BirdBible

Author: Ben Spilsbury

Overview

Business Problem:

There exists a high barrier to entry for birding as a hobby. Help alleviate high barrier to open up birding to larger populations

Objective:

Create multi-class audio recognition model that identifies bird species by recording of its call.

Data Understanding

- Xeno-canto online database
- Over 1,000 sample recordings used
- Recordings extracted in .wav format
- Three bird species considered for proof of concept
- Target variable is bird species

Feature Engineering

- Data initially extracted in standard volume output over time waveform.
- Six key audio characteristics extracted
    - Chroma Vector
    - Spectral Centroid
    - Spectral Bandwidth
    - Spectral Rolloff
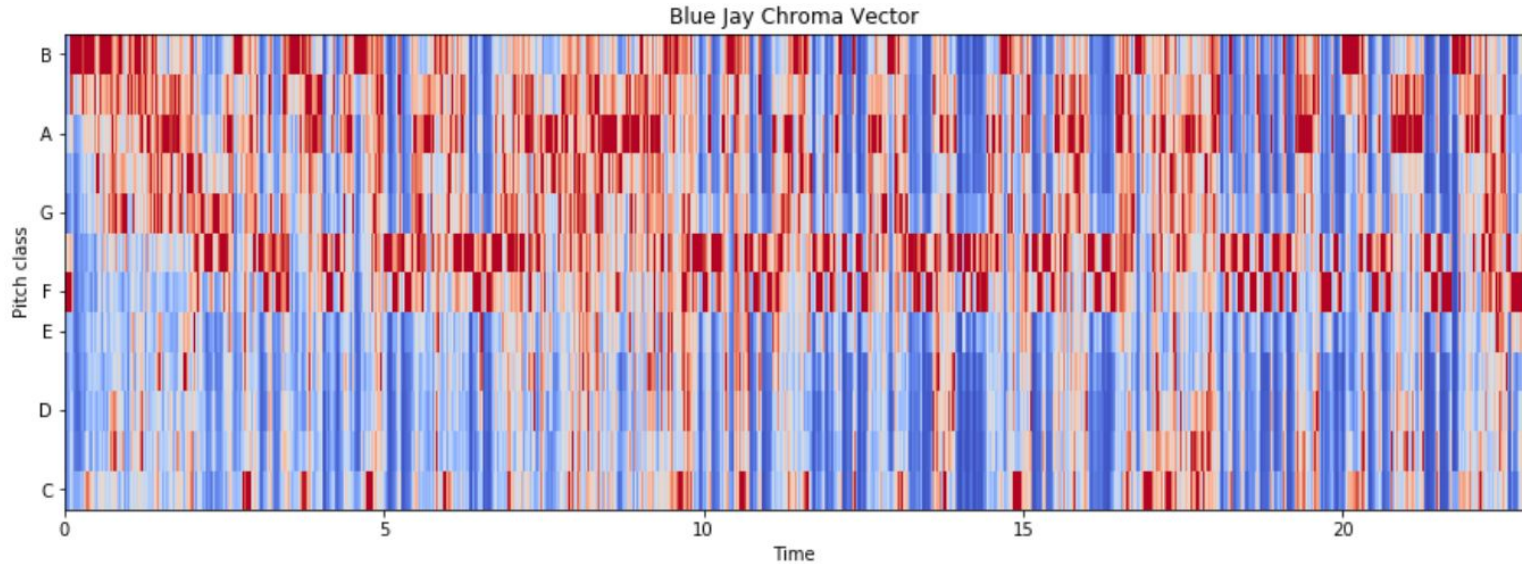   - Zero Crossing Rate
    - Energy Root Mean Square

Energy Root Mean Square

Gauge of general volume output of waveform. Useful for establishing volume output of call.

