Flatiron Capstone Project: LANL Earthquake Detection

David Braslow

May 17, 2019

Overview

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

Predicting "Time to Failure"

Acoustic data and time to failure: 0.1% sampled data
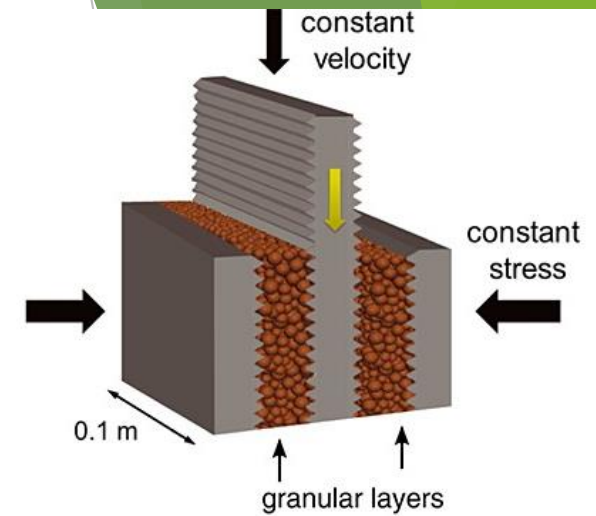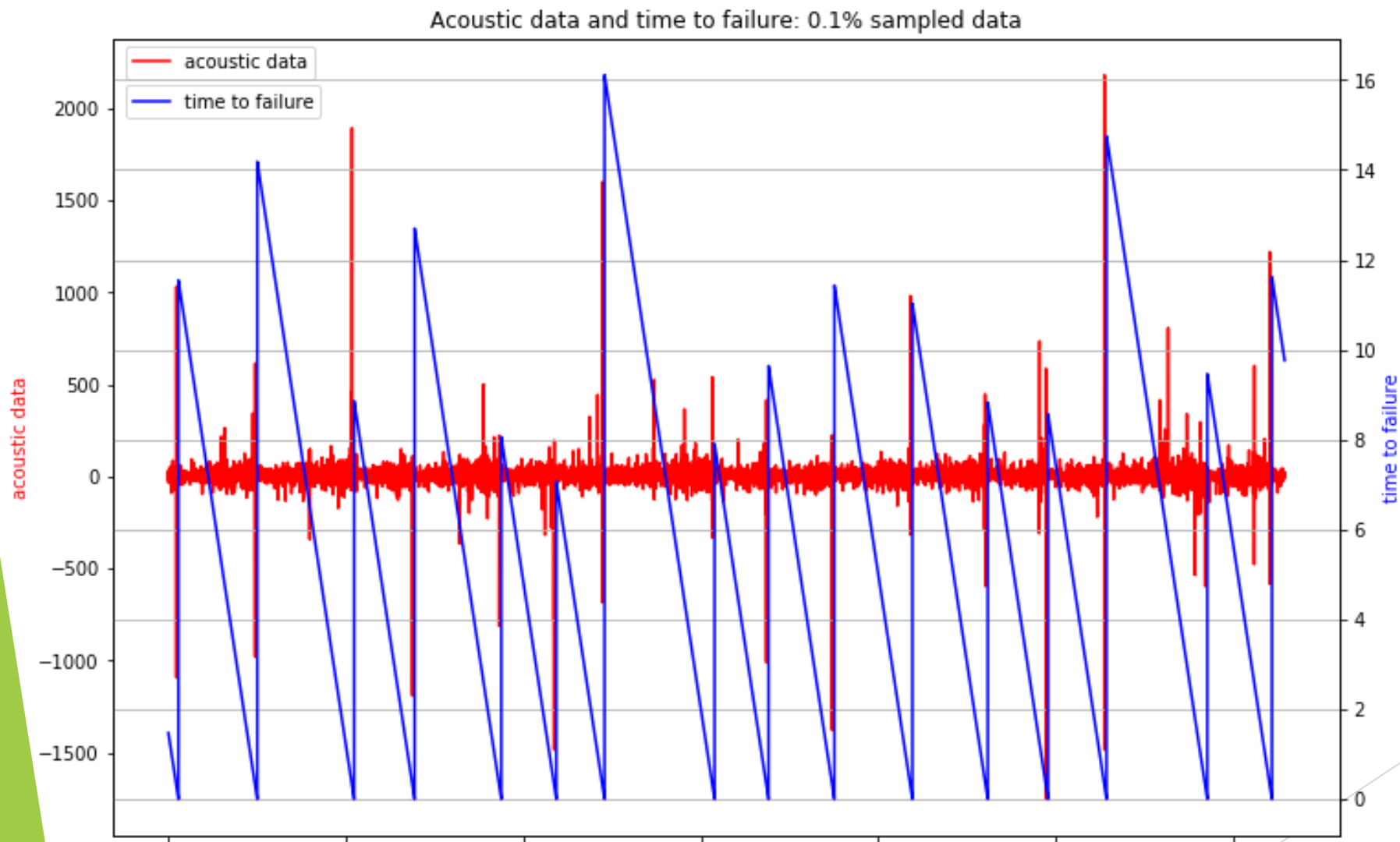
Data Processing

