**LANL Earthquake Prediction**

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

**DataSet**

1. Dataset details, such as number of features, instances, data distribution

**Features:**

**Training Data:**

* acoustic\_data - the seismic signal [int16]
* time\_to\_failure - the time (in milli seconds) until the next laboratory earthquake [float64]

Training Data instances: 629 million points

**Data Distribution:**

| **signal** | **quaketime** |
| --- | --- |
| count | 1.000000e+07 |
| mean | 4.502072e+00 |
| std | 1.780707e+01 |
| min | -4.621000e+03 |
| 25% | 2.000000e+00 |
| 50% | 4.000000e+00 |
| 75% | 7.000000e+00 |
| max | 3.252000e+03 |

**Test Data:**

* seg\_id- the test segment ids for which predictions should be made (one prediction per segment)
* acoustic\_data - the seismic signal [int16] for which the prediction is made.

Test Data instances: 2624 files, with 150,000 instances for each file => 393,600,000 instances

**Techniques we plan to use:**

* SVM
* Gradient Boosting
* Random Forests

**Experimental methodology:**

1. Divide the training data into chunks of 150,000 data points as the test data consists of 150,000 points
2. We are not creating validation dataset as the input dataset is a continguous data from a sensor. Creating validation dataset by choosing the data randomly will not give any good results
3. Scale the data

**Feature Engineering:**

Feature generation: Create several groups of features:

1. Usual aggregations: mean, std, min and max
2. Average difference between the consequitive values in absolute and percent values;
3. Absolute min and max vallues;
4. Aforementioned aggregations for first and last 10000 and 50000 values - I think these data should be useful;
5. Max value to min value and their differencem also count of values bigger that 500 (arbitrary threshold);
6. Quantile features
7. Trend features
8. Rolling features

**Coding Language:**

* Python

**Team-Members**

* Chandra Kiran Saladi ( cxs172130 )
* Shreyash Mane ( ssm170730 )
* Tanya Tukade ( txt171230 )
* Supraja Ponnur ( sxp179130 )