Standard Waveform
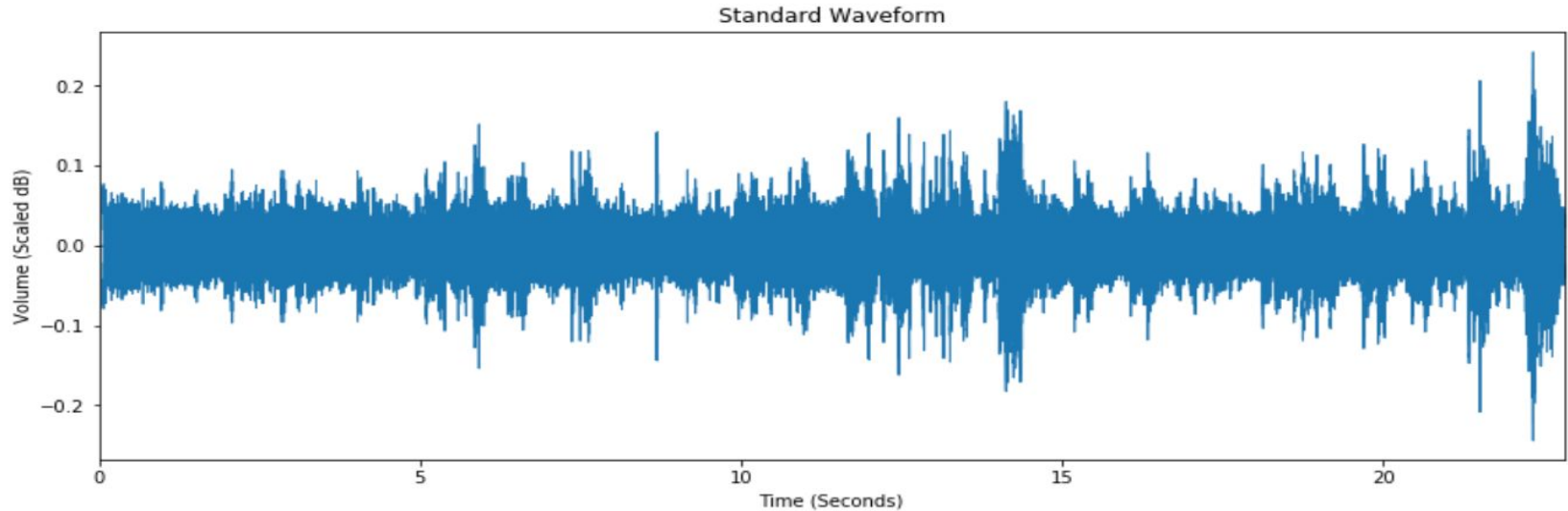
Chroma Vector

Energy output categorized by tone. Useful for establishing general pitch range of calls.


Blue Jay Chroma Vector

Spectral Centroid

Center of mass for frequency output. Weighted according to energy output. Useful for establishing general frequency level of bird call.


Spectral Centroid of a Blue Jay Call Superimposed upon Waveform
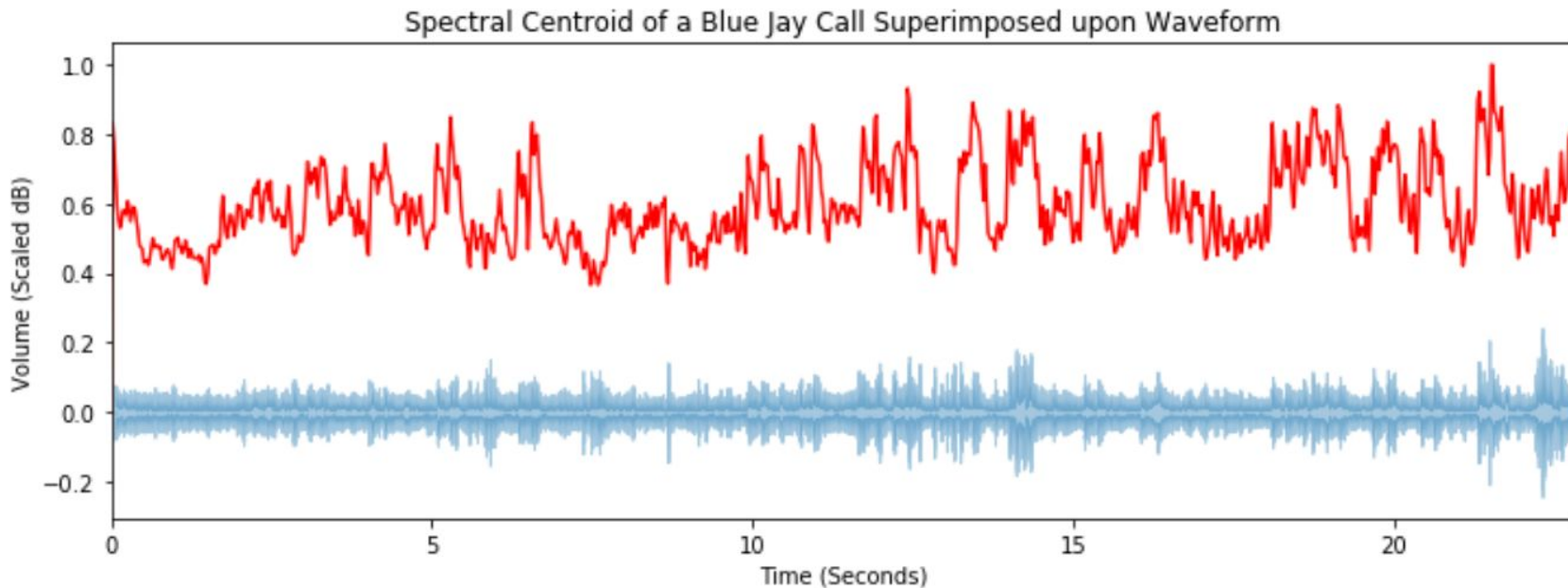
Spectral Bandwidth

