Machine Learning Predicts Aperiodic Laboratory Earthquakes

Olha Tanyuk, Daniel Davieau, Charles South and Daniel W. Engels

Southern Methodist University, Dallas TX 75205, USA otanyuk@mail.smu.edu, danieldavieau@mail.smu.edu, csouth@mail.smu.edu, dwe@lyle.smu.edu

Abstract. Our goal is to find the pattern of aperiodic seismic signals that precede earthquakes at any time in the earthquake's cycle using a small window of time. We use data collected from a laboratory experiment which exhibits similar behavior to natural earthquakes. We apply machine learning to a data set that comes from a classic laboratory experiment having several stick-slip displacements (earthquakes), a type of experiment which has been studied in depth as a simulation of seismologic faults for decades. We show that by listening to the acoustic signal emitted by the laboratory experiment a machine learning algorithm can predict the time remaining before it fails. These predictions are based only on the the acoustic signal and not its history. Los Alamos' initial work showed that the prediction of laboratory earthquakes from continuous seismic data is possible in the case of quasi-periodic laboratory seismic cycles.

1 Introduction

Earthquakes cause mass destruction and loss of life. A traditional method to predict earthquakes is to look to past recurrence intervals. Because the recurrences are not constant, predictions can only be made within broad time windows. One such model predicted that a strong earthquake would occur between 1985 and 1993 in the Parkfield California area but no significant event actually occurred until 2004 [1].

Advances in instrumentation quality and density have fostered hope that progress can be made in forecasting. These advances have led to exciting discoveries of previously unidentified slip processes such as slow slips. Slow Slip Earthquakes (SSE) are fault behaviors that occur slowly enough to make them undetectable without instrumentation. They do not cause immediate widespread destruction like regular earthquakes do. They occur near the boundaries of large earthquake rupture zones [2]. There is evidence to suggest that there is a relationship between slow slip earthquakes and more noticeable regular earthquakes [3].

Researchers imitate natural slow slip earthquakes in the laboratory by placing rocky material between steel blocks and applying shear stress to induce slipping. Recent improvements in the instruments [4] used to measure signals have enabled the collection of larger volume data from more realistic and unpredictable laboratory earthquakes. However, processing the data and detecting patterns in it has become more difficult to work with. In this paper we demonstrate that machine learning can be used to detect patterns in the more realistic data and predict laboratory earthquakes.

In this paper, given seismic signal data with considerably more aperiodic slow-slip laboratory earthquake failures, we find the pattern of acoustic signals to predict the time at which laboratory earthquakes will occur at any time in the slip cycle.

We use acoustic data provided by the Los Alamos National Laboratory (LANL) as part of a 2019 Kaggle competition [4], which also represents laboratory slow-slip earthquakes. The data is very aperiodic and more realistic than the data LANL studied earlier in 2017. [4]. The results of this experiment are potentially applicable to the field of real world earthquakes [5]. Other potential applications include avalanche prediction or failure of machine parts [5].

2 Background

2.1 Statistical Values to Evaluate Predictions

R2	Proportion of the variance in the dependent variable that is predictable from the independent variable	$R^2 \equiv 1 - \frac{SS_{\rm res}}{SS_{\rm tot}}$ where SSres is the sum of the squares of the residuals and SStot is the sum of the squared differences of each observation from the overall mean
MAE	It measures the average magnitude of the errors in a set of predictions, without considering their direction.	$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} \lvert y_j - \hat{y}_j \rvert$ where yj is an actual value, ŷj is a predicted value, n is the total number of observations

Fig. 1. The values used to evaluate predictions in our work are the coefficient of determination r^2 and mean absolute error (MAE).

2.2 Los Alamos National Laboratory's Findings

In 2017 Los Alamos National Laboratory (LANL) researchers discovered a way to successfully predict Slow Slip Earthquakes (SSE) in a laboratory experiment that simulates natural conditions. The team trained a computer to pinpoint and analyze quasi-periodic seismic and acoustic signals emitted during the movements along the fault. They processed massive amounts of data and identified a particular sound pattern previously thought to be noise that precedes an earthquake. The team was able to characterize the time remaining before a laboratory earthquake at all times using time window of 1.8 sec of the data to make each prediction with 89% coefficient of determination [6]. This result was achieved using a Random Forest Regression machine learning technique and quasi-periodic data.

