

Flatiron Capstone Project: LANL Earthquake Detection

David Braslow

May 17, 2019

Overview



The screenshot shows the top banner of the Kaggle competition. It features a dark blue background with a yellow seismic waveform. The text 'Research Prediction Competition' is in the top left. The main title 'LANL Earthquake Prediction' is in large white font, followed by the question 'Can you predict upcoming laboratory earthquakes?'. The prize money '\$50,000' is on the right. The Los Alamos National Laboratory logo and text are on the left. A progress bar at the bottom shows '17 days to go (10 days to go until merger deadline)'. Navigation links include 'Overview', 'Data', 'Kernels', 'Discussion', 'Leaderboard', 'Rules', 'Team', 'My Submissions', and a blue 'Submit Predictions' button.

Research Prediction Competition

LANL Earthquake Prediction

Can you predict upcoming laboratory earthquakes?

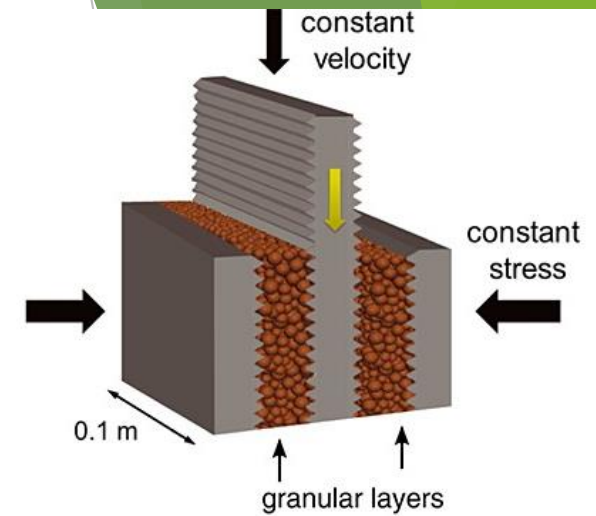
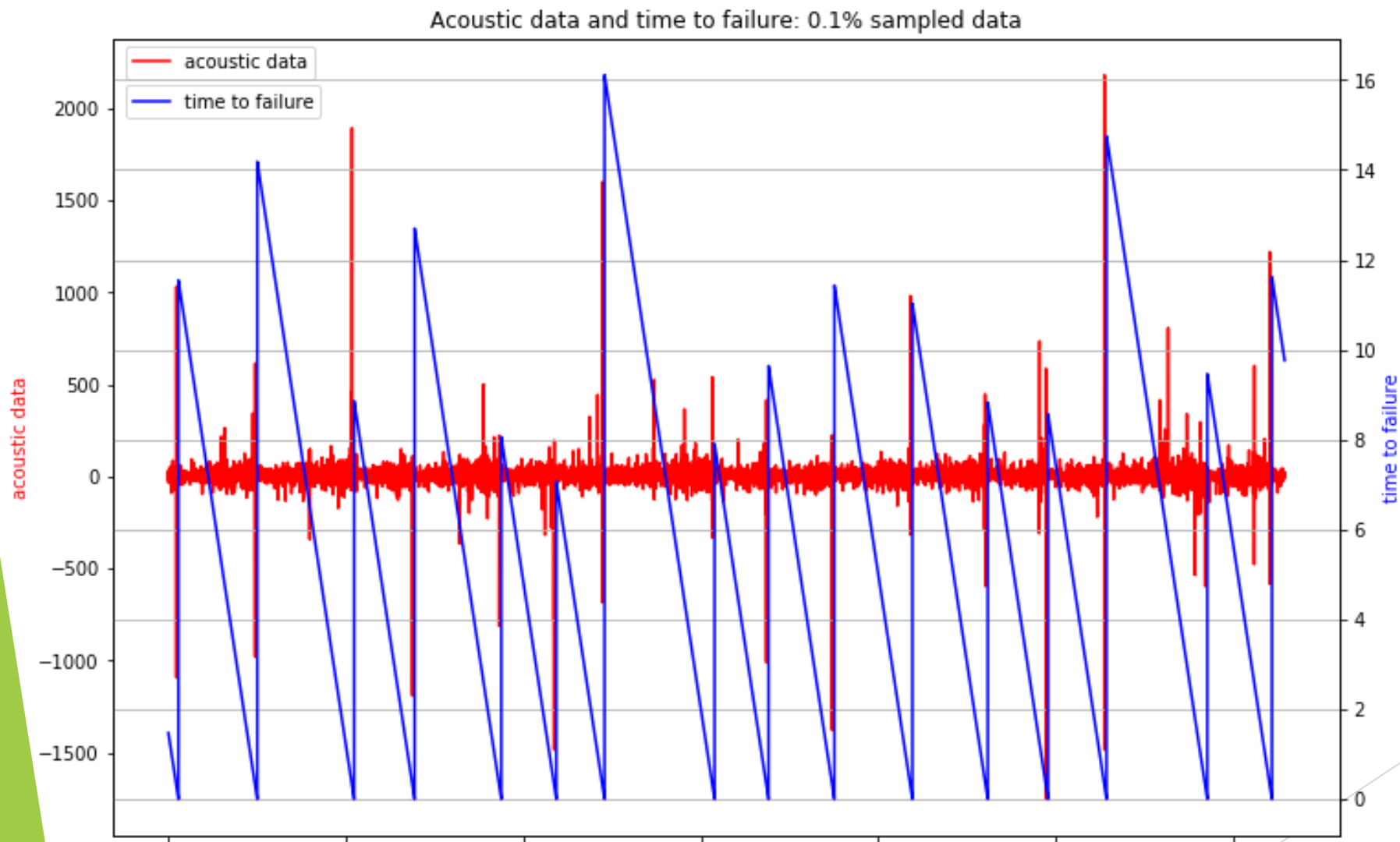
\$50,000
Prize Money

