****

**DESCRIPCIÓN DEL DATASET Y ORIGEN**

**Descripción Completa del Dataset**

En este proyecto se utilizarán varios datasets integrados para construir un **dataset maestro** que permita entrenar modelos de aprendizaje automático enfocados en la **predicción de la producción de trucha en Tierra del Fuego**. A continuación, se detallan los datasets utilizados:

**Dataset 1: Clima y Temperatura en Río Grande (Meteostat)**

* **Descripción:**  
  Dataset histórico de variables climáticas registradas en la estación meteorológica de Río Grande, Tierra del Fuego, Argentina. Contiene información diaria y horaria relevante como temperatura, precipitación, viento, y presión atmosférica.
* **Características:**
  + time (fecha y hora) – **Datetime**
  + temp (temperatura del aire en °C) – **Float**
  + dwpt (punto de rocío en °C) – **Float**
  + rhum (humedad relativa en %) – **Integer**
  + prcp (precipitación en mm) – **Float**
  + snow (nieve en mm) – **Float**
  + wdir (dirección del viento en grados) – **Integer**
  + wspd (velocidad del viento en km/h) – **Float**
  + wpgt (ráfaga máxima de viento en km/h) – **Float**
  + pres (presión atmosférica en hPa) – **Float**
  + tsun (duración de sol en minutos) – **Integer**
  + coco (condiciones de nubosidad) – **Categorical**
* **Cantidad de Instancias:**  
  ~10.000 registros (según frecuencia de datos diaria/horaria entre 1996-2023).
* **Tipo de Archivo:** CSV
* **Fuente:**  
  Meteostat Río Grande y Ushuaia
* **Preprocesamiento Realizado:**
  + Eliminación de registros nulos o incompletos.
  + Agrupación de datos diarios a promedios anuales para correlacionar con producción acuícola.

**Dataset 2: Producción Acuícola Argentina (FAO)**

* **Descripción:**  
  Datos de producción acuícola a nivel nacional en toneladas, discriminados por especie y año. Se extrajo específicamente la información relacionada a la **trucha arcoíris (Oncorhynchus mykiss)**.
* **Características:**
  + COUNTRY.UN\_CODE (Código país ONU) – **Integer**
  + SPECIES.ALPHA\_3\_CODE (Código FAO especie) – **String**
  + PERIOD (Año) – **Integer**
  + VALUE (Toneladas producidas) – **Float**
* **Cantidad de Instancias:**  
  ~70 registros (Argentina, años 1950-2022, filtrando solo Trucha Arcoíris).
* **Tipo de Archivo:** CSV
* **Fuente:**  
  [FAO - Producción Acuícola](https://www.fao.org/fishery/statistics-query/en/aquaculture/aquaculture_quantity)
* **Preprocesamiento Realizado:**
  + Filtro por país = Argentina.
  + Filtro por especie = *Oncorhynchus mykiss*.
  + Eliminación de registros con valores nulos o anómalos.

**Dataset 3: Producción Histórica de Acuicultura Argentina (MAGyP)**

* **Descripción:**  
  Informe histórico compilado por el Ministerio de Agricultura, Ganadería y Pesca de Argentina (MAGyP), contiene la evolución de la acuicultura argentina desde 1996. Incluye datos de producción de trucha a nivel provincial.
* **Características:**
  + Año (Año de producción) – **Integer**
  + Provincia – **String**
  + Especie (nombre común) – **String**
  + Producción en toneladas – **Float**
* **Cantidad de Instancias:**  
  ~250 registros (varios años × provincias).
* **Tipo de Archivo:** PDF (convertido a CSV para análisis).
* **Fuente:**  
  [MAGyP - Acuicultura Histórica](https://www.magyp.gob.ar/sitio/areas/acuicultura/estadisticas/_archivos/000000_Acuicultura%20Argentina%20-%20Recopilaci%C3%B3n%20hist%C3%B3rica%20de%20su%20desarrollo%20.pdf)
* **Preprocesamiento Realizado:**
  + Extracción de tablas del PDF con OCR.
  + Conversión a CSV.
  + Filtrado por provincia = *Tierra del Fuego*.
  + Limpieza de caracteres no válidos y corrección de formatos numéricos.

**Dataset 4: Perfil Sectorial de Pesca y Acuicultura en Argentina (FAO)**

* **Descripción:**  
  Perfil nacional de Argentina sobre la pesca y acuicultura, que contiene información contextual sobre políticas, evolución histórica de especies cultivadas, infraestructura, exportaciones y datos ambientales relevantes.
* **Características:**
  + Informe descriptivo sin estructura tabular estándar.
* **Cantidad de Instancias:**  
  No aplica — es un informe PDF de referencia.
* **Tipo de Archivo:** PDF
* **Fuente:**  
  [FAO Country Profile Argentina](https://www.fao.org/fishery/zh/countrysector/ar/es?lang=es)
* **Preprocesamiento Realizado:**
  + Revisión manual del contenido para extracción de datos relevantes (ej. rangos de temperatura óptimos, principales zonas de cultivo, políticas de fomento).
  + No se convierte a CSV directamente, usado como soporte documental.

**Dataset 5: Información Hidrológica (Instituto Nacional del Agua - INA)**

* **Descripción:**  
  Información hidrológica sobre caudales, niveles de lagos, precipitaciones acumuladas en Argentina, incluyendo la provincia de Tierra del Fuego. Fundamental para evaluar recursos hídricos disponibles para acuicultura.
* **Características:**
  + Información puntual de caudales y niveles.
  + Variables como: caudal medio mensual, precipitaciones acumuladas.
* **Cantidad de Instancias:**  
  Variable, accedida por reporte.
* **Tipo de Archivo:** Online reportes / visualizaciones.
* **Fuente:**  
  [INA Hidrología Argentina](https://www.hidro.gov.ar/Ceado/EF/EF.asp?op=3&v=21)
* **Preprocesamiento Realizado:**
  + Extracción manual de series históricas relevantes para análisis correlativo.
  + Uso como variable auxiliar (p.ej. estacionalidad hídrica).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Fuente Oficial | Formato Original | Fecha de Adquisición | Notas de Pre procesamiento |
| Clima Río Grande y Ushuaia | Meteostat | CSV | Mayo 2025 | Agrupación diaria y anual, limpieza de nulos |
| Producción Acuícola Argentina | FAO | CSV | Mayo 2025 | Filtrado por país y especie |
| Acuicultura Histórica Argentina | MAGyP | PDF → CSV | Mayo 2025 | Extracción de tablas, limpieza manual |
| Perfil de Pesca Argentina | FAO | PDF | Mayo 2025 | Revisión y extracción de datos clave |
| Hidrología Argentina | INA | Online reportes | Mayo 2025 | Captura manual de series hidrológicas |

**CONCLUSIÓN:**

Con la integración de estos datasets se construirá un **dataset maestro** que permitirá analizar la evolución histórica de la producción de trucha arcoíris en Tierra del Fuego, identificar patrones climáticos e hídricos, y entrenar modelos predictivos de regresión y clasificación, cumpliendo los objetivos planteados en el proyecto.