03/11/2021

Mathys Clerget;

Universite du Québec à Chicoutimi

Rapport TP2

Dataset DailyClimate

# Code Permanent :

Mathys Clerget : CLEM14050107

# Introduction :

Ce TP a été réalisé par Mathys Clerget avec comme support : un dataset nommé « Daily Delhi Climate ». C‘est en réalité deux dataset « Daily Delhi Climate Train» et « Daily Delhi Climate Test» qui sont regroupé afin d‘effectué un travail sur l’ensemble des données comprise entre 2013 et 2017.  
(2013-2016-Train et 2017-Test)

Ce dataset possède la température moyenne, l’humidité, la vitesse du vent, et la pression atmosphérique quotidienne de la ville de Delhi entre le 1er janvier 2013 et le 24 avril 2017.

# Description :

## Collecte des données :

Ce dataset provient de Weather Undergroud API.

Il est à disposition de tous sur Kaggle afin de travailler sur les Times Series. Il est à l’origine utilisé par l’Université de Bangalore pour son cours sur l’analyse de données.

Le premier fichier « Daily Delhi Climate Train » est composé de 1463 instances.

Le second fichier « Daily Delhi Climate Test» est lui composé de 115 instances.

## Variables :

Ce dataset possède 5 variables :

|  |  |  |
| --- | --- | --- |
| Variable | Description | Valeurs |
| Date | Date au format YYYY-MM-DD | Date |
| MeanTemp | Température moyenne calculée en fonction des résultats obtenus toutes les trois heures. L’unité n’est pas donnée mais on suppose quelle est en degré Celsius | Float |
| Humidity | L’humidité absolue de la journée. L’unité est le grammes de vapeur d'eau par mètre cube de volume d'air | Float |
| Wind\_Speed | La vitesse du vent. L’unite est le kilomètre par heure | Float |
| Mean\_Pressure | Pression atmosphérique. L’unite est l’atm, atmosphère normale | Float |

## Classes :

J‘ai décidé de séparer les classes par les différentes années que possède ce dataset. Il y a donc 5 classes différentes, les classes de 2013 à 2015 sont composées de 365 valeurs, la classe 2016 de 366 et la classe 2017 est, plus petit, composée de 115 instances.

## Valeurs manquantes ou dupliquées :

Il n‘existe aucune valeur manquante dans ce dataset.

Mais il existe une ligne dupliquée dans ce dataset : Les valeurs du premier janvier 2017. C’est la ligne qui joint les deux fichiers csv

Il a été décidé de conserver la deuxième ligne dupliquée.

# Étapes de création de l‘ensemble de données :

## Fenêtre temporelle glissante :

# Source :

* Dataset :
  + <https://www.kaggle.com/sumanthvrao/daily-climate-time-series-data>
* Autres informations :
  + <https://www.edinformatics.com/math_science/what-is-humidity.html>
  + <https://fr.wikipedia.org/wiki/Atmosph%C3%A8re_(unit%C3%A9)>