

PRÁCTICA 1

Tipología de datos

Edita Caceli Talledo Flores

1. CONTEXTO

EL Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) provee imágenes meteorológicas de 4 satélites: GOES-16, TERRA Y AQUA MODIS, TSM. Para el objetivo del proyecto se decidió extraer las imágenes del GOES-16 del canal infrarrojo. Estas imágenes del GOES están disponibles en 16 canales simples y 4 compuestas [GOES R]. El SENMAHI procesa estas imágenes y las pone a disposición del público a través de su página web [SENAMHI, 2021] para diferentes sectores: todo el globo, Sudamérica, Perú, Perú norte, Perú sur, Perú centro, océano pacífico y otros. Estas imágenes son brindadas en tiempo real con un desfase de 30 minutos y se van actualizando cada 10 minutos.

2. TÍTULO

Imágenes de Satélite del canal IR-GOES/SENAMHI/Perú Obtenidas por Scraping.

3. DESCRIPCIÓN DEL DATASET

El conjunto de datos obtenidos contiene 20 imágenes del canal infrarrojo provenientes del satélite GOES-16 con una diferencia de captura de 10 minutos que van desde las 7 a.m. hasta las 10:10 a.m. Estas imágenes ya han sido procesadas por el SENAMHI y muestran la región de Perú.

4. REPRESENTACIÓN GRÁFICA

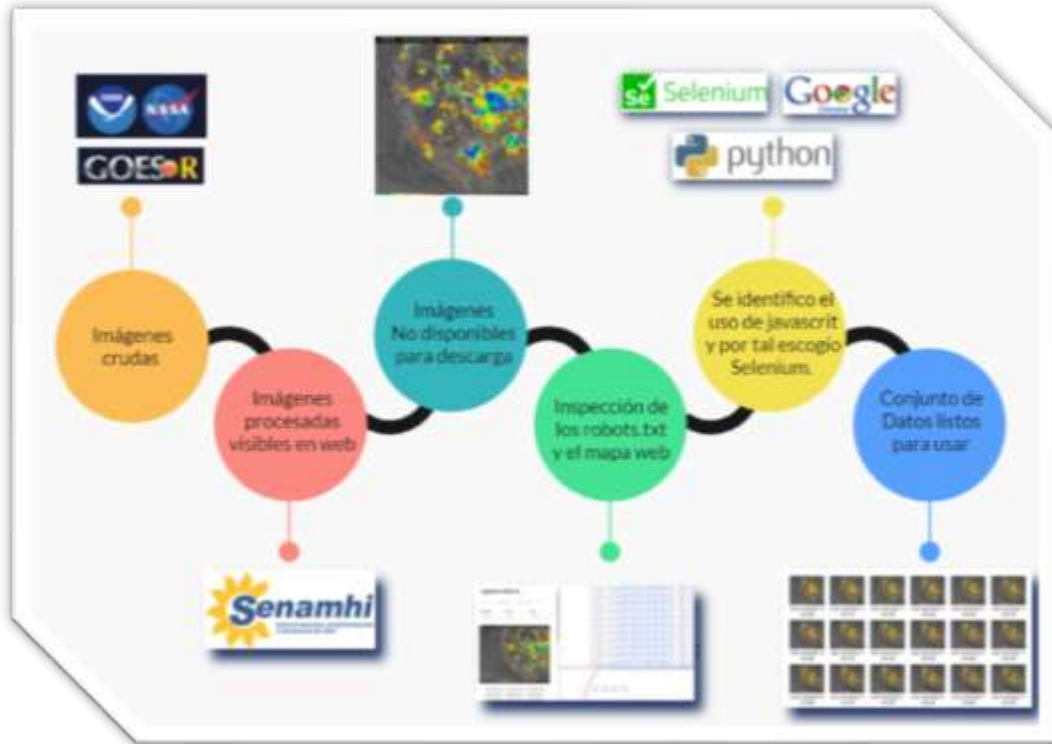


Figura 1. Proceso que atraviesa la imagen.

5. CONTENIDO

El dataset contiene 20 imágenes del canal infrarrojo cuyo nombre identificador PerC14CP202111051200

- Región de captura de la imagen: Per
- Canal empleado: C14
- Fecha de captura de la imagen: CP20211105
- Hora de captura de la imagen: 1200 en GTM

6. AGRADECIMIENTO

- A la Administración Atmosférica y Oceánica Nacional (NOAA por sus siglas en ingles), por las imágenes de sus satélites geoestacionarios GOES-16 distribuidas a través del SENAMHI.
- Al Servicio Nacional de Meteorología e Hidrología del Perú por procesar las imágenes y ponerlas a disposición del público en su página web.