In the lab, the team imitated a real earthquake using steel blocks interacting with rocky material to cause slipping that emitted seismic sounds. An accelerometer recorded the acoustic emission emanating from the sheared layers [6]. For the first time, researchers discovered a pattern that accurately predicted when a laboratory earthquake would occur. The LANL team acknowledges that the characteristics of the lab experiment such as shear stress differ from natural earthquakes but the application of the analysis to the real world to validate their results is ongoing. This method can also be applied outside of seismology to support material failure research in other fields such as aerospace and energy [6]. The lab results reveal that the fault does not fail randomly but in a predictable manner. The observations also demonstrate that the fault's critical stress state which indicates when it might slip can be determined using exclusively an equation of state [6]. So far seismologists and earth scientists have mostly relied on catalogs of historical data to try to characterize the state of faults. These catalogs contain a minute fraction of seismic data with portions discarded during analysis as useless noise. The authors discovered that hidden in the noise-like data are signals that inform them of the state of the fault much more precisely than catalogs [6].

2.3 Experimental Setup

The laboratory system is a two-fault configuration that contains fault gouge material submitted to double direct shear [4].

Two fault gouge layers are sheared simultaneously while subjected to a constant normal load and a prescribed shear velocity [4]. The laboratory faults fail in repetitive cycles of stick and slip that is meant to mimic the cycle of loading and failure on tectonic faults [4]. While the experiment is considerably simpler than a fault in the Earth it shares many physical characteristics [4].

A driving piston displaces at a very constant velocity during the inter-event time and accelerates briefly when a slip occurs [5]. An accelerometer records the acoustic emission emanating from the shearing layers [5]. The steel blocks are extremely stiff therefore the deformation takes place largely in the gouge [5]. Under a broad range of load and shear velocity conditions, the apparatus stick-slips quasi-periodically for hundreds of stress cycles during a single experiment and in general follows predictions from rate and state friction [5]. The rate of impulsive precursors accelerates as failure approaches suggesting that upcoming laboratory earthquake timing could be predicted [5].

The experimental data has 16 earthquakes. The shortest time to failure is 1.5 seconds for the first earthquake, 7 seconds for the 7th and the longest is around 16 seconds.

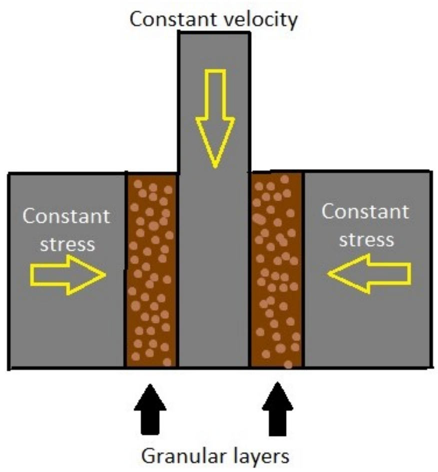

Fig. 2.

2.4 Extra Trees Regressor Overview

The Extra Trees algorithm builds an ensemble of un-pruned regression trees according to the classical top down procedure. Its two main differences from other tree-based ensemble methods are that it splits nodes by choosing cut points fully at random and that it uses the whole learning sample (rather than a bootstrap replica) to grow the trees [7].

The Extra-Trees splitting procedure uses the number of attributes randomly selected at each node (k) and the minimum sample size for splitting a node ($nmin$). It is used several times with the full original learning sample to generate an ensemble model. We denote the number of trees of the ensemble with (M) [7]. The predictions of the trees are aggregated to yield the final prediction by a majority vote in classification and an arithmetic average in regression problems [7]. From the bias-variance point of view, the rationale behind the Extra-Trees method is that the explicit randomization of the cut-point and attribute combined with ensemble averaging should be able to reduce variance more strongly than the weaker randomization schemes used by other methods [7]. The usage of the full original learning sample rather than bootstrap replicas is motivated in order to minimize bias [7]. From a computational point of view, given the simplicity of the node splitting procedure we expect the constant factor to be much smaller than in other ensemble based methods which locally optimize cut-points [7]. The parameters k, $nmin$ and M have different effects: k determines the strength of the attribute selection process, $nmin$ the strength of averaging output noise, and M the strength of the variance reduction of the ensemble model aggregation. These parameters could be adapted to the problem specifics in a manual or an automatic way (e.g. by cross-validation) [7].