Los Alamos National Laboratory · 3,983 teams · 17 days to go (10 days to go until merger deadline)

[Overview](#) [Data](#) [Kernels](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) [Submit Predictions](#)

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

Predicting “Time to Failure”



Data Processing

Extract and augment acoustic snippets from original acoustic file



```
graph TD; A[Extract and augment acoustic snippets from original acoustic file] --> B[Generate over 2000 features for each training and test sample]; B --> C[Select features that are useful for predicting "time to failure"];
```

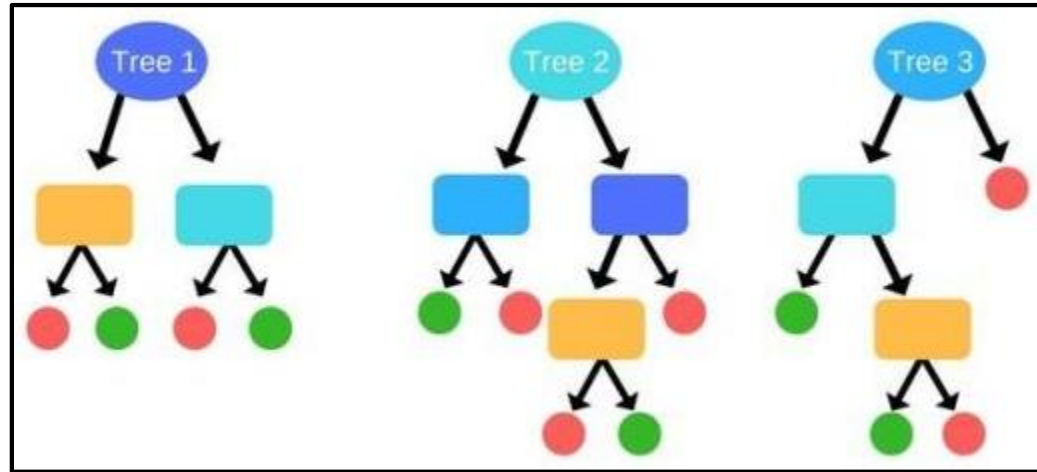
The diagram illustrates a three-step data processing workflow. It begins with a red box at the top, followed by a brown box in the middle, and a dark olive box at the bottom. Each step is connected to the next by a downward-pointing arrow, indicating a sequential process. The background features abstract green and yellow geometric shapes on the right side.

Generate over 2000 features for each training and test sample

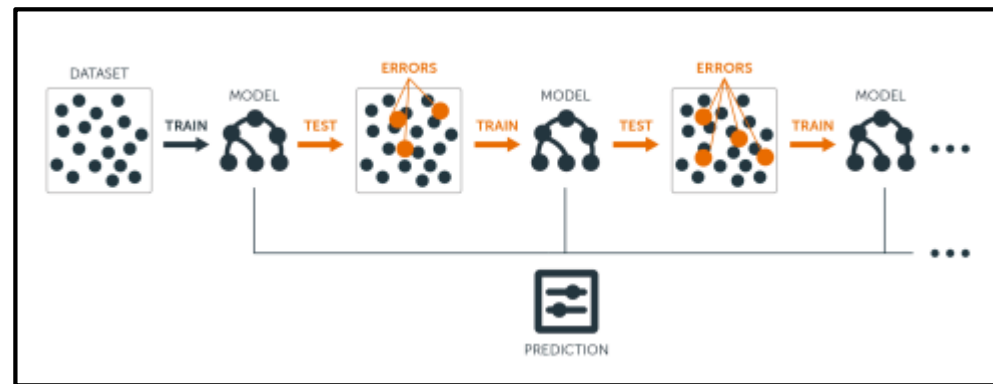
Select features that are useful for predicting “time to failure”

