Contenido

[Contexto 2](#_Toc87122551)

[Objetivo 2](#_Toc87122552)

[Fuente de datos 2](#_Toc87122553)

[Preparación del entorno de ejecución 2](#_Toc87122554)

[1) Importar las bibliotecas necesarias 2](#_Toc87122555)

[2) Importar los datos 2](#_Toc87122556)

[Gráfica de las mediciones de aceite 3](#_Toc87122557)

[Aplicación del algoritmo 3](#_Toc87122558)

[Se seleccionan las variables predictoras (X) y la variable a pronosticar (Y) 3](#_Toc87122559)

[Se entrena el modelo a través de una regresión lineal múltiple 4](#_Toc87122560)

[Se genera el pronóstico 4](#_Toc87122561)

[Valores pronosticados 5](#_Toc87122562)

[Obtención de los coeficientes, intercepto, error y Score 5](#_Toc87122563)

[Comparación del modelo de pronóstico 5](#_Toc87122564)

[Proyección de los valores reales y pronosticados 6](#_Toc87122565)

[Nuevos pronósticos 6](#_Toc87122566)

[Conclusiones 6](#_Toc87122567)

# Contexto

Objetivo: Obtener el pronóstico de la saturación de aceite remanente (ROS, Residual Oil Saturation) a partir de las cuatro mediciones de los registros geofísicos convencionales: (RC1) Registro Neutrón, (RC2) Registro Sónico, (RC3) Registro Densidad-Neutrón, y (RC4) Registro Densidad Corregido por Arcilla.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

## Fuente de datos

Diagrama

Descripción generada automáticamenteSe tienen mediciones de registros geofísicos convencionales: RC1 (Registro Neutrón), RC2 (Registro Sónico), RC3 (Registro Densidad-Neutrón) y RC4 (Registro Densidad -corregido por arcilla-)

* Para la toma de registros se cuenta con cables electromecánicos, sensores, dispositivos eléctricos y sistemas computarizados.
* Se procesan los datos a través de los sensores, para luego ser enviados a la superficie por medio del cable.
* RC1 = Registro Neutrón
* RC2 = Registro Sónico
* RC3 = Registro Densidad-Neutrón
* RC4 = Registro Densidad (corregido por arcilla)

# Preparación del entorno de ejecución

## Importar las bibliotecas necesarias

import pandas as pd               # Para la manipulación y análisis de datos

import numpy as np                # Para crear vectores y matrices n dimensionales

import matplotlib.pyplot as plt   # Para la generación de gráficas a partir de los datos

import seaborn as sns             # Para la visualización de datos basado en matplotlib

%matplotlib inline

## Importar los datos

Fuente de datos: RGeofisicos.csv

# Si se usa Google Colab

#from google.colab import files

#files.upload()

#from google.colab import drive

#drive.mount('/content/drive')

RGeofisicos = pd.read\_csv("RGeofisicos.csv")

RGeofisicos

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico, Calendario

Descripción generada automáticamente Interfaz de usuario gráfica, Aplicación, Tabla, Calendario

Descripción generada automáticamente

# Gráfica de las mediciones de aceite

plt.figure(figsize=(20,5))

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC1'],color='green',marker='o',label='RC1')

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC2'],color='purple',marker='o',label='RC2')

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC3'],color='blue',marker='o',label='RC3')

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC4'],color='pink',marker='o',label='RC4')

plt.xlabel('Profundidad / Pies')

plt.ylabel('Porcentaje / %')

plt.title('Registros geofísicos convencionales')

plt.grid(True)

plt.legend()

plt.show()

Gráfico, Gráfico de líneas

Descripción generada automáticamente

# Aplicación del algoritmo

Se importan las bibliotecas que se van a ocupar:

from sklearn import linear\_model

from sklearn.metrics import mean\_squared\_error, max\_error, r2\_score

### Se seleccionan las variables predictoras (X) y la variable a pronosticar (Y)

X\_train = np.array(RGeofisicos[['Profundidad','RC1','RC2','RC3']]) # Separamos las variables predictoras

pd.DataFrame(X\_train)

Texto, Calendario

Descripción generada automáticamenteInterfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Y\_train = np.array(RGeofisicos['RC4']) # Separamos la variable a pronosticar

pd.DataFrame(Y\_train)

Un conjunto de letras blancas en un fondo azul

Descripción generada automáticamente con confianza media

### Se entrena el modelo a través de una regresión lineal múltiple

RLMultiple = linear\_model.LinearRegression()