The sklearn.ensemble module in the Python programming language includes an averaging algorithm based on randomized decision trees known as the Extra-Trees method [8].

3 Data

The data used in this work is a 157.275 second recording of seismic signals (ordered in sequentially.) It was recorded at 4MHz hence 629,145,480 data points accompanied by the time remaining in seconds until the following lab earthquake.

The seismic signals are recorded using a piezoceramic sensor which outputs a voltage upon deformation by incoming seismic waves (henceforth we will use the term acoustic signal). The seismic data, which serves as the input to our analysis, is this recorded voltage, in integers.

Acoustic signal is voltage upon deformation by incoming seismic waves.

Time to failure is the remaining time in seconds until an actual stick-slip failure occurred.

Table 1.

Acoustic Signal	Time to Failure
12	1.469099998474121
6	1.469099998474121
8	1.469099998474121
5	1.469099998474121
8	1.469099998474121

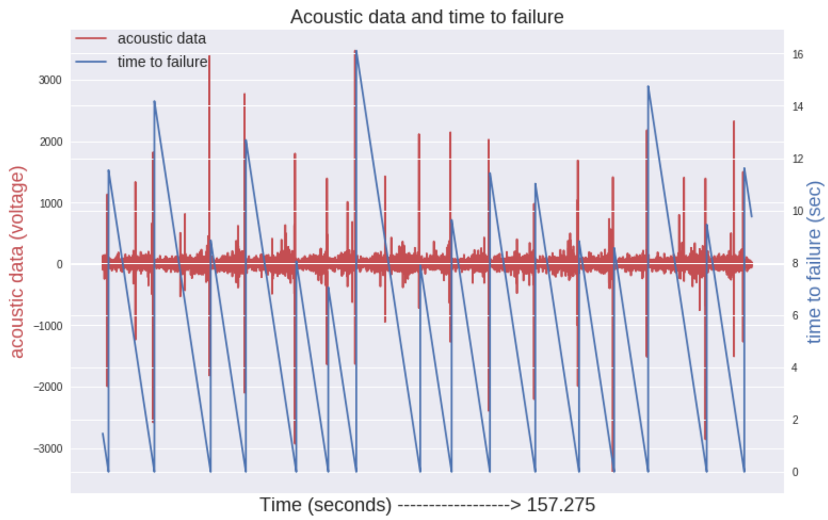

Fig. 3. We can see that the acoustic signal shows huge fluctuations regularly just before the failure. It is also worth noting that failures can be predicted visually as cases when huge fluctuations in the signal are followed by smaller signals.

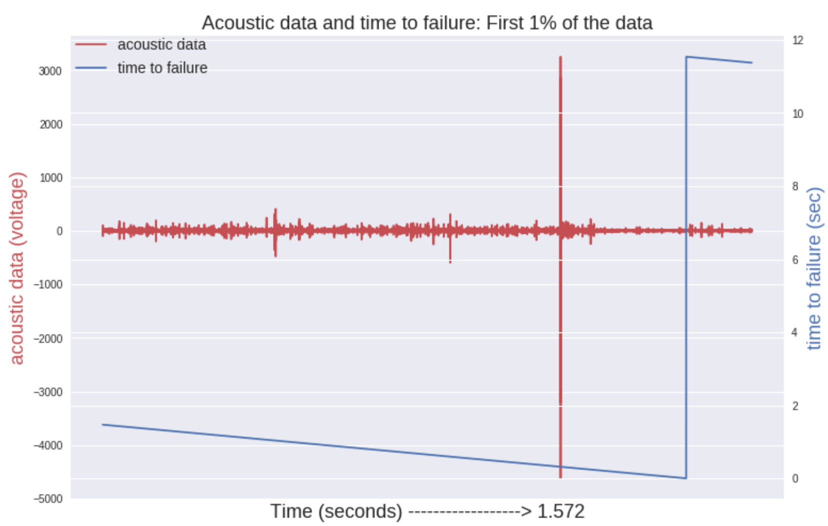

Fig. 4. On this zoomed-in time plot we can see that the large acoustic signal oscillation at the 1.572 second mark is not at the exact time of the failure but just before it. There are trains of intense signal oscillations preceding the large one and some smaller ones after it.

