

# SONAR MINES

## *Mines Vs. Rocks.*

### SONAR

---

El sonar (sound navigation and ranging) utiliza ondas sonoras para detectar objetos, de forma similar a como un murciélago utiliza la ecolocalización para navegar y detectar objetos. Es el mismo principio que se aplica a los datos sísmicos para la exploración subterránea no invasiva de formaciones geológicas con el fin de localizar reservas de petróleo o gas.



En este caso, estamos analizando un conjunto de datos que provienen de un estudio de sonar. Imagina que tenemos un dispositivo de sonar que emite ondas sonoras y luego detecta cómo estas ondas rebotan en diferentes objetos bajo el agua. El estudio se llevó a cabo en un laboratorio, donde se recopilaban muestras de dos tipos de objetos: rocas y tuberías metálicas. Se tomaron muestras desde diferentes ángulos y ubicaciones para asegurarnos de obtener una variedad de datos.

Cada muestra que tenemos está representada por un conjunto de 60 números, y cada número nos dice cuánta energía hay en una banda de frecuencia particular en un momento específico. Por ejemplo, podríamos tener más energía en una banda de frecuencia baja si estamos cerca de un objeto grande como una tubería metálica, o más energía en una banda de frecuencia alta si estamos cerca de una roca más compacta.

Ahora, nuestro objetivo es clasificar estas muestras en dos categorías: roca o tubería metálica. Cada muestra tiene una etiqueta asociada, donde la letra "R" indica que es una roca y la letra "M" indica que es una tubería metálica. Sin embargo, para trabajar con estas etiquetas en nuestros modelos de clasificación, necesitamos convertirlas en una representación binaria, donde "0" podría ser roca y "1" podría ser una tubería metálica. Esto se hace utilizando lo que se llama codificación one-hot.

## PROBLEMA

---

El problema que queremos solucionar con estos modelos sería, que nuestro modelo seleccionado, sea capaz de diferenciar según las métricas si un objeto, rastreado con el sonar, es una mina o una roca.

## SOLUCIÓN

---

Probaré varios modelos favoritos y encontraré el más adecuado:

1. Preparación de los Datos:
  - a. Limpieza de datos, manipulación de valores faltantes (si los hay), codificación de variables categóricas y división de los datos en conjuntos de train y test.
2. Elección de los Modelos:
  - a. Random Forest Classifier.
  - b. Logistic Regression.
  - c. Gradient Boosting.
  - d. KNN.
  - e. Decision Tree.
  - f. Naive Bayes.
  - g. Neural Networks.

3. Entrenamiento y Evaluación de los modelos.
4. Ajuste de los hiperparámetros.
5. Validación cruzada.
6. Selección del modelo final.