2.Realizar EDA detallado, presentando hipótesis y preguntas que se esperan analizar-confirmar

3.Entrenar y validar modelos. Necesario comparar diferentes algoritmos. Validar en elcontexto

4.Desplegar el modelo en un servicio web (AWS, Azure, Google Cloud, etc). Una interfazsencilla (sugerido dash) con la que se pueda utilizar el modelo

5. Realizar pronósticos y análisis de la contribución del modelo(s) en el contexto.

**DESCRIPCIÓN DEL OBJETIVO DEL ESTUDIO, CONTEXTO, RECOLECCIÓN Y EXPLICACIÓN DE VARIABLES**

En el ámbito de los gasoductos, el ROD (rate of drop) se refiere a la velocidad con la que disminuye la presión en un espacio de tiempo determinado, expresado en unidades de psig por minuto. Para salvaguardar los gasoductos, se instalan válvulas automatizadas a lo largo de su infraestructura, y uno de los parámetros clave que se configura en estos dispositivos es el ROD. En esencia, cuando el equipo detecta que este valor se supera, toma automáticamente la decisión de cerrar o seccionar el gasoducto.

El transporte de gas a través de tuberías a alta presión es un proceso caracterizado por cambios lentos en su comportamiento. En condiciones normales de operación, no se anticipan variaciones abruptas en la presión. Generalmente, el ROD puede originarse debido a roturas o fugas, la entrada o salida de equipos compresores, o cambios súbitos en las condiciones de flujo en los puntos de entrada y salida.

Por tanto, es crucial configurar adecuadamente el ROD para garantizar que las protecciones se activen frente a condiciones anómalas de operación, preservando así la seguridad de todas las partes involucradas.

El presente dataset corresponde a valores de ROD máximo y volumen a través de una fuga obtenidos a partir de las combinaciones de las siguientes variables:

* leak\_dia: diámetro en pulgadas de una rotura.
* leak\_distance: distancia de la rotura hasta el punto de medición del ROD.
* pressure: presión de entrada al escenario en psig.
* flow\_pipe: flujo a través de la tubería en Mpcd.

El propósito de este estudio es examinar las relaciones entre las variables previamente descritas y su capacidad para determinar el Rate of Drop (ROD) que se generará. Esta información se utilizará para establecer los valores óptimos de ROD a configurar en las líneas de transporte. El modelo busca responder preguntas específicas, tales como:

* ¿Para una línea operando a una presión y flujo determinado, cuál es el ROD al que se debe configurar la protección para detectar una fuga de un diámetro equivalente en pulgadas? ¿Este ROD configurado corresponde a un flujo que pueda cambiar en condiciones normales de operación, como la entrada o salida de compresores, e incluso cambios en el flujo de las salidas o entradas de gas?
* En caso de una rotura o fuga, ¿a qué ROD equivalente debo configurar la protección si no deseo ventear a la atmósfera una cantidad específica de gas? ¿A qué diámetro equivalente de fuga corresponde este volumen, considerando las condiciones operativas planteadas?

Este estudio pretende proporcionar una comprensión detallada de cómo las variables interrelacionadas afectan el ROD, permitiendo así la determinación precisa de valores óptimos para la configuración de protecciones en las líneas de transporte de gas.

**3.Entrenar y validar modelos. Necesario comparar diferentes algoritmos. Validar en el contexto. 5. Realizar pronósticos y análisis de la contribución del modelo(s) en el contexto.**

Finalmente se construyeron 2 modelos:

1. Regresión Polinomica grado 3: explica la relación entre XXX XXX XXX y como explica el cálculo del ROD.
2. Regresión líneal simple o polinomica: entre las variables rod y leak\_flow, que explica como estas 2 variables presentan correlación.

**Caso de uso 1:**

¿Qué valor de ROD se debe configurar en una protección para detectar una rotura de 3” en una línea de transporte que opera a 900 psig de presión? ¿Qué cambio de volúmenes abruptos puede tolerar la protección sin que actúe?

**Caso de uso 2:**

Si se tiene un cliente que tiene cambios de flujos abruptos de 0 a 70 Mpcd, ¿Qué valor de ROD se debe configurar en una protección si no se desea que esta se cierre bajo esta condición? ¿Suponiendo que esta línea opera a 800 psig, a que diámetro de rotura aproximadamente corresponde este ROD?

**Caso de uso 3:**

La alta gerencia ha decidido que no se toleraran fugas de gas que emitan más de 50 Mpcd de gas a la atmosfera, ¿A qué valor de ROD se debe configurar la protección? ¿A que diámetro de fuga equivalente aproximada corresponde esta directriz para una línea que opera a 700 psig?

**MODELO DE REGRESIÓN PARA ROD Y LEAK\_FLOW**

rod\_Y = rod['rod']

rod\_X = rod['leak\_flow']

X = rod\_X.values.reshape(-1, 1)

Xtrain, Xtest, ytrain, ytest = **train\_test\_split**(X, rod\_Y, *test\_size*=0.2, *random\_state*=123)

model = Pipeline([('poly', PolynomialFeatures(*degree*=3)),    ('linear', LinearRegression(*fit\_intercept*=False))])

model.**fit**(Xtrain, ytrain)

model.**score**(Xtrain, ytrain)

model.**score**(Xtest, ytest)

Con grado 3 da de Score 0.89, una regresión líneal da 0.84