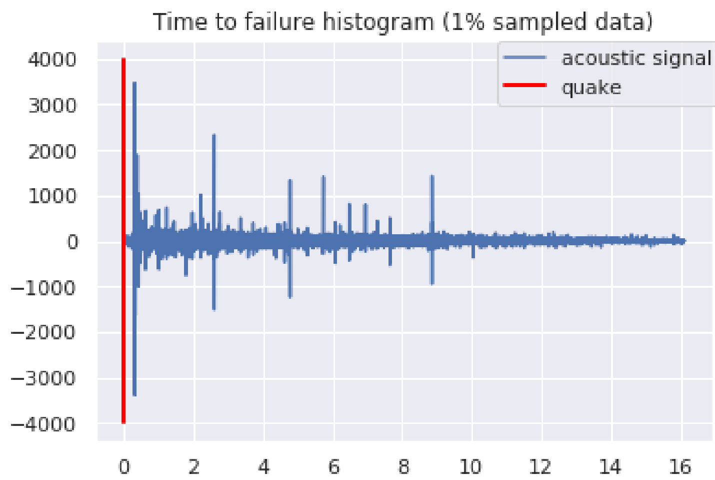

Fig. 5. In this 1% sample of the data we can see that the voltage amplitude of acoustic precursors accelerates as failure approaches, suggesting that upcoming laboratory earthquake timing could be predicted. The red line indicates that a quake occurs when the time to failure approaches 0. The minimum time remaining until the quake is -5.5150e+03 sec.

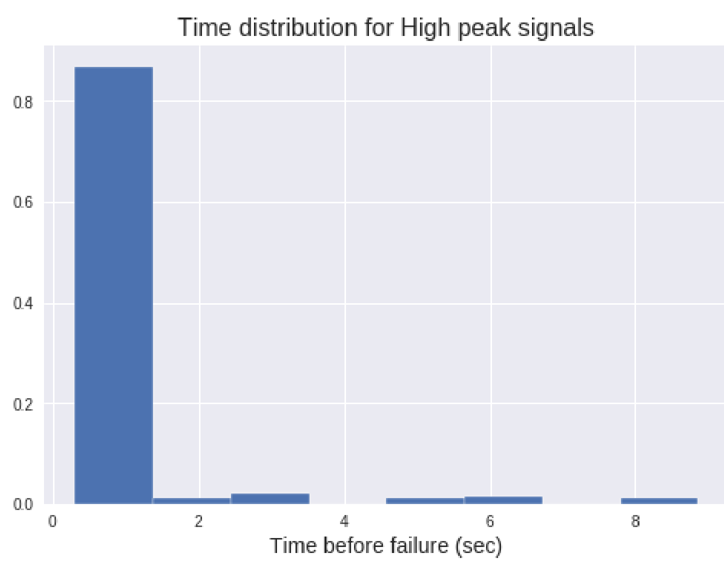

Fig. 6. We found that more than 90% of high acoustic signal values (absolute value greater than 1000) are around 0.31 seconds before an earthquake!

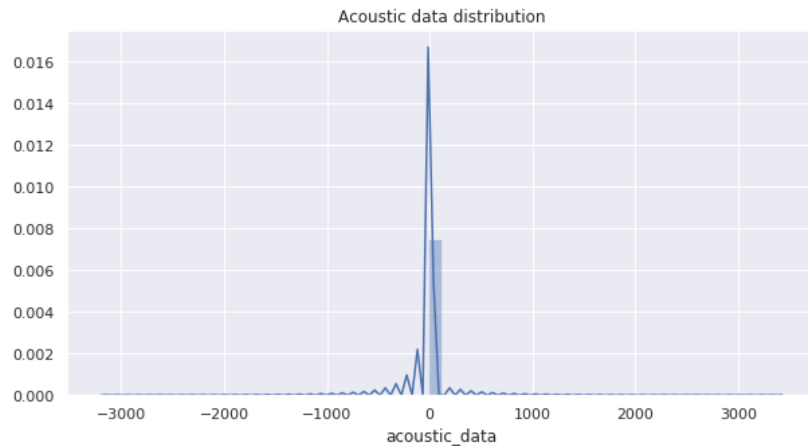

Fig. 7. The distribution of the acoustic signals have a very high peak and we see outliers in both directions