Modeling Approaches

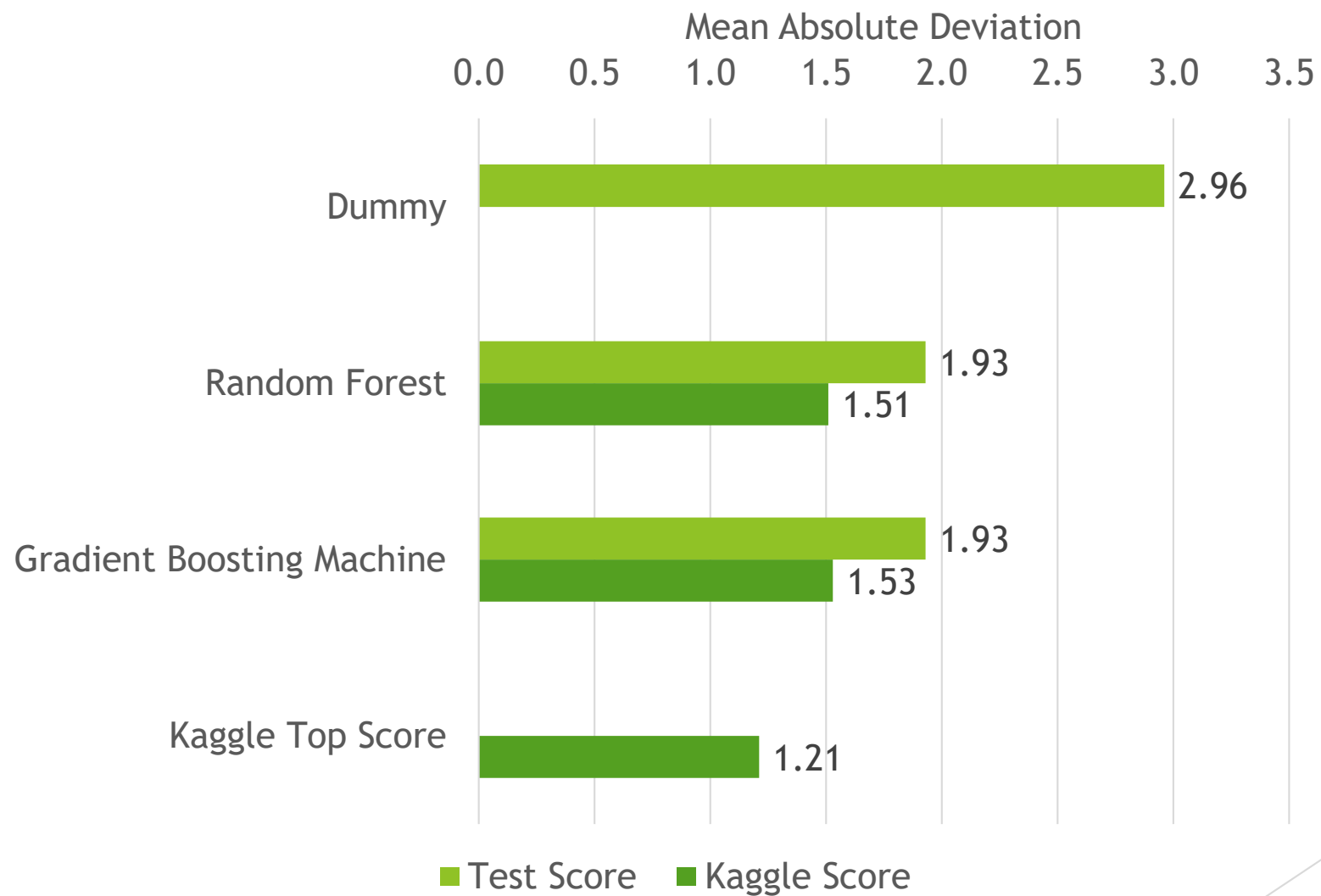
► Random Forest



► Gradient-Boosting Machine



Results



Conclusion

I can predict the timing of aperiodic earthquakes moderately well using tree-based approaches

This can be useful for advancing seismology research and for improving public earthquake warning systems

Model improvement may be possible using continuous acoustic data, rather than relying on acoustic snippets

Additional computing resources could yield further improvements with more feature engineering and model testing