

# Descripción del Dataset para el trabajo

Para esta ocasión, se va a contar con datos meteorológicos obtenidos a partir de la plataforma Open-Meto, el cual es una página que aporta de forma libre y gratuita datos históricos meteorológicos.

La misma plataforma permite seleccionar las variables meteorológicas que se requieren para el armado del modelo y el lugar o ubicación en donde se quieren obtener

En cuyo caso, para el dataset inicial, se contó con 6 variables las cuales conformaran las 6 columnas del dataframe. Estas características o features son:

- **Date:** Corresponde a la fecha que se tomó dichos valores meteorológicos y su formato de fecha es AAAA-MM-DD:HH:MM:SS
- **Humedad del suelo:** Esta variable describe la humedad volumétrica del suelo desde los 0 a 1cm de distancia y su unidad o formato es:  $\text{mm}^3/\text{mm}^3$
- **Temperatura:** Se caracteriza por describir la temperatura del ambiente expresada en  $^{\circ}\text{C}$
- **Humedad relativa:** Muestra el porcentaje de humedad que hay en el ambiente
- **Precipitación:** Esta característica describe la cantidad de lluvia que cayó en la hora y fecha tomada y su unidad es en mm (milímetros).
- **Velocidad del viento:** Esta última variable expresa que tan rápido fluye el viento en el ambiente y esta expresada en km/h.

Por lo que se puede demostrar que estos tipo de datos se corresponde a datos estructurados y numéricos.

Además, estos datos fueron obtenidos seleccionando las ubicaciones de la isla de Tierra del Fuego y las localidades del lado Argentino: Rio Grande, Tolhuin y Ushuaia.

El rango de datos que se utilizó para este trabajo abarca desde la fecha del 16 de Noviembre del año 2022 hasta el 21 de Mayo del 2025 (Nov2022-May2025) y la datación del mismo se obtuvo en cada hora durante todo ese periodo, brindando una cantidad total de 22032 filas en el dataframe.

A continuación se cita el enlace de la plataforma Open-Meteo con las variables características y las locaciones seleccionadas.

[https://open-meteo.com/en/docs/historical-forecast-api?latitude=-54,-53.7877,-54.5108,-54.8108&longitude=-69,-67.7097,-67.1955,-68.3159&hourly=temperature\\_2m,soil\\_moisture\\_0\\_to\\_1cm,relative\\_humidity\\_2m,precipitation,wind\\_speed\\_10m&start\\_date=2022-12-16&end\\_date=2025-05-21&models=best\\_match&timezone=auto#hourly\\_weather\\_variables](https://open-meteo.com/en/docs/historical-forecast-api?latitude=-54,-53.7877,-54.5108,-54.8108&longitude=-69,-67.7097,-67.1955,-68.3159&hourly=temperature_2m,soil_moisture_0_to_1cm,relative_humidity_2m,precipitation,wind_speed_10m&start_date=2022-12-16&end_date=2025-05-21&models=best_match&timezone=auto#hourly_weather_variables)

Una ventaja que tiene esta plataforma es que te genera un script en Python que permite extraer los datos con las características que seleccionaste para empezar a trabajar o desarrollar con un modelo de Machine Learning o de Ciencia de Datos.

La idea con estos datos es poder generar la etiqueta o variable objetivo que describa el nivel de riesgo de incendio, para ello analizará con un criterio o regla general, los datos meteorólogos y asignará un cierto nivel del riesgo de incendio.

De esta forma, los datos quedan de la forma supervisada por lo que se procede con aplicar alguna técnica de clasificación.