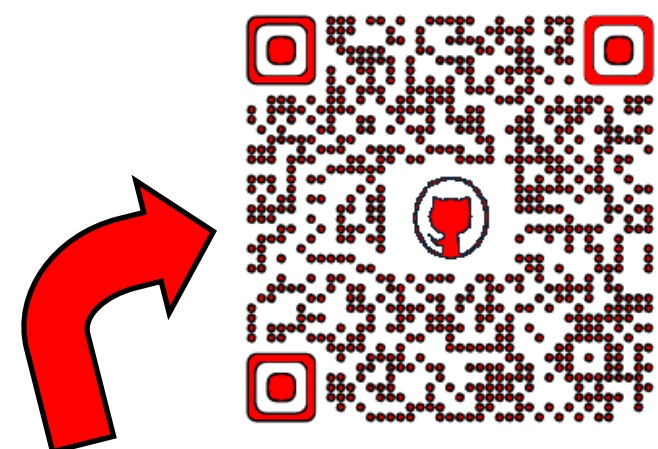


Projet AirIQ

Réalisé par Mohamed BOULANOUAR, Maxime THOOR et Alexandre VEREPT

Prédiction de l'indice de qualité de l'air à Lille



Code source sur
notre GitHub

Pourquoi ce projet ?

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.

Aujourd'hui, de plus en plus de mesures de pollution sont réalisées en France et en Europe. Grâce au **Machine Learning**, nous pouvons désormais prédire un indice de qualité de l'air dans les métropoles.

Ainsi, l'**objectif** de notre projet est de **déterminer un indice fiable avec 2 facteurs mesurables facilement à l'échelle locale: la température et l'humidité**. Ces mesures sont apportées par la ruche située sur le toit de l'ISEN à Lille.

ISEN
ALL IS DIGITAL!
LILLE



1 Choix de l'algorithme de prédiction

Tout d'abord, une petite analogie pour comprendre le Machine Learning :

Imaginons que l'on veuille **apprendre les couleurs à un enfant** qui sait déjà parler et qui connaît les noms des couleurs, mais qu'il ne sache pas quel nom associer avec quelle couleur. Nous lui faisons passer un test:

Quelles sont ces couleurs ?

Couleur 1
Couleur 2
Couleur 3
Couleur 4



Vert foncé X
Gris X
Magenta X
Orange X

Si vous n'avez pas de déficience visuelle et que vous connaissez vos couleurs, vous vous rendez compte rapidement que l'enfant a répondu **au hasard**, et c'est normal, il n'a encore jamais associé la couleur à son nom !

Lors de l'**entraînement**, nous lui montrons de nombreuses couleurs associées à leur nom.

Ces couleurs sont **Orange, Bleu clair, Vert clair, Rouge, Bleu foncé, Jaune, ...**

Si nous lui montrons suffisamment d'exemples, il sera capable de comprendre ce que les couleurs représentent, alors même que **nous n'avons pas défini clairement ce que pouvait être le jaune ou le rouge !**

Lorsque nous lui faisons un second test, il est capable de répondre, et ce même sur certaines couleurs qu'il n'a jamais vues, mais qui restent proches de celles qu'il a déjà rencontrées :

Quelles sont ces couleurs ?