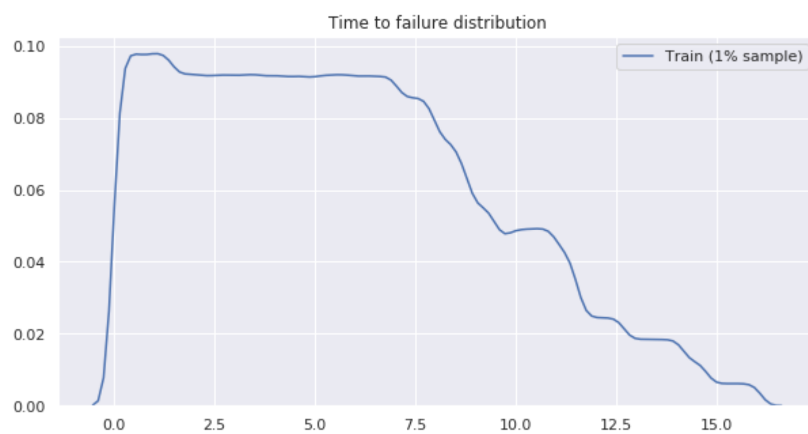

Fig. 8. The distribution of the time to failure seems right skewed. It should dieally be normally distributed.

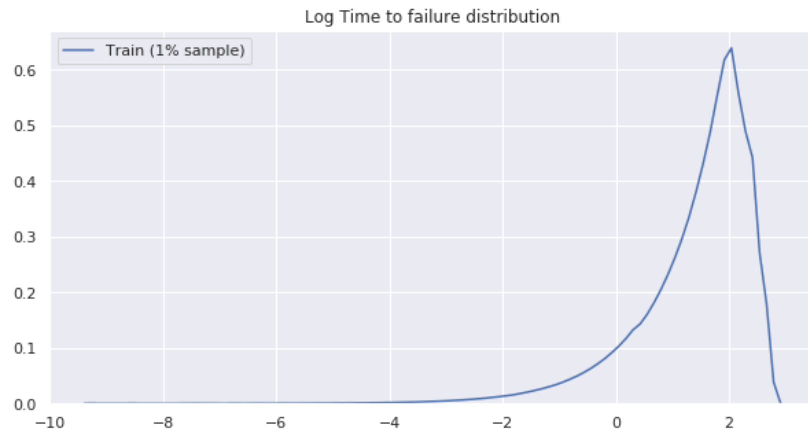

Fig. 9. In this plot we can see that applying a logarithmic transform to the time to failure results in a left skewed distribution.

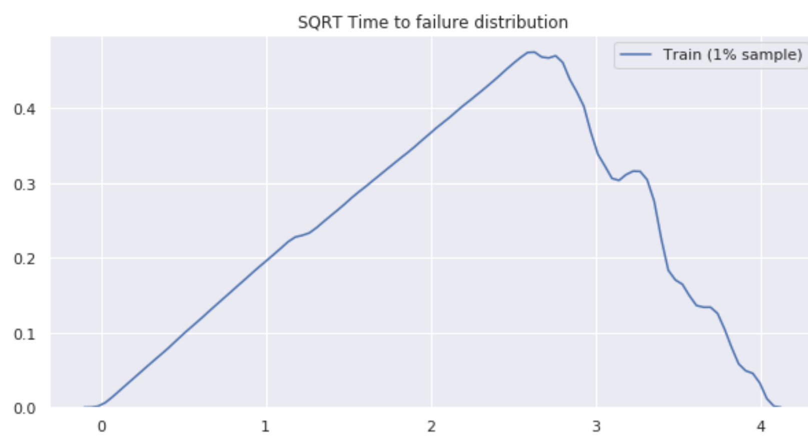

Fig. 10. In this plot we can see that applying a square root transformation to the time to failure is still not normal but improved the distribution significantly.

4 Feature Engineering

Working with quasi periodic seismic signals LANL achieved a 0.89 coefficient of determination. They divided the data into 1.8 second time windows and used a Random Forest technique. [5]. The most important features in the LANL model were variance, kurtosis and threshold.