Extract and augment acoustic snippets from original acoustic file

Generate over 2000 features for each training and test sample

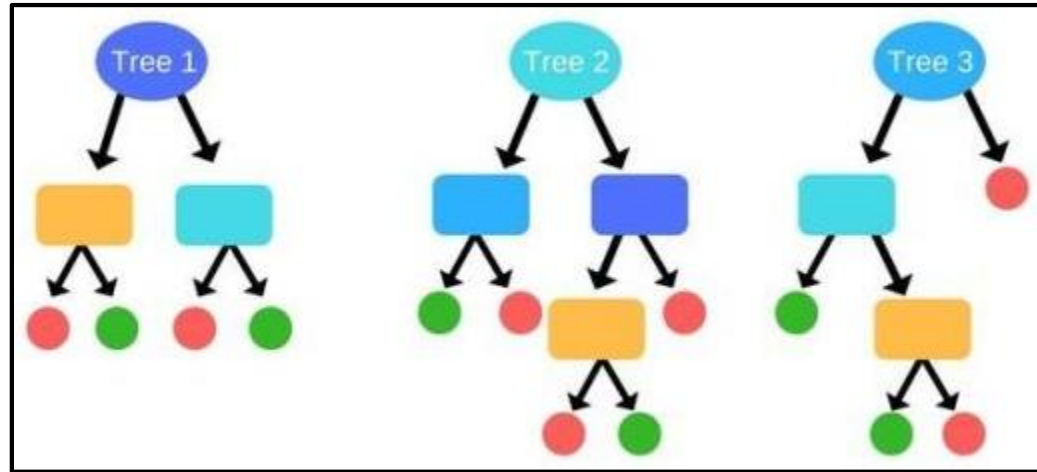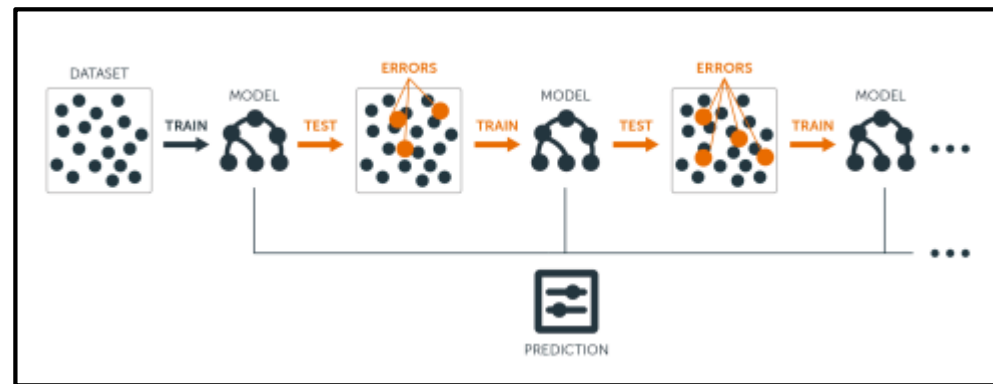Select features that are useful for predicting "time to failure"

