# LANL Earthquake Prediction

## **Can you predict upcoming laboratory earthquakes?**

## https://www.kaggle.com/c/LANL-Earthquake-Prediction

Problemas en el abordaje actual:

1. Los modelos tienden a capturar la periodicidad del conjunto de entrenamiento y esto perjudica las predicciones que se hacen sobre conjuntos aleatorios de secciones de la serie

Preguntas:

1. La cantidad de tiempo que se debe esperar para que se produzca una falla depende de la historia del sistema? I.e. la cantidad de fallas que se han producido antes, la cantidad de tiempo que tardó en generarse cada falla, y la magnitud de la señal acústica en las fallas anteriores influye en la cantidad de tiempo que falta para la siguiente falla
2. Es posible usar la transformada de fourier para analizar la señal acústica y extraer de esta forma información relevante, como lo puede ser la una frecuencia específica producida por la formación de clusters de arena que luego generarán una falla, o usarla para filtrar características de los datos y extraer aquellas que sean relevantes, como puede ser el caso de extraer la periodicidad de los datos.
3. Las fallas en el experimento son deslizamientos de la placa central, cuando la arena no se encuentra compactifica o no hay muchos clusters de arena compactificada la placa de mueve de forma constante, al compactificarse más allá de cierto límite no permite el movimiento de la placa central y entonces en cuando se producen los desplazamientos; los desplazamientos rompen los clusters de arena y la devuelven a un estado desordenado que permite que la placa se desplace a una velocidad constante. Teniendo esto en mente, es posible aplicar mecánica estadística para predecir características del sonido emitido por la formación o rompimiento de clusters de cierto tamaño, y quizás usar percolación para estudiar la formación del cluster crítico que precede al deslizamiento.

Propuestas:

1. Las ventana de la serie de tiempo usadas para extraer las características tienen información de posiblemente varios deslizamientos, luego, buscar una forma de predecir la cantidad de deslizamientos en una ventana de tiempo dada puede ser útil para predecir la cantidad de tiempo faltante para la siguiente falla. Adicionalmente entrenar un predictor que solamente use la información de una ventana de tiempo donde no se hayan producido fallas puede ser útil para generar clasificadores débiles que luego se combinarán para generar un predictor fuerte para la cantidad de tiempo faltante para la siguiente falla.

Kernels:

La siguiente es una lista de kernels interesantes que presentan información útil sobre la forma estándar de analizar la serie de tiempo:

1. <https://www.kaggle.com/artgor/earthquakes-fe-more-features-and-samples>
2. <https://www.kaggle.com/andrekos/basic-feature-benchmark-with-quantiles>
3. <https://www.kaggle.com/jsaguiar/baseline-with-multiple-models>
4. <https://www.kaggle.com/wimwim/rolling-quantiles>