Instantaneous frequency range. Useful for establishing tone character.

Spectral Bandwidth of a Blue Jay Call at Three Different Standard Deviation Thresholds
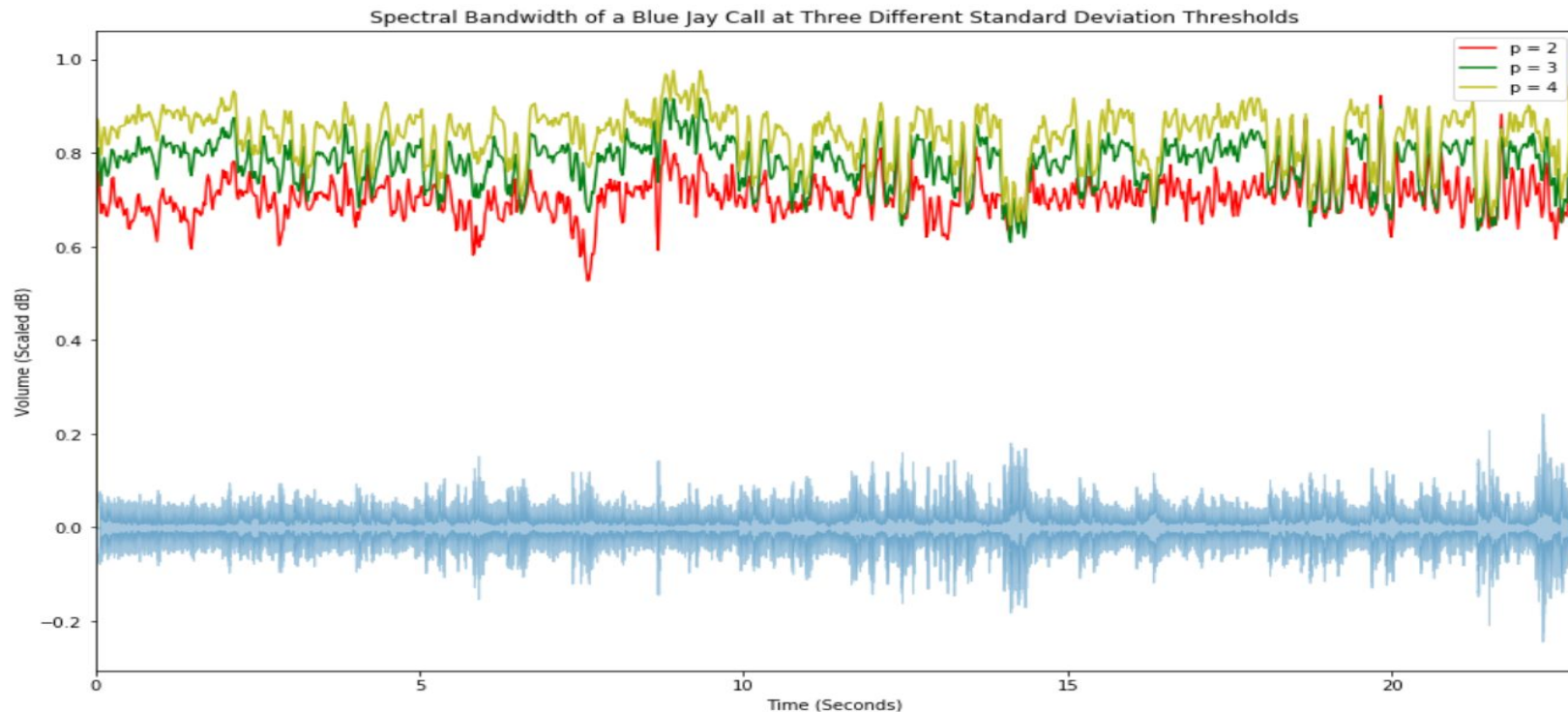
Spectral Rolloff