Modeling Approaches

- Random Forest
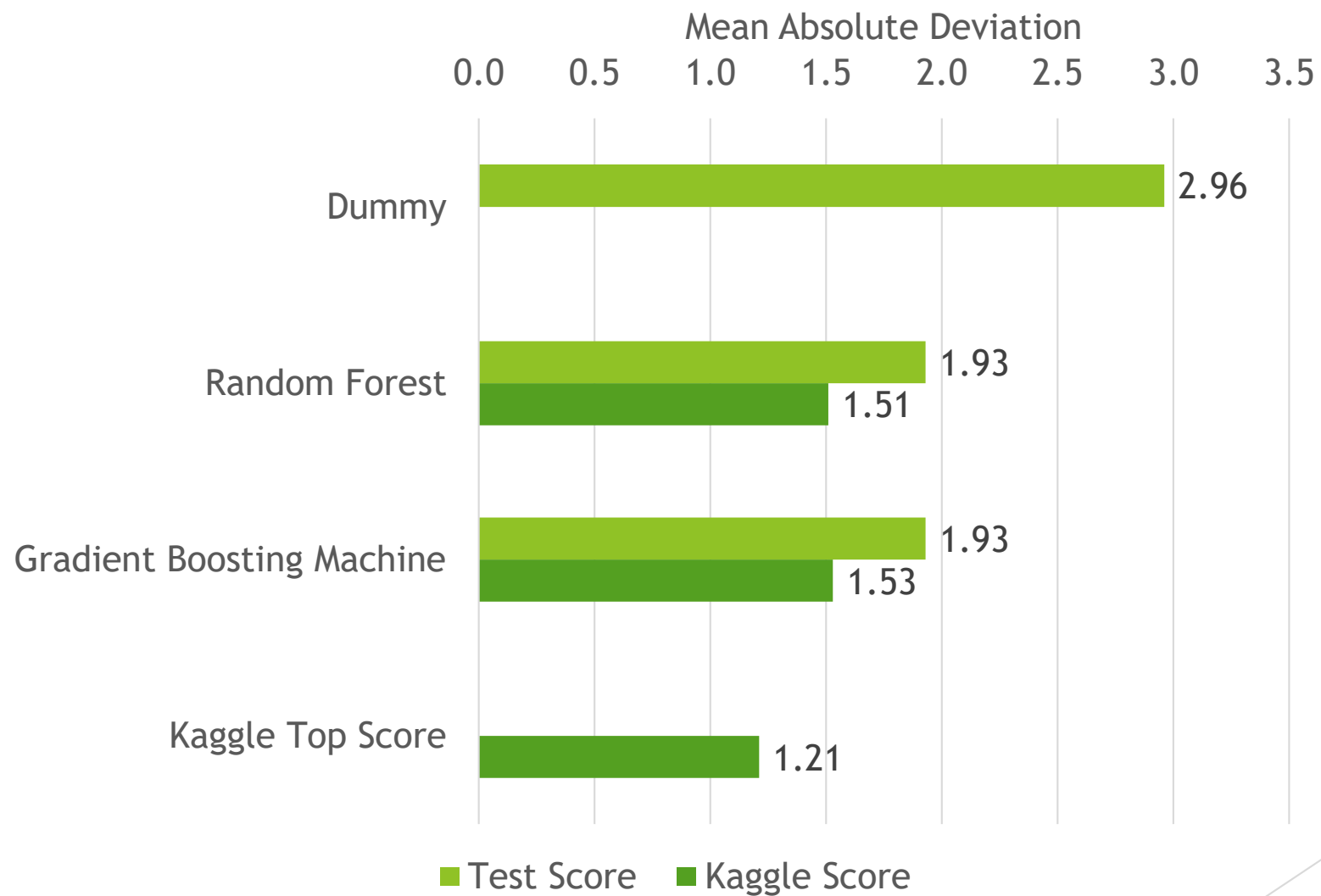

- Gradient-Boosting Machine

Results

Mean Absolute Deviation

	Test Score	Kaggle Score
Dummy	2.96	
Random Forest	1.93	1.51
Gradient Boosting Machine	1.93	1.53
Kaggle Top Score		1.21

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