

**Prévisions météo en Australie**

**Cursus concerné : Data Engineer**

**Description du projet**

Votre jeu de données contient des observations météorologiques quotidiennes provenant de nombreux endroits en Australie.

L’objectif de votre projet est de développer un modèle de classification qui prédira s’il y aura de la pluie le jour suivant. La variable cible est RainTomorrow et contient deux modalités :

* oui : il a plu
* non : il n’a pas plu

**Base de données à télécharger**

Vous pouvez télécharger votre jeu de données en [cliquant ici](https://assets-datascientest.s3-eu-west-1.amazonaws.com/de/total/rains.csv).

**Quelles sont les conditions de validation ?**

Pour valider ce projet, vous devez rendre un jupyter notebook contenant trois parties distinctes. Ces parties correspondent au plan de votre projet. Voici un détail des attendus :

* Audit, exploration et nettoyage des données
  + description détaillée de la base de données et des variables
  + statistiques descriptives
  + traitement des valeurs manquantes et des outliers
  + …
* Visualisation des données
  + entre quatre à six visualisations
  + les visualisations doivent être commentées, légendées et claires
  + …
* Entraînement et évaluation du modèle
  + implémenter au moins deux modèles de classification
  + optimisation
  + comparaison des résultats

**Description des variables**

| **Nom de la colonne** | **Description** |
| --- | --- |
| Date | date |
| Location | lieu |
| MinTemp | température minimale |
| MaxTemp | température maximale |
| Rainfall | précipitation |
| Evaporation |  |
| Sunshine |  |
| WindGustDir |  |
| WindGustSpeed |  |
| WindDir9am | direction du vent à 9h |
| WindDir3pm |  |
| WindSpeed9am | vitesse du vent à 9h |
| WindSpeed3pm |  |
| Humidity9am | humidité à 9h |
| Humidity3pm |  |
| Pressure9am | pression à 9h |
| Pressure3pm |  |
| Cloud9am | nuages à 9h |
| Cloud3pm |  |
| Temp9am | température à 9h |
| Temp3pm |  |
| RainToday | pluie dans la journée |
| RainTomorrow | variable cible: a-t-on de la pluie le lendemain ? |

| **Piste de travail** | - compléter les valeurs manquantes / est-ce que je supprime la colonne  - normaliser/binariser  - supprimer les variables inutiles  - construire des variables temporelles passées |
| --- | --- |