Higher bound to spectral bandwidth. Useful for identifying upper limit of bird's vocal range.

Spectral Rolloff of a Blue Jay Call at Superimposed upon Waveform
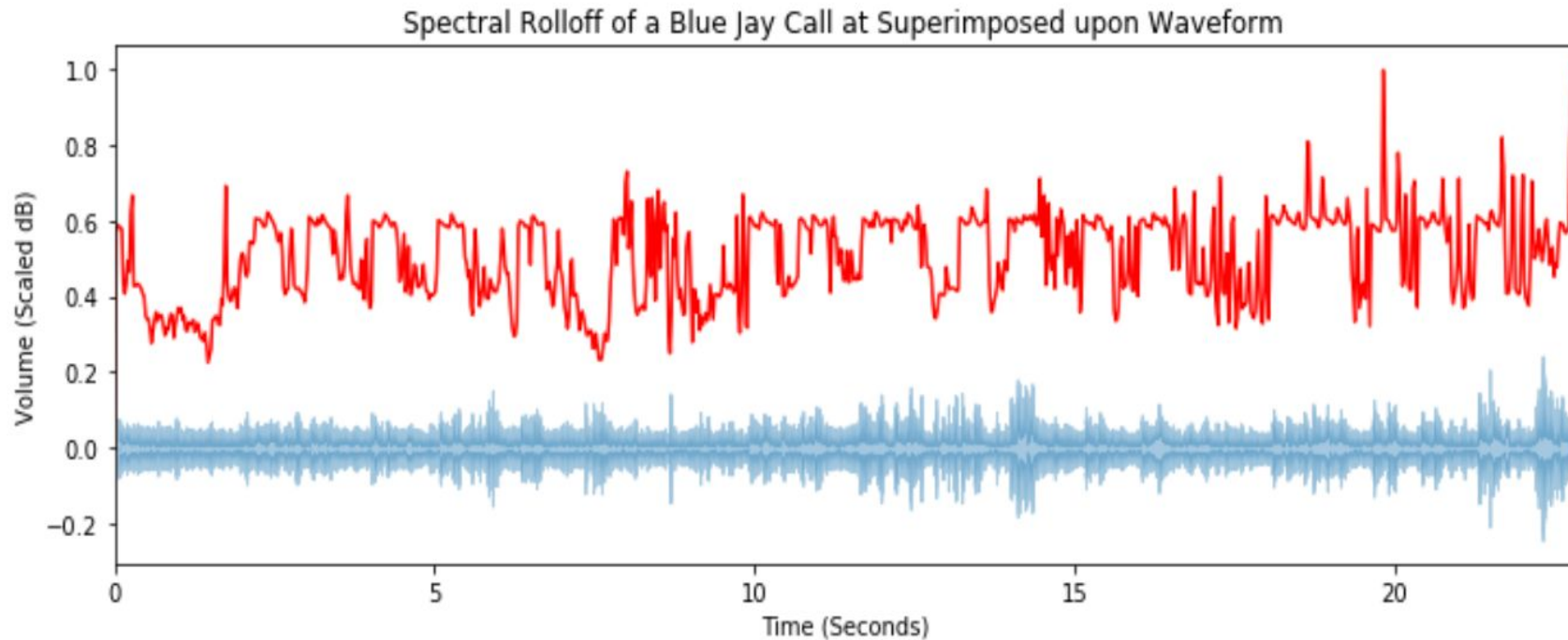
Zero Crossing Rate

Rate that the standard waveform passes from positive to negative decibels. Useful for establishing call's tonal quality.

