

# ClimateGuard

## Catalogue des données

Le catalogue des données décrit en détail les différents ensembles de données météorologiques, de qualité de l'air, de qualité des sols, de consommation d'électricité, ainsi que les informations générées par l'API. Ce catalogue est élaboré conformément aux principes du RGPD pour garantir la protection des données.

## Sources de Données

### 1. Données Météorologiques et Qualité des sols:

- Time: Timestamp de la mesure.
- Temperature\_2m (°C): Température à 2 mètres au-dessus du sol.
- Relative\_Humidity\_2m (%): Humidité relative à 2 mètres.
- Dew\_Point\_2m (°C): Point de rosée à 2 mètres.
- Apparent\_Temperature (°C): Température ressentie.
- Precipitation (mm): Précipitations.
- Rain (mm): Quantité de pluie.
- Snowfall (cm): Quantité de neige tombée.
- Snow\_Depth (m): Profondeur de neige.
- Weather\_Code (WMO Code): Code météorologique WMO.
- Pressure\_MSL (hPa): Pression au niveau de la mer.
- Surface\_Pressure (hPa): Pression de surface.
- Cloud\_Cover (%): Couverture nuageuse totale.
- Cloud\_Cover\_Low (%): Couverture nuageuse basse.
- Cloud\_Cover\_Mid (%): Couverture nuageuse moyenne.
- Cloud\_Cover\_High (%): Couverture nuageuse haute.
- Et0\_FAO\_Evapotranspiration (mm): Évapotranspiration FAO.
- Vapour\_Pressure\_Deficit (kPa): Déficit de pression de vapeur.
- Wind\_Speed\_10m (m/s): Vitesse du vent à 10 mètres.
- Wind\_Speed\_100m (m/s): Vitesse du vent à 100 mètres.
- Wind\_Direction\_10m (°): Direction du vent à 10 mètres.
- Wind\_Direction\_100m (°): Direction du vent à 100 mètres.
- Wind\_Gusts\_10m (m/s): Rafales de vent à 10 mètres.
- Soil\_Temperature\_0\_to\_7cm (°C): Température du sol de 0 à 7 cm.
- Soil\_Temperature\_7\_to\_28cm (°C): Température du sol de 7 à 28 cm.
- Soil\_Temperature\_28\_to\_100cm (°C): Température du sol de 28 à 100 cm.
- Soil\_Temperature\_100\_to\_255cm (°C): Température du sol de 100 à 255 cm.
- Soil\_Moisture\_0\_to\_7cm (m<sup>3</sup>/m<sup>3</sup>): Teneur en eau du sol de 0 à 7 cm.
- Soil\_Moisture\_7\_to\_28cm (m<sup>3</sup>/m<sup>3</sup>): Teneur en eau du sol de 7 à 28 cm.
- Soil\_Moisture\_28\_to\_100cm (m<sup>3</sup>/m<sup>3</sup>): Teneur en eau du sol de 28 à 100 cm.
- Soil\_Moisture\_100\_to\_255cm (m<sup>3</sup>/m<sup>3</sup>): Teneur en eau du sol de 100 à 255 cm.

## 1. Données Qualite de l'air:

- Time: Timestamp de la mesure.
- PM10 ( $\mu\text{g}/\text{m}^3$ ): Particules fines de diamètre inférieur à 10  $\mu\text{m}$ .
- PM2\_5 ( $\mu\text{g}/\text{m}^3$ ): Particules fines de diamètre inférieur à 2.5  $\mu\text{m}$ .
- Carbon\_Monoxide ( $\mu\text{g}/\text{m}^3$ ): Monoxyde de carbone.
- Nitrogen\_Dioxide ( $\mu\text{g}/\text{m}^3$ ): Dioxyde d'azote.
- Sulphur\_Dioxide ( $\mu\text{g}/\text{m}^3$ ): Dioxyde de soufre.
- Ozone ( $\mu\text{g}/\text{m}^3$ ): Ozone.

## 1. Données Consommation d'électricité:

- Timestamp: Horodatage de la mesure.
- ZoneKey: Clé de la zone géographique.
- CountryCode: Code du pays.
- Consommation: Valeur de la consommation en MW avec unités.
- Production: Liste des sources de production avec valeurs et unités.
- Utilisateur: Informations sur les utilisateurs, leur nombre, type, comportement, etc.

## 1. Données Générées par l'API:

- Endpoint: /consommation-electricite
- Route pour récupérer les données d'électricité pour une ville spécifique.
- Clés: zoneKey, countryCode, consommation, production, timestamp, utilisateur.

# Utilisation des Données

## 3. Données Transformées (Jointure et Nettoyage):

- **Jointure:**
- Les données météorologiques et de qualité de l'air sont jointes sur la colonne de date.
- Conversion de Date:
- La colonne 'time' est convertie en type 'datetime' puis formatée en 'YYYY-MM-DD'.
- Mapping des Codes Météorologiques:
- Les codes météorologiques (WMO) sont transformés en descriptions plus explicites.
- Suppression des Colonnes Inutiles:
- Certaines colonnes inutiles, telles que 'dew\_point\_2m', 'apparent\_temperature', etc., sont supprimées.
- Remplacement des Valeurs Manquantes:
- Les valeurs manquantes sont remplacées par 0.
- 4. Exportation des Données:
- Destination:
- Les données transformées sont exportées au format CSV vers `"/home/FIL_ROUGE/Historical/data/processed/weatherAirSoil.csv"`.

## 1. Traitement de Spark (sparkStreaming)

- **Lecture des Données:**
  - Les données météorologiques sont lues en streaming à partir du topic Kafka "weather-data".
  - Les données de consommation d'électricité sont lues en streaming à partir du topic Kafka "electricity\_consumption".
  - **Jointure des Données:**
  - Les données météorologiques et de consommation d'électricité sont jointes par la colonne 'name' et 'zoneKey' respectivement.
  - **Transformation des Colonnes:**
  - Les colonnes du DataFrame résultant sont sélectionnées et renommées.
  - **Conversion des Horodatages:**
  - Les horodatages sont convertis en format lisible.
- 
- **Exportation vers MongoDB:**
  - Destination:
  - Les données résultantes sont exportées vers une base de données MongoDB nommée "weatherElectro" et une collection nommée "weather\_electricity\_data".
- 
- **Sécurité des Données:**
  - Sécurité des Données:
  - Les procédures de sécurité sont mises en œuvre pour protéger l'intégrité et la confidentialité des données.
  - Une gestion des accès bien implémentée

## Utilisation des Données

1. Entraînement du Modèle de prédiction des pluies : Les données de météo sont utilisées pour entraîner le modèle
2. Visualisations: L'ensemble des données transformées est utilisé à des fins de visualisations et d'analyses