7. INSPIRACIÓN

Las imágenes de satélite de temperatura de brillo son herramientas importantes en el monitoreo de la atmosfera y el seguimiento de eventos extremos. Con ellas se pueden realizar validaciones de modelos numéricos, como también predicciones empleando técnicas de deep learning usando redes neuronales recurrentes.

8. LICENCIA

La licencia seleccionada para el conjunto de datos es **CC BY-SA 4.0 License**. La elección de la licencia se debe a que esta concuerda con los términos y condiciones que establecen los propietarios de la fuente de información. Tal cual lo indican en las siguientes líneas extraídas de su web:

... La Información y Contenidos a los que el usuario puede acceder en la página web son de acceso público y gratuito. El usuario puede utilizar la información para realizar acciones con o sin fines de lucro, siempre aceptando la obligación de reconocer de manera

*expresa que el SENAMHI es el propietario de la información en todo soporte en que dicha información sea usada. Para ello debe indicarse como fuente en todo tipo de soporte la siguiente leyenda: “**Información recopilada y trabajada por el Servicio Nacional de Meteorología e Hidrología del Perú. El uso que se le da a esta información es de mi (nuestra) entera responsabilidad**” [SENAMHI] ...*

Por tal:

- Se debe otorgar el crédito correspondiente; es decir proveer el nombre de creador del conjunto de datos, indicando los cambios que se han realizado sobre el conjunto de datos y asumiendo la responsabilidad que conlleve esta acción.
- Se permite un uso comercial, siempre y cuando se reconozca de manera expresa al propietario original del conjunto de datos.
- Los trabajos derivados o contribuciones realizadas posteriormente sobre el conjunto de datos deben publicarse utilizando esta misma licencia. Así se asegura que el trabajo del autor original se distribuya bajo los términos que el mismo planteo.

9. CÓDIGO

Aquí presento una imagen del código, para acceder al código escrito seguir el siguiente enlace:

<https://github.com/masterscience/WebScraping-Images/blob/main/ScrapingImages.py>

```
@author: edita talledo
"""

from selenium import webdriver
import time
import requests

#-----
def load_requests(source_url):
    r = requests.get(source_url, stream = True)
    if r.status_code == 200:
        aSplit = source_url.split('/')
        ruta = "D:/1_UOC/UOC/1_Tipologia&CicloDeVidaDatos/Practica1/"+aSplit[len(aSplit)-1]
        print(ruta)
        output = open(ruta,"wb")
        for chunk in r:
            output.write(chunk)
        output.close()
# -----

PATH = "C:\Program Files (x86)\chromedriver.exe"
driver = webdriver.Chrome(PATH)

driver.get("https://www.senamhi.gob.pe/?p=satelites-goes16")

images = driver.find_elements_by_tag_name('img')

imagenes = []
i = 0
for imgs in images:
    imagenes.append(imgs.get_attribute('src'))
    print(imagenes)
    load_requests(imagenes[i])
    i = i+1

driver.quit()
```

10. DATASET

El conjunto de datos extraídas por el método de web scraping puede ser obtenidos del siguiente enlace:

<https://github.com/masterscience/WebScraping-Images/blob/main/ImagenesSatelite.rar>

11. REFERENCIAS

- SENAMHI. *Términos y Condiciones de Uso y Privacidad*.
<https://www.senamhi.gob.pe/?p=terminos-condiciones>.
- SENAMHI. 2021. *Tiempo/Satelite. Satélite GOES 16*.
<https://www.senamhi.gob.pe/?&p=satelites-goes16>.
- GOES R. *Geostationary Operational Environmental Satellites – R Series, A collaborative NOAA & NASA program*. <https://www.goes-r.gov/>.
- NOAA. *National Oceanic and Atmospheric Administration*.
<https://www.noaa.gov/>

12. RECURSOS

- Subirats M. y Calvo G. *Web scraping*. Universitat Oberta de Catalunya.
- Mitchell R. 2018. *Web Scraping with Python*. O'Reilly Media, Inc.