Couleur 1
Couleur 2
Couleur 3
Couleur 4

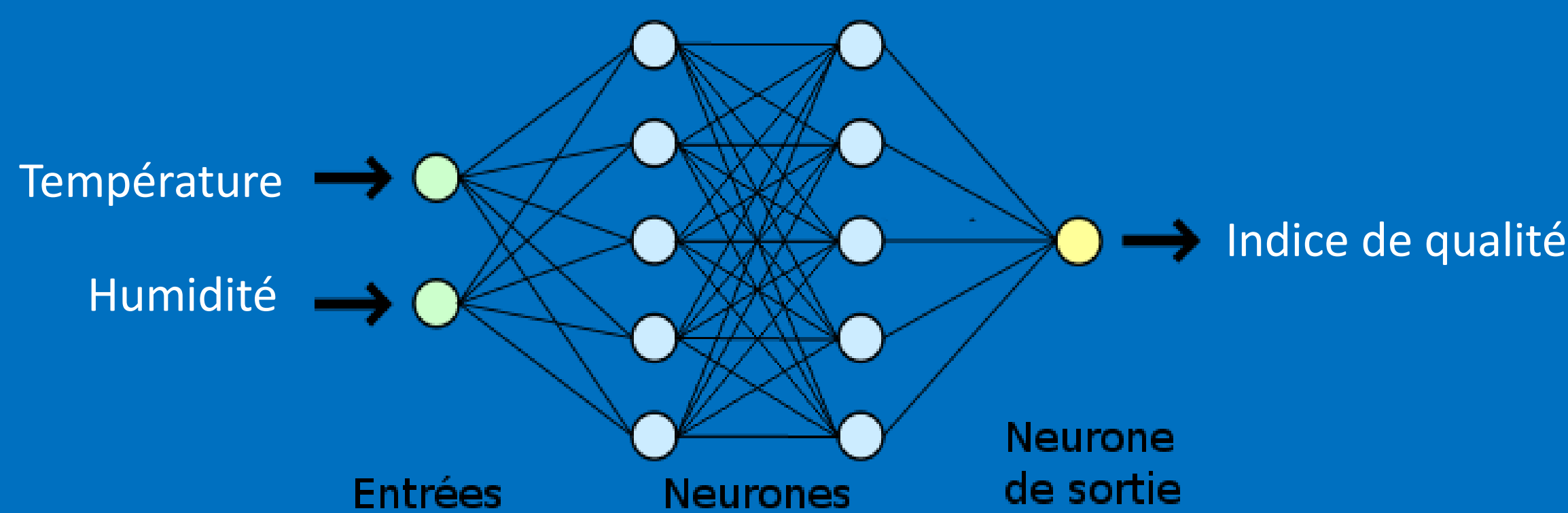


Orange O
Bleu clair O
Vert foncé O
Rouge foncé O

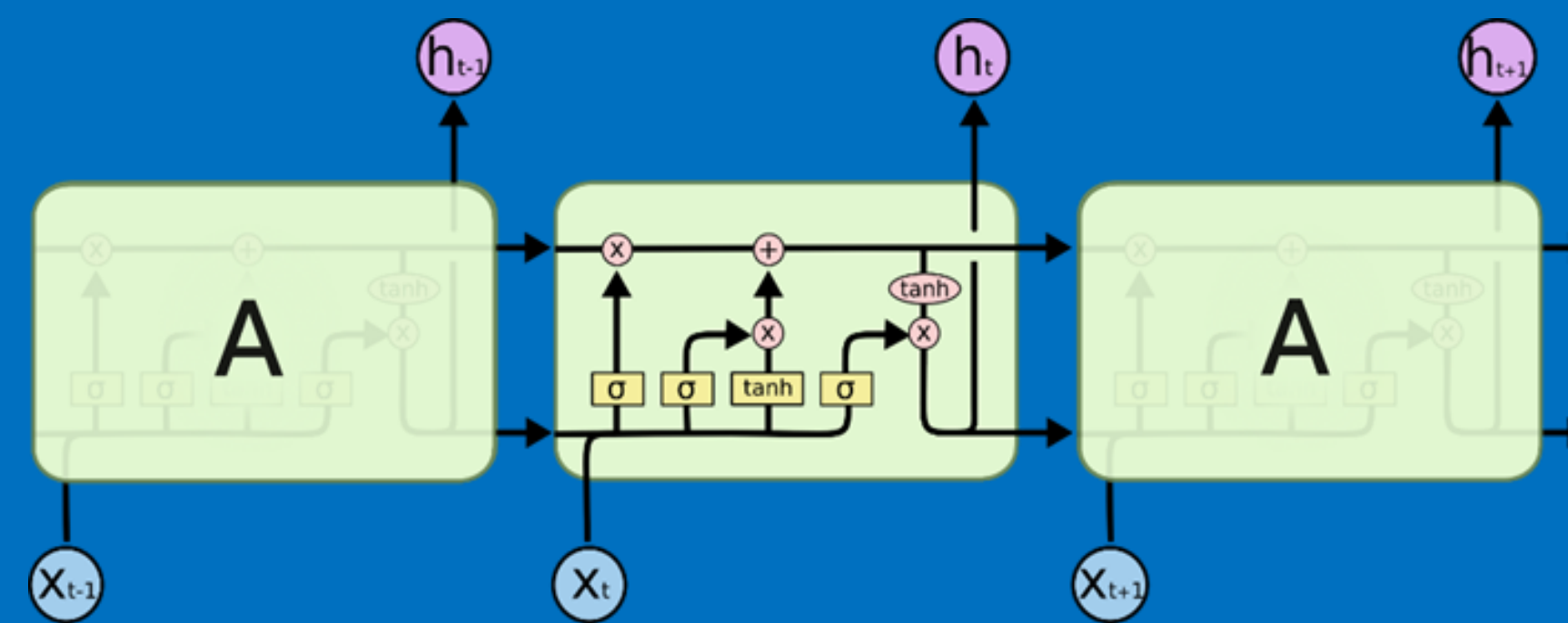
Le petit bonhomme a appris à reconnaître les couleurs en regardant les exemples. Eh bien, le **Machine Learning** est **basé exactement sur le même principe**, mais notre élève est remplacé par un **réseau de neurones** artificiel, et il est un peu moins intelligent que notre humain. Il nécessite donc énormément d'exemples pour s'améliorer.

