



Estimación del tiempo restante de explotación de yacimientos de turba en Tierra del Fuego mediante Aprendizaje Automático.

ENTREGA 2

Descripción del Dataset y Origen

Cátedra: Aprendizaje Automático – 1C 2025

Docente: Martín Mirabete

**Alumno
Sergio Andrés SANCHEZ**

Proporcione una descripción completa del dataset, incluyendo la cantidad de instancias, características (columnas), tipos de datos, y cualquier información relevante.

El Dataset provisto contempla expedientes tramitados por explotación de Turba Rubia y Negra por cada uno de los yacimientos desde enero de 2020 a diciembre de 2024.

Este Dataset tiene 1997 instancias de las cuales 23 instancias son las que están en duda y por eso se solicito una verificación / explicación de a que se refieren esas instancias en el Dataset, ya que en la mayoría de las características de la línea están sin información.

También se observó que no se tiene el listado de los yacimientos con la superficie total y lo restante por extraer como para poder comenzar con el trabajo del proyecto.

Este Dataset cuenta con 27 características, de las cuales 2 son de tipo categóricas y las restantes son de tipo numéricas. De las 2 de tipo categóricas, una solo posee el tipo de embalaje y nombre científico de la turba, que además es el único que se extrae (Turba embolsada - Musgus Sphagnum. La otra característica categórica, refiere a si la extracción mencionada fue informada su comercialización o no.

Las características que posee este Dataset son las siguientes:

Variable	Tipo	Definición
Area	Numérica	Superficie del yacimiento, en hectáreas
Vol_N	Numérica	Volumen de turba negra a 2023
Vol_R	Numérica	Volumen de turba rubia a 2023
Vol_total	Numérica	Volumen total de turba a 2023
Fecha	Numérica	Año de cuantificación (2023)
id	Numérica	valor de identificación (no tener en cuenta)
uid	Numérica	valor de identificación (no tener en cuenta)
poligono	Numérica	número del polígono
id_2	Numérica	valor de identificación (no tener en cuenta)
producto	Categórica	tipo de mineral extraído
num_expte	Numérica	número de expediente de regalías mineras
ano	Numérica	año de declaración jurada
mes	Numérica	mes de declaración jurada
volumen_produccion	Numérica	volumen de mineral extraído
trimestre	Numérica	trimestre de la declaración jurada
regalias	Numérica	importe de regalías mineras liquidado
tasa_insp_fisc	Numérica	importe de tasas de fiscalización liquidado
ano_comer	Numérica	año de comercialización territorio nacional continental
mes_comer	Numérica	mes de comercialización a territorio nacional continental
productor_comer	Numérica	número de productor
producto_comer	Numérica	producto comercializado
volumen_comercializado	Numérica	volumen comercializado
valor_fob_usd	Numérica	valor fob expresado en dólares (valor de mineral puesto en aduana)
valor_fob_ars	Numérica	valor fob expresado en pesos (valor de mineral puesto en aduana)
tasas_comercial	Numérica	tasas abonadas en concepto de emisión de certificado exportación
tasa_cambio	Numérica	relación peso dólar, al momento de emitir el certificado de exportación
sin_comercializacion	Categórica	valor TRUE (no comercializó) valor FALSE (comercializó)

Al parecer para anonimizar el Dataset, en id debería figurar el Identificador que ellos poseen del yacimiento sin mencionar cual es y por un error el mismo no salió en el Dataset.

En el transcurso de la semana del 02/06 al 06/06 se estaría solucionando el inconveniente.

Lamentablemente esto sucede por trabajar con información que no se tiene en Dataset ya armados y por eso estas demoras en poder hacer los análisis, pero con la idea y lo ventajoso de realmente estar trabajando con datos reales provenientes del gobierno y actualizados a la fecha.

Informe sobre el origen del dataset, es decir, de dónde provienen los datos. Esto puede incluir la fuente, la fecha de adquisición y cualquier proceso de recopilación o preprocesamiento que haya realizado.

Origen: El Dataset proviene de una base de datos en Postgre de la Dirección General de Desarrollo Minero dependiente del Ministerio de Producción y Ambiente del Gobierno de la Provincia de Tierra del Fuego, Antártida e Islas del Atlántico Su.

Fecha de disponibilidad del Dataset: Viernes 16 de mayo de 2025.

Recopilación y Preprocesamiento: Por el momento se realizó un análisis superficial de los datos y se encontró que hay datos faltantes en relación al tamaño total de cada yacimiento, por lo que se solicitó al área que proveyó el Dataset proporcionar esa información faltante y se está esperando que ese nuevo Dataset con la información solicitada pueda ser entregado para poder continuar con el análisis.

También estarían faltando los Id de cada uno de los yacimientos para así mantener anonimizado el Dataset. Este Id es el que una vez finalizado el análisis y el Modelo de Aprendizaje Automático, el mismo se pueda ceder al Ministerio de Producción y Ambiente y puedan unir los datos a través de ese Id y así poder utilizar el Modelo de acá en más para diferentes análisis, así como poder ir ajustando el modelo si así se lo requiriese.