



**CENTRO POLITÉCNICO SUPERIOR
MALVINAS ARGENTINAS**

Trabajo Final

Aprendizaje Automático

Regalías de Petróleo Crudo

Materia: Aprendizaje Automático

Alumno: Biason Franco 35.631.316

Año: Mayo 2025



Contexto

La Provincia de Tierra del Fuego es una de las principales regiones productoras de hidrocarburos en Argentina. Los ingresos provenientes de las regalías petroleras (pagos realizados por las empresas que explotan el recurso) constituyen una fuente crítica para el financiamiento del presupuesto provincial. Estas regalías están directamente relacionadas con factores como el volumen de producción, el tipo de crudo y el precio de venta, los cuales presentan variaciones mensuales.

En este contexto, contar con herramientas que permitan anticipar el comportamiento futuro de las regalías resulta fundamental para la gestión pública y la planificación financiera de la provincia. Si bien existen métodos tradicionales para realizar estimaciones económicas, este proyecto propone abordar el problema desde un enfoque basado en Aprendizaje Automático.

Objetivo del Proyecto

El objetivo del proyecto es predecir el monto mensual de regalías petroleras que recibirá la Provincia de Tierra del Fuego, utilizando técnicas de Aprendizaje Supervisado y tomando como base datos históricos provenientes del portal oficial del Ministerio de Economía (datos.energia.gob.ar). La intención es capturar patrones que relacionan producción, tipo de crudo y precios con los ingresos por regalías, para después proyectar posibles escenarios futuros.

Objetivo General

Desarrollar un modelo de Aprendizaje Automático capaz de predecir el monto mensual de regalías petroleras recibidas por la Provincia de Tierra del Fuego, utilizando datos históricos de producción, tipo de crudo y precios.



Objetivos Específicos

- Filtrar y preparar un dataset con registros mensuales exclusivamente de Tierra del Fuego.
- Analizar las variables que afectan directamente el monto de regalías (producción, tipo de crudo, precio).
- Aplicar modelos de regresión supervisada para predecir valores continuos de ingresos mensuales.
- Evaluar el desempeño de los modelos mediante métricas apropiadas (MAE, MSE, R^2).
- Interpretar los resultados obtenidos para su posible uso en la planificación presupuestaria.

Tipo de Problema

Se trata de un problema de regresión supervisada, ya que el objetivo es predecir el monto mensual de regalías petrolera, que es una variable numérica continua.

Modelos Posibles a Utilizar

Para abordar el problema, se consideran los siguientes modelos de regresión incluidos en la librería scikit-learn:

- Regresión Lineal.
- Árbol de Decisión para Regresión.
- Bosques Aleatorios.
- Gradient Boosting Regressor.
- Support Vector Regressor.

Durante el desarrollo del proyecto, voy a evaluar el rendimiento de cada uno de estos modelos para determinar cuál se adapta mejor al problema planteado.



Fuente de Datos

El dataset fue extraído del portal de datos abiertos del Ministerio de Economía:

<http://datos.energia.gob.ar>

Archivo seleccionado: "Informe Regalías CRUDO.xlsx"

Éste archivo contiene información mensual por provincia sobre:

- Fecha (AÑO / MES)
- Provincia
- Precio (por m³)
- Producción (m³)
- Monto total de regalías
- Tipo de crudo



Descripción del Dataset

Cantidad de instancias (filas): 294

Cantidad de características (columnas): 6

Columna	Tipo de dato	Descripción
FECHA	Fecha (AAAAMM)	Mes y año del registro
PROVINCIA	Texto	Provincia argentina que registra el ingreso de regalías
PRECIO	Numérico (float)	Precio del crudo en pesos argentinos por m³
PRODUCCIÓN	Numérico (float)	Volumen de crudo producido en m³
MONTO_REGALÍAS	Numérico (float)	Monto total en pesos argentinos recibido en concepto de regalías
TIPO_CRUDO	Texto	Tipo de crudo extraído (Ej: Escalante, Medanito, etc.)

Preprocesamiento Realizado

Para esta entrega se aplicaron las siguientes acciones de limpieza y filtrado:

- Se filtraron los registros exclusivamente de la provincia de Tierra del Fuego, que es el foco del análisis.
- Se eliminaron filas incompletas o con valores nulos en las columnas clave (PRECIO, PRODUCCIÓN, MONTO_REGALÍAS).
- Se verificó la consistencia temporal de la columna FECHA para asegurar un orden cronológico correcto.
- Se preparó el dataset para su lectura con bibliotecas como pandas, facilitando la visualización y análisis de patrones.



Conclusión

El dataset resultante permite trabajar con un enfoque de regresión supervisada, siendo la variable objetivo (MONTO_REGALÍAS) un valor continuo. Esto facilita la implementación de diversos modelos de regresión (como Random Forest, Gradient Boosting, etc.) y el posterior análisis del desempeño mediante métricas como MAE, RMSE o R2.