Standard Waveform Zero Crossings
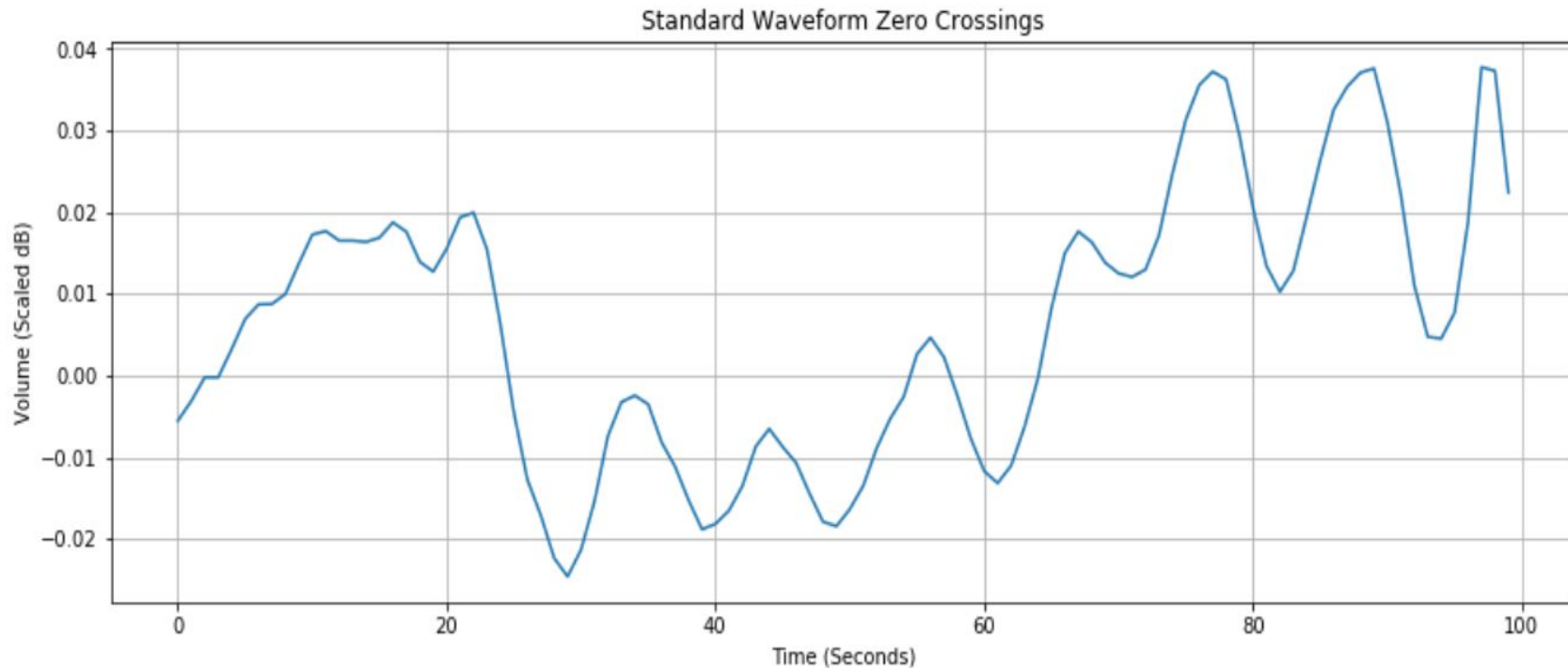
Modeling and Interpretation

Three models tested:

- Baseline decision tree
- Ensemble decision tree and random forest
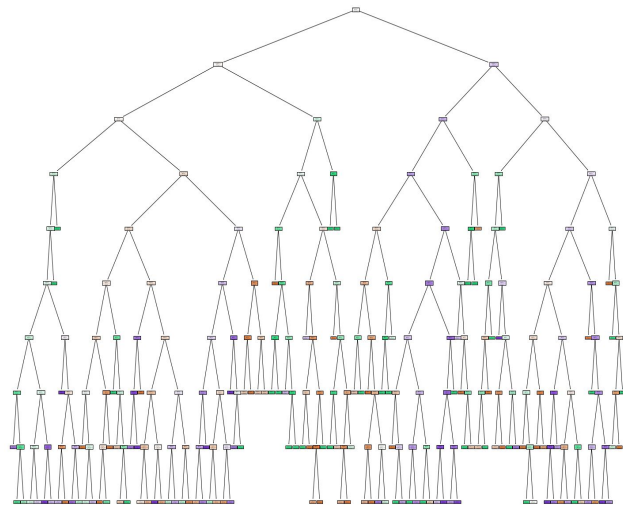- Sequential neural network

Accuracy used as success metric

- No real class imbalance
- No particular danger in miscategorization

Highest Performer: Ensemble Model
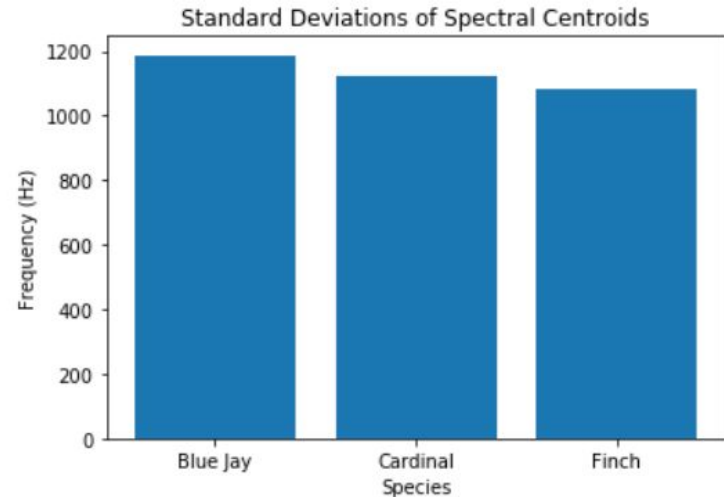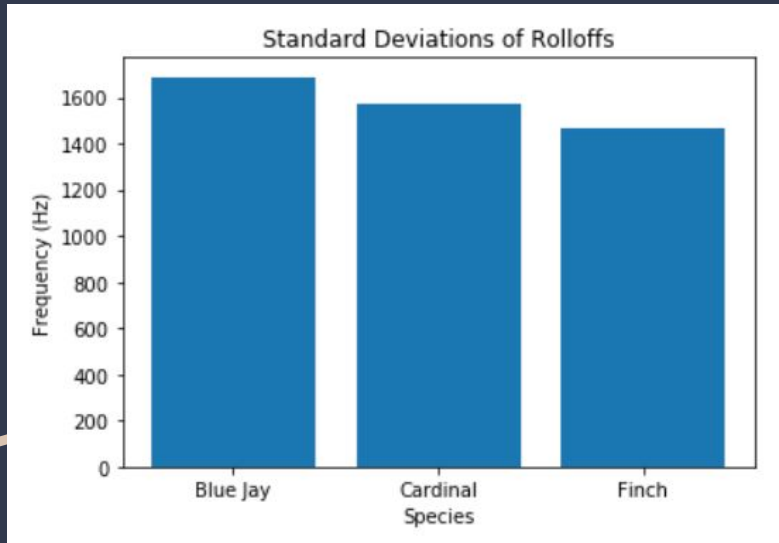
Ensemble model:

- Accuracy of 0.47
- Beats decision tree baseline of 0.44
- Underperforms compared to ~0.49% share of majority class
- Interpretability of model is lacking

Error Attribution:

- Standard deviations of characteristics examined
- Rolloffs and spectral centroids showed percentage difference of ~10% and ~15%

→ *Low recording integrity implied*

Standard Deviations of Rolloffs

Standard Deviations of Spectral Centroids

Conclusion and Next Steps

Results

- Model performs poorly
    - Ensemble of random forest and decision tree best
    - Only 47% accuracy
    - Majority class percent share at ~49%

Next Steps

- Improve integrity of sample data set
    - Remove white noise
    - Cull succinct "paragon" calls with fewer samples
- Continue to fine tune neural network to improve accuracy