

## Descripción del Dataset y Origen

**Alumno:** GARCÍA BARQUET, Darío Agustín

**Materia:** Aprendizaje Automático

**Profesor:** MIRABETE, Martín

### Desarrollo:

El conjunto de datos que voy a utilizar para desarrollar mi proyecto fueron provistos por la Secretaría de Energía de la Nación Argentina. Estos datos son públicos y de uso libre.

El mismo especifica los detalles mensuales de la producción anual de gas y de petróleo de los pozos ubicados en Argentina. Se utilizará el set mas actualizado, que corresponde a lo que va del año 2023.

El data set contiene 716462 entradas y 38 columnas. A continuación, un detalle del dataset “crudo”, aclarando cada una de sus columnas con su referencia:

'idempresa': De tipo objeto. Es el identificador único que se le da a cada una de las empresas concesionadas e involucradas en la extracción del hidrocarburo.

'anio': De tipo entero. El año en el cual se registraron los datos. En éste caso, al tratarse de un dataset de este mismo año, todos los valores corresponden al 2023.

'mes': De tipo entero. El mes es en el cual se registraron los datos. Refiere a en qué mes se registró el dato de productividad.

'idpozo': De tipo entero. Identificador único para cada pozo productor.

'prod\_pet': De tipo flotante. Cantidad de petróleo producido por el pozo.

'prod\_gas': De tipo flotante. Cantidad de gas producido por el pozo.

'prod\_agua': De tipo flotante. Cantidad de agua producida por el pozo.

'iny\_agua': De tipo flotante. Cantidad de agua inyectada en el pozo.

'iny\_gas': De tipo flotante. Cantidad de gas inyectado en el pozo.

'iny\_co2': De tipo flotante. Cantidad de dióxido de carbono inyectado en el pozo.

'iny\_otro': De tipo flotante. Cantidad de otro tipo de fluido inyectado en el pozo.

'tef': De tipo flotante.

'vida\_util': De tipo flotante.

'tipoextraccion': De tipo objeto. El sistema de extracción que se está utilizando para facilitar la producción.

'tipoestado': De tipo objeto. El estado en el que se encuentra el pozo en el registro.

'tipopozo': De tipo objeto. Hace referencia a para qué es utilizado el pozo. En nuestro caso usaremos únicamente los productores de gas o petróleo.

'observaciones': De tipo objeto. Observaciones varias que se realizan sobre el pozo.

'fechaingreso': De tipo objeto. La fecha de ingreso de los datos al registro.

'rectificado': De tipo objeto. Indica si el pozo recibió un proceso de rectificación.

'habilitado': De tipo objeto. Identifica si el pozo está habilitado al momento del registro.

'idusuario': De tipo entero. Identificador del usuario que cargó el dato.

'empresa': De tipo objeto. Empresa concesionaria encargada de la extracción.

'sigla': De tipo objeto. Sigla utilizada para identificar a la concesión.

'formprod': De tipo objeto. Identificador abreviado para cada una de las formaciones productivas relacionadas al sistema petrolero del cual se realiza la explotación de hidrocarburos.

'profundidad': De tipo flotante. Profundidad máxima o productiva del pozo.

'formacion': De tipo objeto. Identificador completo para cada una de las formaciones productivas relacionadas al sistema petrolero del cual se realiza la explotación de hidrocarburos.

'idareapermisiconcesion': De tipo objeto. Identificador del área en la cual se es dado el permiso o la concesión a cada una de las empresas respectivamente.

'areapermisiconcesion': De tipo objeto. Área en la cual se es dado el permiso o la concesión a cada una de las empresas respectivamente.

'idareayacimiento': De tipo objeto. Identificador del área en la cual se encuentra el yacimiento productivo.

'areayacimiento': De tipo objeto. Área en la cual se encuentra el yacimiento productivo.

'cuenca': De tipo objeto. Cuenca a la que pertenece cada una de las formaciones productivas.

'provincia': De tipo objeto. Provincia en la que se encuentra el pozo productor.

'tipo\_de\_recurso': De tipo objeto. Hace referencia a si se utiliza metodo convencional o no convencional.

'proyecto': De tipo objeto.

'clasificacion': De tipo objeto. Hace referencia a para qué está siendo utilizado el pozo al momento del registro. Nosotros utilizaremos aquellos pozos que sean utilizados para explotación y exploración.

'subclasificacion': De tipo objeto.

'sub\_tipo\_recurso': De tipo objeto. Hace referencia a si se trata de una producción de shale o tight gas.

'fecha\_data': De tipo objeto. Fecha a la que corresponde el dato. Nótese que la fecha anterior ("fechaingreso") es la fecha en la que se realizó el registro en el sistema únicamente.

Cabe resaltar que muchos de estos campos o columnas no serán utilizados para el proyecto porque no son relevantes. También, el dataset será sometido a un severo proceso de manejo de outliers, normalización, binarización y de preprocesamiento en general.