RLMultiple.fit(X\_train,Y\_train) # Se entrena el modelo

#LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=None, normalize=False)

### Se genera el pronóstico

Y\_pronostico = RLMultiple.predict(X\_train)

pd.DataFrame(Y\_pronostico)

Pantalla azul con letras blancas

Descripción generada automáticamente con confianza mediaPantalla azul con letras blancas

Descripción generada automáticamente con confianza media

### Valores pronosticados

RGeofisicos['Pronostico'] = Y\_pronostico

RGeofisicos

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

# Obtención de los coeficientes, intercepto, error y Score

print('Coeficientes: \n',RLMultiple.coef\_)

print('Intercepto: \n',RLMultiple.intercept\_)

print('Residuo: %.4f' % max\_error(Y\_train,Y\_pronostico))

print('MSE: %.4f' % mean\_squared\_error(Y\_train,Y\_pronostico))

print('RMSE: %.4f' % mean\_squared\_error(Y\_train,Y\_pronostico,squared=False)) #True devuelve MSE y False devuelve RMSE

print('Score (Bondad de ajuste): %.4f' % r2\_score(Y\_train,Y\_pronostico))

Interfaz de usuario gráfica

Descripción generada automáticamente

# Comparación del modelo de pronóstico

* Se tiene un **Score de 0.8581**, el cual indica que el pronóstico de la saturación de aceite remanente (SOR), en un determinado nivel de profundidad, se logrará con un **85.81% de efectividad** (grado de intensidad).
* Además, los pronósticos del modelo final se alejan en promedio 0.0004 y 0.0195 unidades del valor real, esto es, MSE y RMSE, respectivamente.

# Proyección de los valores reales y pronosticados

plt.figure(figsize=(20,5))

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC1'],color='green',marker='o',label='RC1')

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC2'],color='purple',marker='o',label='RC2')

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC3'],color='blue',marker='o',label='RC3')

plt.plot(RGeofisicos['Profundidad'],RGeofisicos['RC4'],color='pink',marker='o',label='RC4')

plt.plot(RGeofisicos['Profundidad'],Y\_pronostico,color='red',marker='o',label='Pronostico')

plt.xlabel('Profundidad / Pies')

plt.ylabel('Porcentaje / %')

plt.title('Registros geofísicos convencionales')

plt.grid(True)

plt.legend()

plt.show()

Gráfico, Gráfico de líneas

Descripción generada automáticamente

plt.figure(figsize=(20,5))

plt.plot(RGeofisicos['Profundidad'],Y\_pronostico,color='red',marker='o',label='Pronostico')

plt.xlabel('Profundidad [Pies]')

plt.ylabel('Porcentaje [%]')

plt.title('Registros geofísicos convencionales')

plt.grid(True)

plt.legend()

plt.show()

Gráfico, Gráfico de líneas

Descripción generada automáticamente

# Nuevos pronósticos

ROS = pd.DataFrame({'Profundidad': [5680.5], 'RC1': [0.45], 'RC2': [0.64], 'RC3': [0.5]})

RLMultiple.predict(ROS)



En este caso, al dar estos nuevos valores y mediciones, se obtiene un Score de 0.4541, es decir, un 45.41% de efectividad.

# Conclusiones

A lo largo de esta práctica, a través de cuatro mediciones de los registros geofísicos convencionales, los cuales son: (RC1) Registro Neutrón, (RC2) Registro Sónico, (RC3) Registro Densidad-Neutrón, y (RC4) Registro Densidad Corregido por Arcilla, se obtuvo el pronóstico de la saturación de aceite remanente (ROS, Residual Oil Saturation), esto gracias a la aplicación del algoritmo de regresión lineal múltiple (ya que se tienen más de dos variables independientes), que pertenece a la categoría de aprendizaje supervisado, el cual su principal objetivo es predecir valores desconocidos o faltantes de una función de valor continuo.

Como se mencionó anteriormente, al aplicar este algoritmo, se obtuvo un **Score de 0.8581**, el cual indica que el pronóstico de la saturación de aceite remanente (SOR), en un determinado nivel de profundidad, se logrará con un **85.81% de efectividad** (grado de intensidad).

Por ende el modelo de pronóstico quedó de la siguiente manera:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

En adición a esto los pronósticos del modelo final se alejan en promedio **0.0004** y 0.0195 unidades del valor real, esto es, **MSE** y RMSE, respectivamente.