**Intro :**

La qualité de l’air est un sujet qui passionne et qui devient une métrique importante dans notre société dans un contexte de changement climatique. En effet ceci a des conséquences importantes à court terme comme la gratuité des transports, la circulation alternée ou l’augmentation des pathologies respiratoires lors des pics de pollution.

Il y a de plus en plus de mesure de qualité de l’air en France et en Europe, mais nous pourrions nous servir du Machine Learning pour pouvoir prédire un indice de qualité dans des zones ou la mesure n’est pas faite, et c’est l’objectif de notre projet : prédire un indice plutôt fiable, avec peu de données.

**Nos objectifs :**

Le but de ce projet est donc de réussir à prédire la pollution de l’air du lendemain à partir des données de la ruche située sur le toit de l’ISEN.

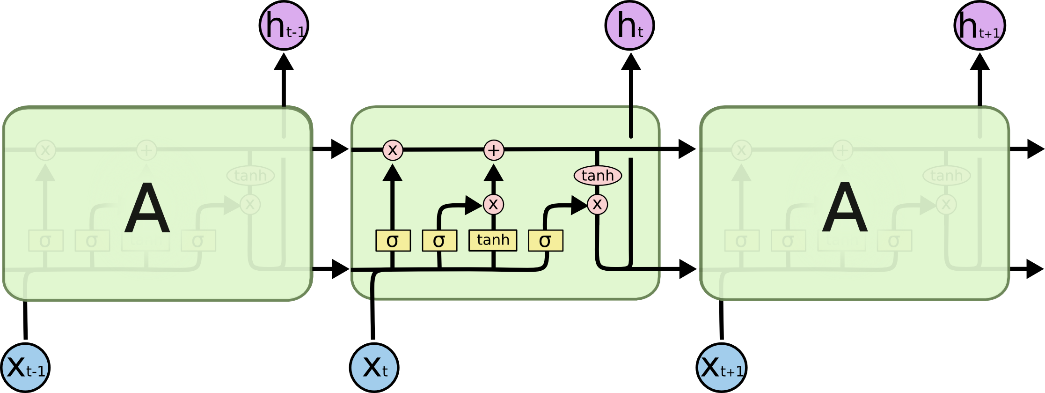
Cette dernière est équipée de capteurs qui mesurent la température et l’humidité de l’air (entre autres). Nous utilisons également l’indice de qualité de l’air fourni par ATMO pour nos prédictions.

Avec ce type de données, il est évident que nous ne pouvons pas prédire l’indice de qualité de l’air avec 100% de précision, mais nous devons donner une approximation suffisamment fiable pour pouvoir être utilisée concrètement.

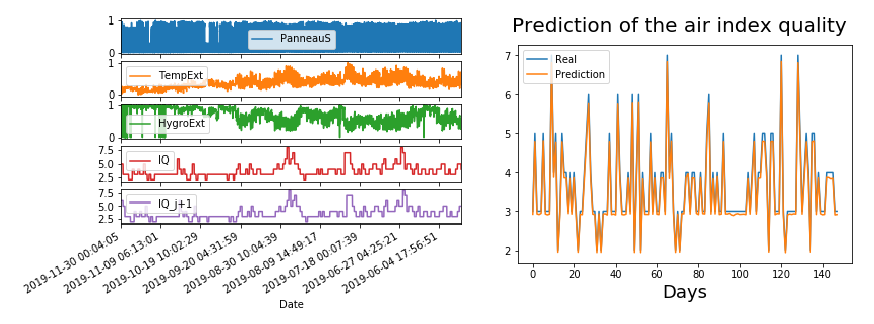
**Notre travail :**

1. **Créer un premier modèle pour la ruche**

Notre premier objectif était de créer le programme permettant de prédire la qualité de l'indice de l'air. Nous avons donc fait nos recherches et nous avons rapidement découvert la nécessité d’utiliser un Réseau de Neurones Récurrents (RNN), car ce type de modèle nous permet de conserver la temporalité des données et trouver des liens entre différents instants. Plus précisément nous utilisons un réseau nommé LSTM (pour Long Short Term Memory).



Nous avons obtenu le résultat suivant :

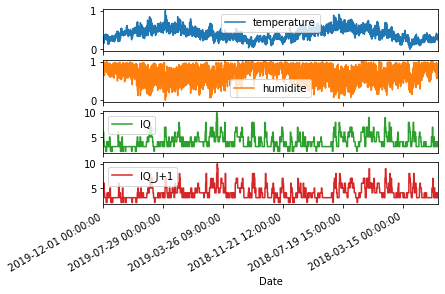


Cela n’avait pas l’air si mauvais, mais la précision maximale que nous avions sur les données de validation était d’environ 35%, ce qui n’est pas très satisfaisant.

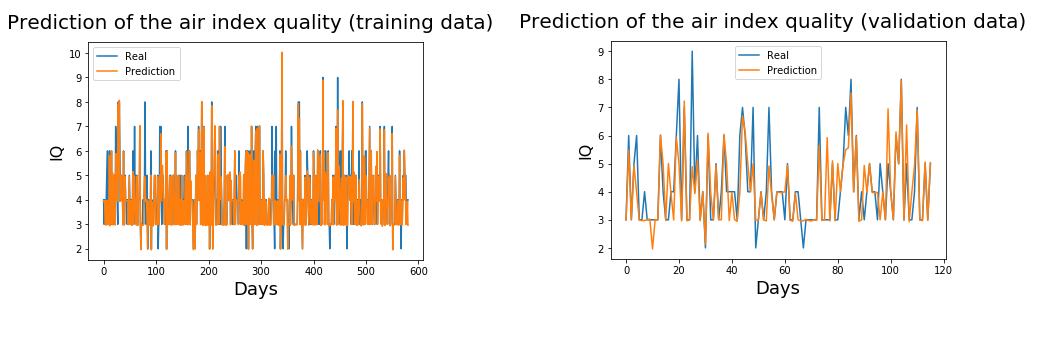
Nous en avons tiré plusieurs conclusions, mais la plus importante est que nous n’avons clairement pas assez de données avec 130 jours de data.

1. **Elargir notre dataset**

Après ces résultats mitigés mais encourageants, nous avons décidé de former notre modèle aux données publiques, disponibles gratuitement en ligne. Nous choisissons l'ensemble de données météorologiques sur le site Web de Météo France :



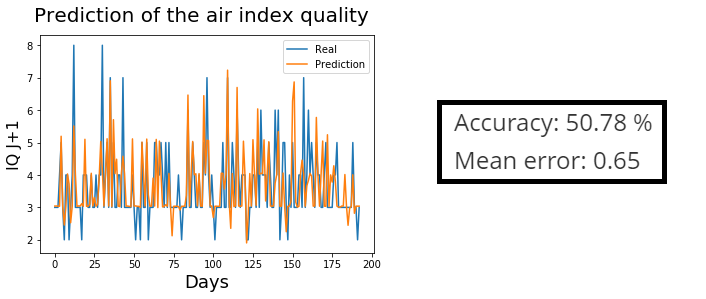
L’entrainement sur ces données a donné des résultats beaucoup plus satisfaisants :



Avec 69% d’accuracy et 0,49 d’erreur moyenne sur les données de validation.

1. **Appliquer notre modèle à la ruche**

Nous testons maintenant le modèle que nous avons formé sur les données de Météo France pour prédire la qualité de l’air avec les données de la ruche. Voici le premier résultat que nous avons eu :



**Conclusion et généralisation :**