We used a similar approach in this study. Our goal is to predict the time remaining before the next failure using only moving time windows of the acoustic data. We divided the data into 0.3 second time windows (1,500,000 observations) which is small enough relative to the lab quake cycle which spans 8 to 16 seconds. As indicated in figure 6 more than 90% of high acoustic values (absolute value greater than 1000) are around 0.31 seconds before an earthquake. It makes sense to divide our data by 0.3 sec windows to reduce error at the end of the quake cycle.

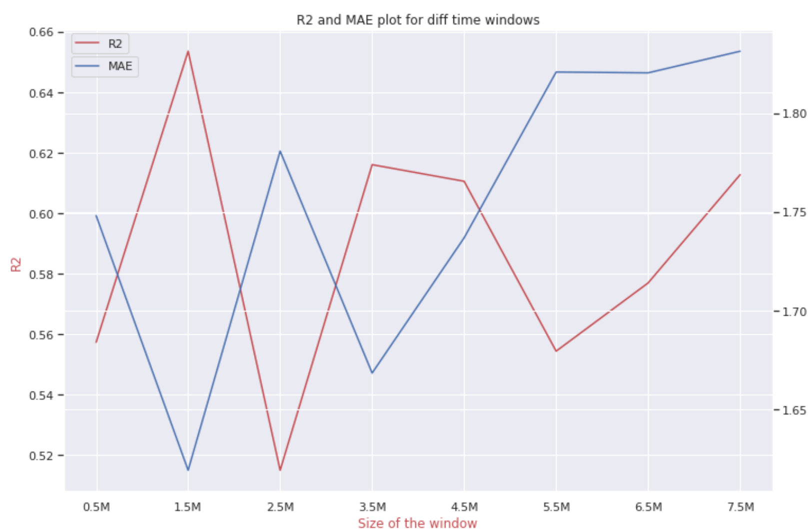

Fig. 11. Checking how sensitive our results are to the size of the time window we find that the highest r^2 and smallest mean absolute error we were able to achieve is with 1.5M observations in each time window.

Our resulting transformed data set consists of 419 time windows (0.3 seconds each). From each time window we compute a set of 95 potentially relevant statistical features (e.g., mean, variance, kurtosis, min/max, threshold and so on). Using feature importance techniques we found that only the following (in Table 3) are considerably important:

Table 2. List of Engineered Features

Standard Deviation
90% Quantile
95% Quantile
99% Quantile
Absolute Standard Deviation
Average Rolling Standard Deviation for 100, 1000 and 10000 observations
Variance of Rolling Standard Deviation for 100, 1000 and 10000 observations
Minimum Rolling Standard Deviation for 100, 1000 and 10000 observations
1% Quantile of rolling standard deviation for 100, 1000, 10000 observations
5% Quantile of rolling standard deviation for 100, 1000, 10000 observations
10% Quantile of rolling standard deviation for 100, 1000, 10000 observations
90% Quantile of rolling standard deviation for 100, 1000, 10000 observations
95% Quantile of rolling standard deviation for 100, 1000, 10000 observations
99% Quantile of rolling standard deviation for 100, 1000, 10000 observations
Variance of Rolling Absolute Mean for 100, 1000 and 10000 observations

We apply different machine learning techniques such as the Random Forest Regressor, XGB Regressor, Decision Tree Regressor, LGBM Regressor and Extra Trees Regressor to the new continuous values that we created analyzing acoustic time series data.

To avoid correlation between the new features we applied a principal component analysis technique. The principle components allows us to reduce the number of features from 35 to only 5 which represents 99.9% of the full data variation.

We use a 50/50 continuous split of the data for use as training and testing data sets respectively. Contiguity of the train and test data sets is important to minimize contamination of the training data with information about the test data.

We selected regularization hyper-parameters for each machine learning algorithm using random grid search technique based on a 3-fold cross-validation.

5 Results

We run different techniques on a training data set (50% of the full data) before principal component analysis and after. Principal component analysis did not allow for any significant improvement in our results. We apply our model to generate predictions on the test data and measure the accuracy of them using r^2 (the coefficient of determination) and MAE (mean absolute error).

