Introduction :

Le coronavirus (COVID-19) a récemment provoqué une préoccupation mondiale majeure.  
Alors que le nombre de cas de coronavirus augmenterait, la propagation de COVID-19 est une menace sérieuse pour la santé mondiale.

Donc, bel et bien il s’agit d’un sujet d’actualité dans toute la force du terme.

Analyse de ce qui est demande :

Enonce :

**Proposer un modèle de prédiction, permettant de prédire l’évolution des décès et le nombre des contaminations pour chaque pays.**

Remarques :

* C'est bien entendu du machine Learning, c'est un cas de time séries- Les données qu’on a sont dans l’API. Json, sont des séries temporelles qui veut dire que leur valeur change au cours du temps-.
* Notez bien que, L’IA peut être utilisée pour de nombreux problèmes de prédiction et peut également être utilisée ici pour les prédictions liées au virus Corona. Cependant, comme les données ne sont pas fiables et cohérentes et qu'elles changent à mesure que les gouvernements modifient les politiques, il ne sera pas possible d'élaborer un modèle de prévision précis.

Réflexions et motivation :

Proposition 1 :

* Les données qu’on a sont dans l’API. Json, sont des séries temporelles qui veut dire que leur valeur change au cours du temps.
* Dans notre cas je dois prédire deux valeurs : **le nombre de décès** et **le nombre de contaminations** à partir de l'historique. Donc il s’agit d'une régression de plusieurs valeurs
* En machine Learning, il y a en générale deux types de problèmes :
* **Classification : classifier les données en des catégories**
* **Régression : prédire une valeur réelle**
* Donc ici il s’agit d’un problème de régression
* J’ai quatre variables pour chaque pays
* J’essaie d’abord par penser à un modèle pour chaque pays
* Après le généraliser.
* C’est beaucoup de travail mais ça peut être fait !!

Proposition 2 :

* Pourquoi TensorFlow**??**
* Méthodes prédictives ... processus stochastique existant ... Deep Learning est mal adapté aux prévisions à court terme.
* Donc…sauf, si on veut s’amuser avec les réseaux de neurones.
* Sachant que, les modèle non apprentissages peuvent avoir une solution, mais il faut savoir que le nombre des données est souvent un handicape pour les méthodes traditionnelles.

Proposition 3 :

* Faire des simulations graphiques sur les données que plusieurs sites proposent (le nombre de cas confirmés porteurs du virus) puis étudier un peu l'évolution du virus au sein d'un pays (phase de déclenchement, d'activation, d'inflexion vers un déclin probable des contaminations (cas de la Chine) ou une hausse incontrôlée (cas de l'Italie)). Cela vous donnera une idée générale sur la propagation du virus (en gros, quelle allure : linéaire, exponentielle ...) et donc sur le modèle statistique sur lequel vous pourrez vous basez par la suite (régression, SVM, réseaux de neurones, ...).
* En effet, cette dernière proposition je la trouve personnellement judicieuse, il est toujours intéressant de voir les données de chaque pays sur un graphe, surtout que tous les pays ne sont pas au même stade. Deuxièmement, distinguer les phases du virus est une bonne idée.

Choix d’un modèle :

* Vu le temps consacre a cette partie, je dois donc plutôt me baser sur un modèle existant et essayer de voir sa performance sur les données a disposition.
* Dans ce travail, nous allons essayer de prédire la propagation du coronavirus pour chacune des régions infectées. Ajustement des analyses de séries chronologiques et des algorithmes statistiques pour produire les meilleures prévisions à court terme et à long terme. Un filtre adaptatif de **Kalman en ligne** nous fournit une très bonne prévision sur une journée pour chaque région.

Présentation du modèle :

* Le filtre de Kalman a été lancé par Rudolf Emil Kalman en 1960, initialement conçu et développé pour résoudre le problème de navigation dans Apollo Project.
* Depuis lors, il a de nombreuses applications dans les technologies telles que le guidage, la navigation et le contrôle de véhicules, le suivi d'objets par vision par ordinateur, l'optimisation de trajectoire, l'analyse de séries chronologiques dans le traitement du signal, l'économétrie et plus encore.
* Le filtre de Kalman est un algorithme récursif qui utilise une mesure chronologique dans le temps, contenant du bruit statistique et produisant des estimations de variables inconnues.

Prédictions

* Comme expliqué précédemment, pour la prédiction sur une journée, nous utilisons le filtre de Kalman, tandis que pour la prévision à long terme, nous ajustons un modèle linéaire où ses principales caractéristiques sont les prédicteurs de Kalman, le taux d'infection par rapport à la population, les caractéristiques dépendant du temps et l'historique météorologique et prévision.
* La prédiction d'un jour de Kalman est très précise et puissante tandis qu'une prédiction sur une période plus longue est plus difficile mais fournit une tendance future.
* La prédiction à long terme ne garantit pas une exactitude complète mais fournit une estimation juste suivant la tendance récente. Le modèle devrait être réexécuté quotidiennement pour obtenir de meilleurs résultats.

**Reference :**

* <https://github.com/Rank23/COVID19/blob/master/w.csv>
* <https://towardsdatascience.com/using-kalman-filter-to-predict-corona-virus-spread-72d91b74cc8>
* <https://towardsdatascience.com/using-kalman-filter-to-predict-corona-virus-spread-72d91b74cc8>
* <https://towardsdatascience.com/machine-learning-methods-to-aid-in-coronavirus-response-70df8bfc7861>