

Entrega 2: Descripción del Dataset y Origen

Predicción del Riesgo Operativo en Aerogeneradores en Condiciones Climáticas Extremas en Tierra del Fuego

Descripción de los Datasets

1. Dataset Operativo (SCADA)

- Descripción:**
Contiene registros operativos de aerogeneradores, medidos a intervalos de 10 minutos, con información sobre producción de energía y condiciones de operación.
- Cantidad de instancias:**
50,530 filas.
- Características (columnas):**

Variable	Descripción	Tipo de dato
Date/Time	Fecha y hora de la medición	datetime
LV ActivePower (kW)	Potencia activa medida	float64
Wind Speed (m/s)	Velocidad del viento medida	float64
Theoretical Power Curve (kWh)	Potencia teórica esperada	float64
Wind Direction (°)	Dirección del viento medida	float64
- Tipo de datos:**
Mixtos: datetime y float64.
- Observaciones relevantes:**
 - No hay valores nulos en el dataset después del preprocesamiento.
 - Se eliminaron filas duplicadas.
 - Se realizaron conversiones de fecha y hora al tipo datetime.

2. Dataset Meteorológico

- **Descripción:**
Datos climáticos históricos de Río Grande, Tierra del Fuego, obtenidos de Meteostat.org, con variables meteorológicas diarias.
- **Cantidad de instancias:**
352 días.
- **Características (columnas):**

Variable Descripción		Tipo de dato
date	Fecha de la medición	datetime
tavg	Temperatura promedio diaria (°C)	float64
tmin	Temperatura mínima diaria (°C)	float64
tmax	Temperatura máxima diaria (°C)	float64
prcp	Precipitación (mm)	float64
wdir	Dirección promedio del viento (°)	float64
wspd	Velocidad promedio del viento (km/h)	float64
pres	Presión atmosférica promedio (hPa)	float64

- **Tipo de datos:**
datetime y float64.
- **Observaciones relevantes:**
 - Se eliminaron variables con muchos valores nulos como snow y tsun.
 - Se eliminaron duplicados.
 - Se estandarizaron los tipos de datos.

Origen de los Datos

Dataset	Fuente	Fecha de adquisición	Proceso de recopilación
SCADA	https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset	31 de mayo de 2025	Datos históricos de sensores SCADA de aerogeneradores.
Meteorológico	https://meteostat.net/es/place/ar/rio-grande?s=87934&t=2018-01-01/2018-12-18#google_vignette	31 de mayo de 2025	Extracción de datos meteorológicos públicos vía plataforma web.

Notas de Preprocesamiento

- Se convirtieron las columnas de fecha y hora a datetime.
- Se eliminaron duplicados en ambos datasets.
- Se eliminaron valores nulos de columnas irrelevantes para el modelo.
- Se unificaron ambos datasets por fecha (Date/Time y date) para su posterior análisis y modelado.
- **Análisis Exploratorio:** pendiente de realizar en la siguiente etapa del proyecto.

Repositorio del Proyecto

El proyecto y los datasets están almacenados en el repositorio GIT:

 https://github.com/sfvillegas/riesgo_operativo_aerogeneradores