d: Le nombre de fois que les observations brutes sont différenciées, également appelé degré de différenciation.
q: La taille de la fenêtre de moyenne mobile, également appelée ordre de moyenne mobile.
Un modèle de régression linéaire est construit comprenant le nombre et le type de termes spécifiés, et les données sont préparées par un degré de différenciation afin de le rendre stationnaire, c'est-à-dire de supprimer les structures de tendance et saisonnières qui affectent négativement le modèle de régression.

Une valeur de 0 peut être utilisée pour un paramètre, ce qui indique de ne pas utiliser cet élément du modèle. De cette façon, le modèle ARIMA peut être configuré pour remplir la fonction d'un modèle ARMA, et même d'un simple modèle AR, I ou MA.

L'adoption d'un modèle ARIMA pour une série chronologique suppose que le processus sous-jacent qui a généré les observations est un processus ARIMA. Cela peut sembler évident, mais contribue à motiver la nécessité de confirmer les hypothèses du modèle dans les observations brutes et dans les erreurs résiduelles des prévisions du modèle.

3.1.3 SVM

Les machines à vecteurs de support ou séparateurs à vaste marge (en anglais support Vector machine, SVM) sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des « classifiers » linéaires.

Les séparateurs à vaste marge ont été développés dans les années 1990 à partir des considérations théoriques de Vladimir Vapnik sur le développement d'une théorie statistique de l'apprentissage : la théorie de Vapnik-Tchervonenkis. Ils ont rapidement été adoptés pour leur capacité à travailler avec des données de grandes dimensions, le faible nombre d'hyperparamètres, leurs garanties théoriques, et leurs bons résultats en pratique.

Les SVM ont été appliqués à de très nombreux domaines (bio-informatique, recherche d'information, vision par ordinateur, finance1…). Selon les données, la performance des machines à vecteurs de support est de même ordre, ou même supérieure, à celle d'un réseau de neurones ou d'un modèle de mélanges gaussiens
Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. La résolution de ces deux problèmes passe par la construction d'une fonction h qui a un vecteur d'entrée x fait correspondre une sortie y

3.2 Les outils utilisés

3.2.1 Langage Python

Python est un langage de programmation de haut niveau interprété pour la programmation à usage général. Créé par Guido van Rossum et publié pour la première fois en 1991, Python a une philosophie de conception qui met l'accent sur la lisibilité du code, notamment en utilisant des espaces importants. Il fournit des constructions qui permettent une programmation claire à la fois à petite et à grande échelle[W30].

Figure 11: Logo de Python.

Python dispose d'un système de type dynamique et d'une gestion automatique de la mémoire. Il prend en charge plusieurs paradigmes de programmation, notamment orientés objet, impératifs, fonctionnels et procéduraux, et dispose d'une bibliothèque standard large et complète. Beaucoup de code écrit en Python s'est construit au fil des décennies et, étant un langage de programmation open source, une grande partie de celui-ci a été publié pour que d'autres puissent l'utiliser. La quasi-totalité est collectée sur https://pypi.python.org, prononcée «pie-pee-eye» ou, plus communément appelée «CheeseShop». Vous pouvez installer ce logiciel sur votre système pour être utilisé par vos propres projets. Par exemple, si vous souhaitez utiliser Python pour créer des scripts avec des arguments de ligne de commande, vous devez installer la bibliothèque "click", puis l'importer dans vos scripts et l'utiliser. Il existe des bibliothèques pour pratiquement tous les cas d'utilisation que vous pouvez créer, de la manipulation d'images aux calculs scientifiques en passant par l'automatisation des serveurs.

De plus, les interprètes Python sont disponibles pour de nombreux systèmes d'exploitation. CPython, l'implémentation de référence de Python, est un logiciel open source [29] et possède un modèle de développement basé sur la communauté, comme le font presque toutes ses

variantes d'implémentation. CPython est géré par la fondation à but non lucratif Python Software Foundation [W31].
Python est plus efficace que le langage R dans les approches d'apprentissage en profondeur. Il peut également gérer efficacement les mégadonnées. Nous avons utilisé la bibliothèque python Keras pour implémenter notre LSTM.

3.2.2 Jupyter Notebook

Jupyter Notebook est un outil incroyablement puissant pour développer et présenter de manière interactive des projets de science des données. Un bloc-notes intègre le code et sa sortie dans un document unique qui combine des visualisations, du texte narratif, des équations mathématiques et d'autres médias riches. Le flux de travail intuitif favorise un développement itératif et rapide, faisant des blocs-notes un choix de plus en plus populaire au cœur de la science des données contemporaine, de l'analyse et de plus en plus de la science en général. Mieux encore, dans le cadre du projet open source Jupyter, ils sont entièrement gratuits.

Figure 12: Logo de Jupyter.

Le projet Jupyter est le successeur du précédent IPython Notebook, qui a été publié pour la première fois en tant que prototype en 2010. Bien qu'il soit possible d'utiliser de nombreux langages de programmation différents dans les ordinateurs portables Jupyter [W34].

3.2.3 Flask

Flask est un framework open-source de développement web en Python. Son but principal est d'être léger, afin de garder la souplesse de la programmation Python, associé à un système de templates

Figure 13: Logo Flask.

3.4 Conclusion

Dans ce chapitre, nous avons vu un aperçu des algorithmes candidats qu'on a choisi pour faire notre prédiction. Dans les chapitres qui suivent nous essayerons de trouver une bonne Dataset pour notre expérimentation ensuite nous retiendrons un algorithme parmi ceux vu pour l'adopter à notre sujet et obtenir les résultats.

CHAPITRE 4.

DESCRIPTION ET PRÉTRAITEMENT DES DONNÉES

Ce chapitre est consacré à la description de notre Dataset choisi.

4.1. xxxxx

In

4.1.1. xxxxxx

A xxxxxx

4.2. xxxxxx

xxxxxx

4.5 Conclusion

Pour conclure...

CHAPITRE 5. RÉSULTATS & DÉPLOIEMENT

Xxxxxxxxxxxxxxxxxxxx

5.1. Xxxxxxxxxxxxx

xxxxxxxx

5.2. xxxxxx

xxxxxxxx

5.3. Conclusion

xxxxxxxxxxxxxxxx

CONCLUSION

xxxxxxxxxxx

BIBLIOGRAPHY

[1] W

WEBOGRAPHY

[W1] U.S. News. 2012, March 14. “10 States Most at Risk of Flooding.”, Available on: *https://www.usnews.com/news/slideshows/10-states-most-at-risk-of-flooding*

[W2]Texas Development Water Board. 2016. “$3.5 million in flood protection grants approved by the TWDB, [Online], Available on: *https://www.twdb.texas.gov/newsmedia/press_releases/2016/08/flood.asp*

[W3] Bilder, Mike and Ed Johnson. 2012. “State of the Enterprise - National Weather Service.”, [Online], Available on: *https://www.ametsoc.org/cwwce/index.cfm/reportsand-studies/general-reports-and-studies/state-of-the-national-weather-service-in2012/*

https://www.lemagit.fr/actualites/252480603/Coronavirus-les-data-scientists-se-mobilisent-pour-aider-les-chercheurs

http://www.emro.who.int/fr/health-topics/corona-virus/about-covid-19.html