The most accurate results were achieved by the Extra Trees Regressor algorithm with 1,61 seconds MAE and 0.65 r^2. The parameters used are displayed in table 3. The hyper-parameters used with the algorithm were:

Table 3. Extra Trees Regressor Parameters

Parameter	**Setting**
Maximum Depth	25
Maximum Features	Auto
Minimum Samples Leaf	16
Minimum Samples Split	30
Minimum Weight Fraction Leaf	0.0623
Number of Estimators	400

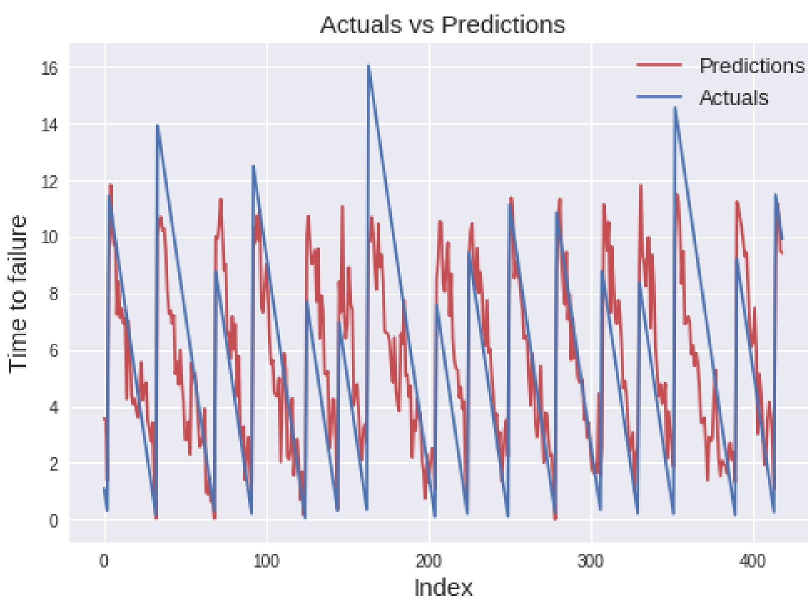

Fig. 12. We emphasize that there is no past or future information considered in calculating the predictions (red line). Each prediction uses only the acoustic signal information within one single time window.

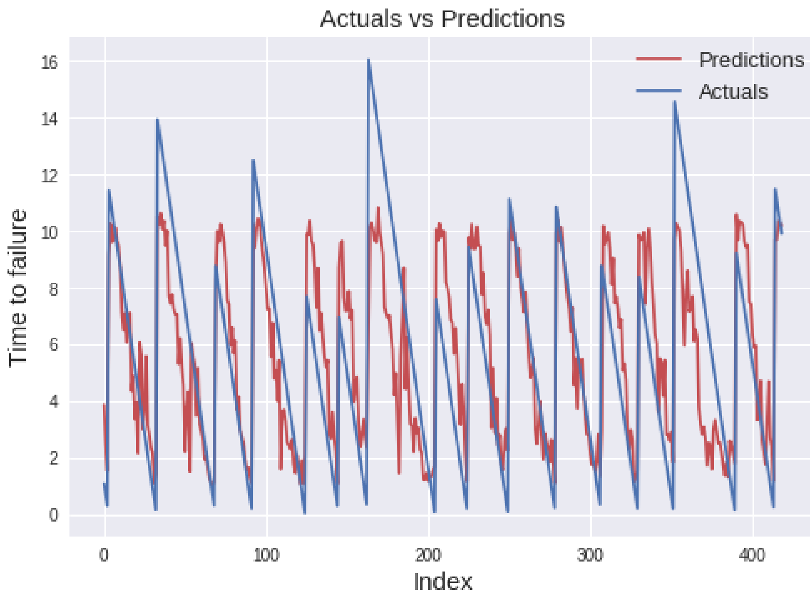

Fig. 13. Results achieved with the Ada Boost Regressor algorithm are 1.67 MAE and 0.62 r^2 score. The hyper parameters used are learning rate = 0.01826, loss = square, number of estimators = 500 and base estimator= Ridge(alpha=1).

6 Analysis

The most accurate results with coefficient of determination 0.65 and mean absolute error 1.61 seconds we achieved using the Extra Trees Regressor. The most important features are shown on figure 14. The top 5 are 90% and 95% quartile rolling standard deviations, standard deviation of rolling absolute mean and average rolling standard deviation.