Au début notre réseau de neurones répond au hasard comme l'enfant, puis à force de lui montrer des informations de météo en entrée, associées à la qualité de l'air mesurée, notre modèle s'améliore et finit par être suffisamment précis !

Voici à quoi ressemble un réseau de neurones basique, avec des entrées, des neurones et une sortie:



Pour les plus connaisseurs, nous utilisons un **réseau de neurones récurrents** RNN (plus précisément, une LSTM pour Long Short-Term Memory) pour **conserver la temporalité des données** et trouver des liens entre différents instants en choisissant quelles informations garder en mémoire ou non. Ceci permet de traiter des données qui forment une suite et qui sont dépendantes les unes des autres.

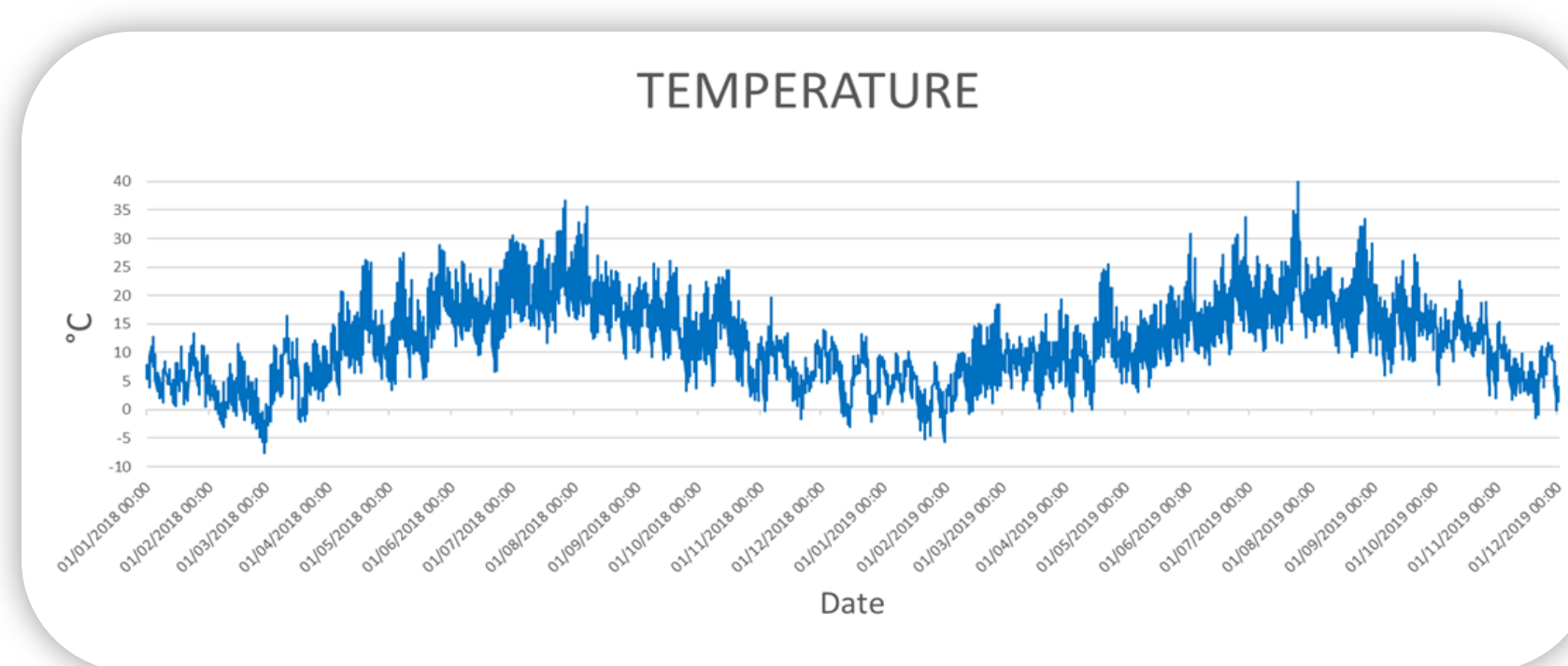


Le terme récurrent ici signifie que chaque sortie d'un de nos neurones est réinjectée dans le suivant, et nous n'obtenons le résultat final qu'une fois que notre information a parcouru chaque neurone de la chaîne.

2 Les données utilisées

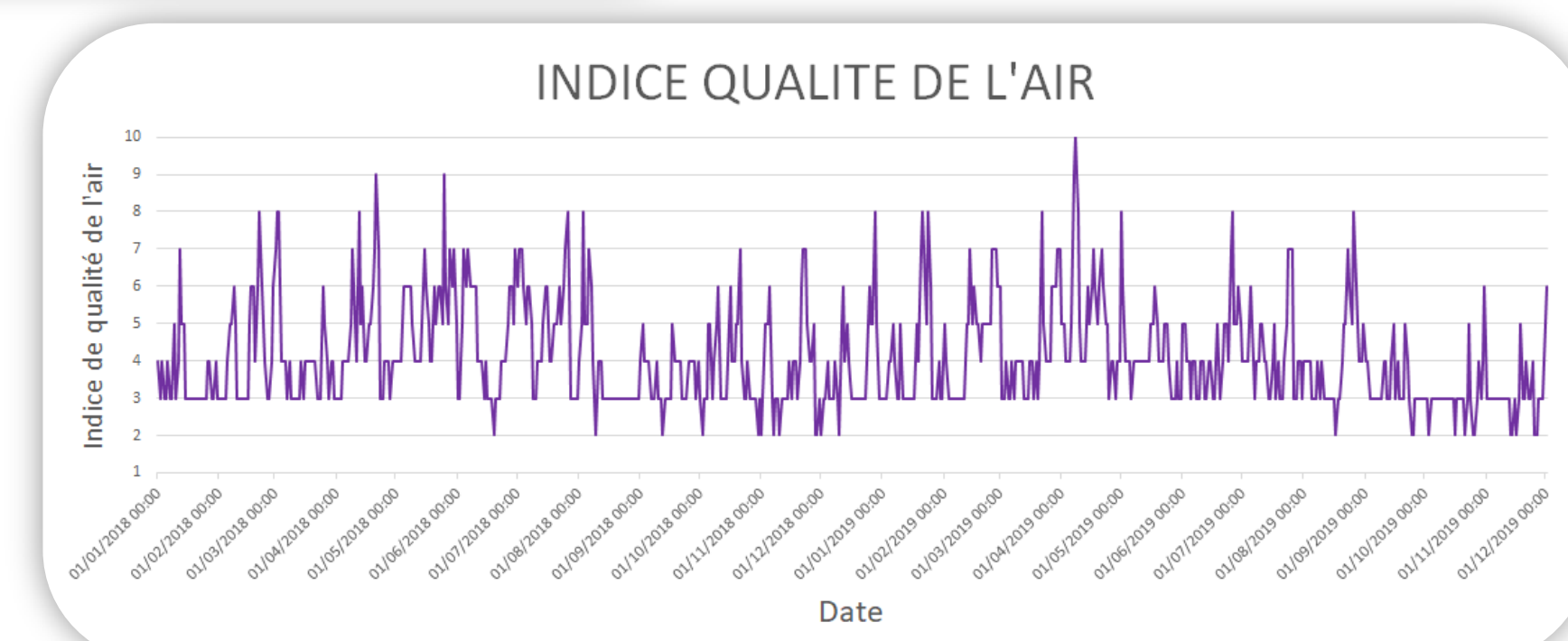
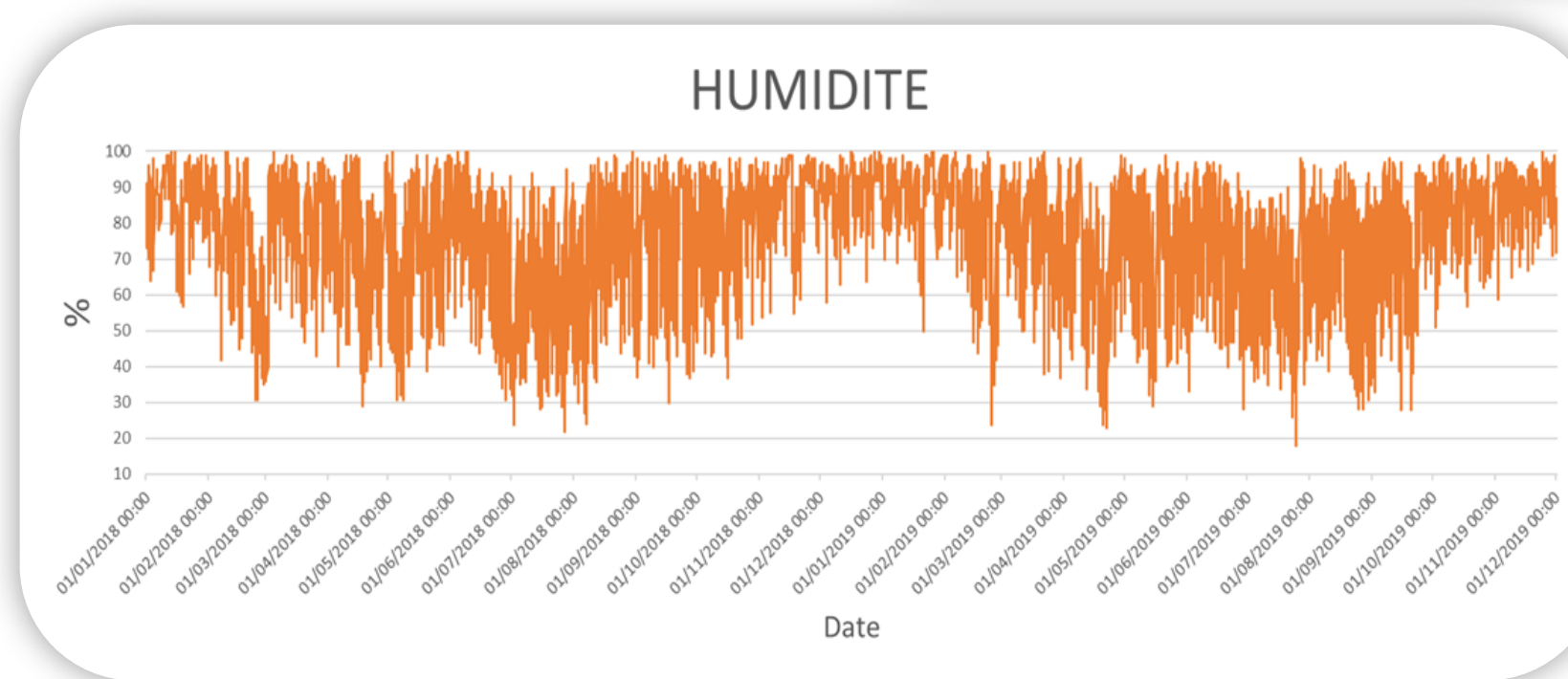
Notre **modèle** **nécessite énormément de données**: plus il a de données pour s'entraîner, plus il sera précis.

Les **capteurs à l'ISEN** ne sont **présents que depuis peu**, et ne fournissent pas une quantité de données satisfaisante: nous utilisons donc **des données publiques de météo France et de la MEL pour l'entraînement**.



MEL
MÉTROPOLÉ
EUROPÉENNE DE LILLE

METEO
FRANCE



3 Application réelle

Après s'être entraîné sur les données météo d'archive, notre **modèle est appliqué aux données collectées par la ruche** sur le toit de l'ISEN.

Nous obtenons un **indice correct dans la plupart des cas**, et en moyenne la différence entre la valeur réelle et la valeur que nous prédisons est de 0,7.

Une fois le modèle obtenu était devenu satisfaisant, nous avons automatisé le processus:

Récupérer les données météo du jour en ligne sur les différentes bases de données, mettre en forme les informations pour notre réseau, et **faire une prédiction tous les jours pour le lendemain**.

Voyez par vous-même le résultat final !

