

Alumno: Loza Ever

Profesor: Martin Mirabete

Materia: Aprendizaje Automático

Tema: Actividades de la clase 9

El presente informe tiene como objetivo presentar un análisis detallado del dataset meteorológico obtenido de la Estación Astronómica Río Grande (EARG), una institución científica de la Universidad Nacional de La Plata. Este conjunto de datos, que abarca desde enero de 2016 hasta agosto de 2024, proporciona información valiosa sobre diversas variables meteorológicas, tales como temperatura, presión, y velocidad del viento.

El dataset fue obtenido a través de técnicas de scraping, lo que permitió recopilar datos de boletines meteorológicos publicados en línea. La información contenida en este informe es esencial para la modelización de fenómenos climáticos y el desarrollo del proyecto de aprendizaje automático enfocado en la predicción de condiciones meteorológicas. A través de este análisis, se busca facilitar la comprensión de las características del dataset y su aplicabilidad.

Descripción del Dataset

Cantidad de Instancias y Características

El dataset contiene un total de 4230 instancias (filas), donde cada instancia representa un registro de datos meteorológicos diario. A continuación, se detallan las características del dataset:

Variable	Descripción	Tipo de Dato
A	Año	int
M	Mes	int
D	Día	int
temp media	Temperatura media en °C	float
temp max	Temperatura máxima en °C	float
H tmax	Hora de Temperatura Max	Int
M tmax	Hora de Temperatura Max	int
temp min	Temperatura mínima en °C	float
H tmin	Hora de Temperatura Min	Int
M tmin	Minuto de Temperatura Min	int
Viento medio	Viento en km/h	float
dir	Dirección del viento	categorical
rfga	Ráfaga de viento en km/h	float
dir	Dirección de la rafaga	categorical

H	Hora del viento	Int
M	Minuto del viento	int
pres. media	Presión media x hPa	float
lluvia diaria	Lluvia en mm	float
rad max	Radiación en W/m²	float
H rad	Hora de rad max	Int
M rad	Minuto de rad min	Int
UV max	Índice UV	float
H UVmax	Hora de UV max	Int
M UVmax	Minuto de UV max	int

Origen del Dataset

El dataset fue obtenido a través de scraping de la página oficial de la Estación Astronómica Río Grande (EARG), perteneciente a la Universidad Nacional de La Plata. La fuente principal de los datos es el boletín meteorológico disponible en línea. <http://earg.fcaglp.unlp.edu.ar/meteorologia/boletin>

Fecha de Adquisición

Los datos fueron adquiridos mediante un script en Python ejecutado el 27 de octubre de 2024. Este script accede a las URLs correspondientes a los boletines meteorológicos generadas para el rango de fechas ingresado por el usuario, que abarca desde enero de 2016 hasta agosto de 2024.

Proceso de Recopilación y Preprocesamiento

El proceso de recopilación implicó las siguientes etapas:

- **Generación de URLs:** Se generaron automáticamente las URLs de los boletines meteorológicos desde enero de 2016 hasta agosto de 2024, mediante scraping, insertando el año y el mes a extraer los datos necesarios para el futuro modelo.
- **Scraping de Datos:** Se utilizó la librería requests para realizar solicitudes HTTP y BeautifulSoup para parsear el contenido HTML, extrayendo los datos relevantes sobre el clima.
- **Almacenamiento:** Finalmente, los datos fueron almacenados en un archivo Excel y guardados en el repositorio de GitHub [Politecnico Malvinas-Proyecto ML](#), facilitando su uso en el proyecto de Aprendizaje Automático. Los archivos **scraping.py** y **boletin_2016_01-2024_08.xlsx** fueron subidos a la carpeta **data/raw** del repositorio. Mientras que los documentos de las entregas se encuentran en **docs/**.

Consideraciones Finales

Este dataset es de dominio público y se obtiene de una fuente confiable, lo que lo hace adecuado para su uso en proyectos académicos y de investigación. La información meteorológica puede ser de gran utilidad para modelar fenómenos climáticos y desarrollar proyectos de aprendizaje automático relacionados con la predicción de condiciones meteorológicas.