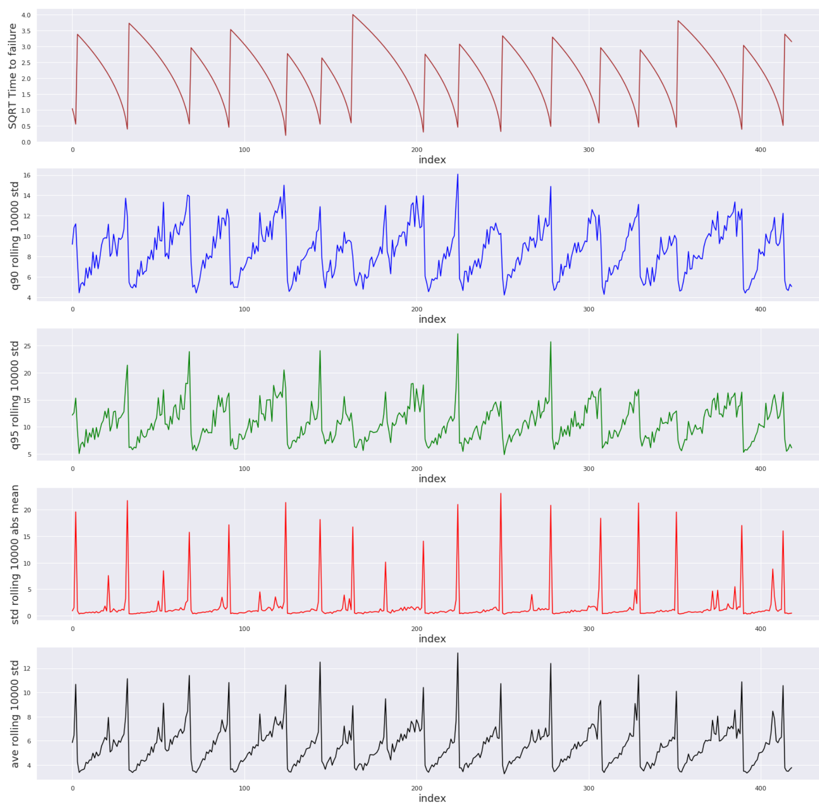

Fig. 14.

7 Ethics

Our responsibility to report our results has to be weighed with our responsibility not to cause social disturbance. If the method demonstrated in this paper is scaled and applied to predict natural earthquakes this balance must be considered. Unwarranted predictions could have affects on personal property value while failure to report warranted predictions could result in loss of life or avoidable property damage [9].

8 Conclusions

Given the more realistic data Machine Learning can provide failure forecasts based on small windows of time. The acoustic signal measurement is an indicator of imminent failure. The prior recurrence interval is not needed to make a prediction. The Extra Trees Regressor machine learning technique provides the most accurate results.

In this study we are using only 157 seconds of data. Future work should introduce higher volumes of data to determine if accuracy can be improved. Also the combination of the acoustic signal and the recurrence interval should be tested to determine it's affect on accuracy.

9 Acknowledgment

This research was supported by Dr. Michael L. Blanpied, U.S. Geological Survey. We thank Dr. Michael L. Blanpied for comments that greatly improved our work.

References

1. Jackson, D.D.: The 2004 parkfield earthquake, the 1985 prediction, and characteristic earthquakes: Lessons for the future. Bulletin of the Seismological Society of America **96**(4B) (2006) 397–409 https://doi.org/10.1785/0120050821.
2. Ikari, M.J.: Slip weakening as a mechanism for slow earthquakes. Nature Geoscience **6** (May 2013) 468 https://doi.org/10.1038/ngeo1818.
3. Rousset, B., Bürgmann, R., Campillo, M.: Slow slip events in the roots of the san andreas fault. Science Advances **5**(2) (2019)
4. Kaggle: Lanl earthquake prediction. https://www.kaggle.com/lanl (March 2019)
5. Rouet-Leduc, B.: Machine learning predicts laboratory earthquakes. Geophysical Research Letters **44**(18) (August 2017)
6. Rouet-Leduc, B.: Los Alamos machine learning discovers patterns that reveal earthquake fault behavior. Los Alamos National Laboratory, P.O. Box 1663 Los Alamos, NM 87545. Science briefs 2018 edn. (March 2018)
7. Geurts, P.: Extremely randomized trees. Springer Science + Business Media, Inc, (March 2006)
8. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
9. Ayhan Sol, H.T.: The ethics of earthquake prediction. Science and Engineering Ethics **10**(4) (December 2004) 655–666