Music Genre Classification with Machine Learning

Christiaan Dageforde

Goals

- Utilize Librosa methods for extracting spectral features from audio
- Build a model that can accurately predict the musical genre of an audio signal

Key Definitions

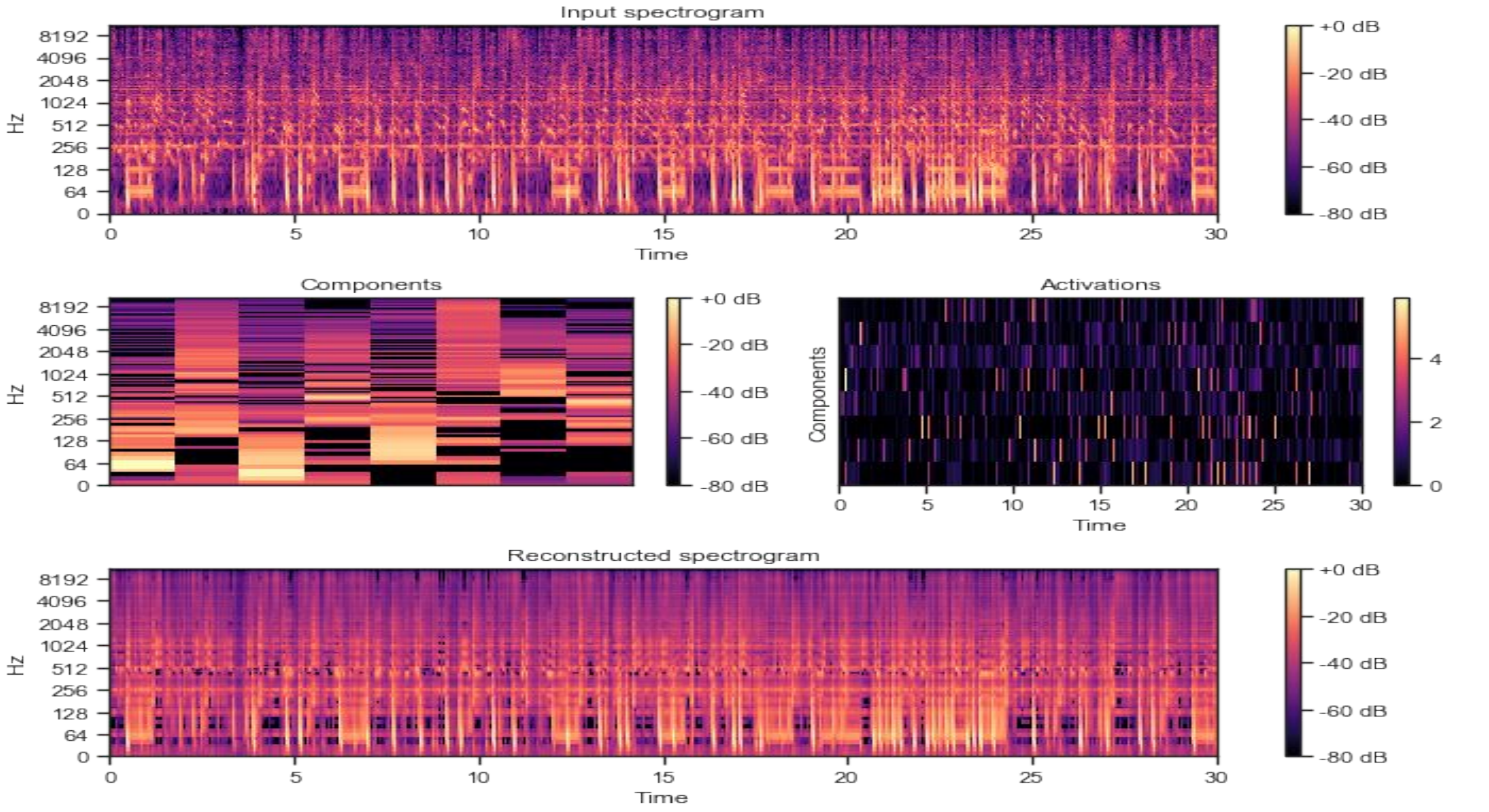
- Digital Signal Processing
  - The process of transforming analog signals into digital ones
    - Audio/speech
    - Sonar & radar
    - Digital images
- Librosa
  - A Python package for audio analysis. Librosa provides us with the methods to build music information retrieval systems
- Musical Information Retrieval (MIR)
  - A broad field describing the process of retrieving information from music
  - Applications in psychoacoustics, musicology, machine learning, and so, SO much more!

Data

- GTZAN
  - 1000 audio files
  - 30 seconds each
- Ten Genres [Label Values]
  - Blues
  - Classical
  - Country
  - Disco
  - Hip-hop
  - Jazz
  - Metal
  - Pop
  - Raggae
  - Rock
- Extracted spectral features with Librosa

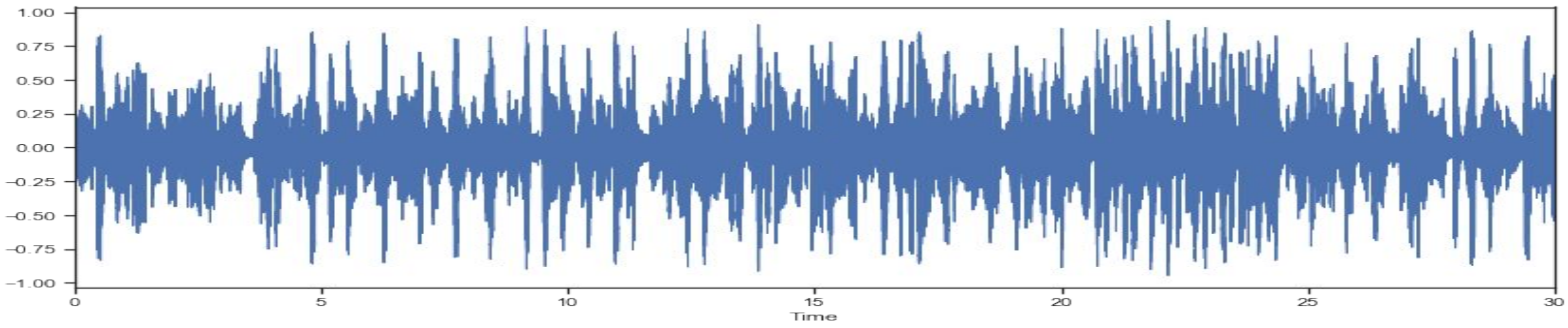
Spectrogram

- A visual representation of the spectrum of frequencies of sound
- Frequency
    - The rate at which a sound wave repeats over time
- Frequency vs Amplitude
    - Size (Amp) vs Speed (Freq)
- Spectrogram
    - X-Axis: Time
    - Y-Axis: Frequency
    - Shading: Amplitude

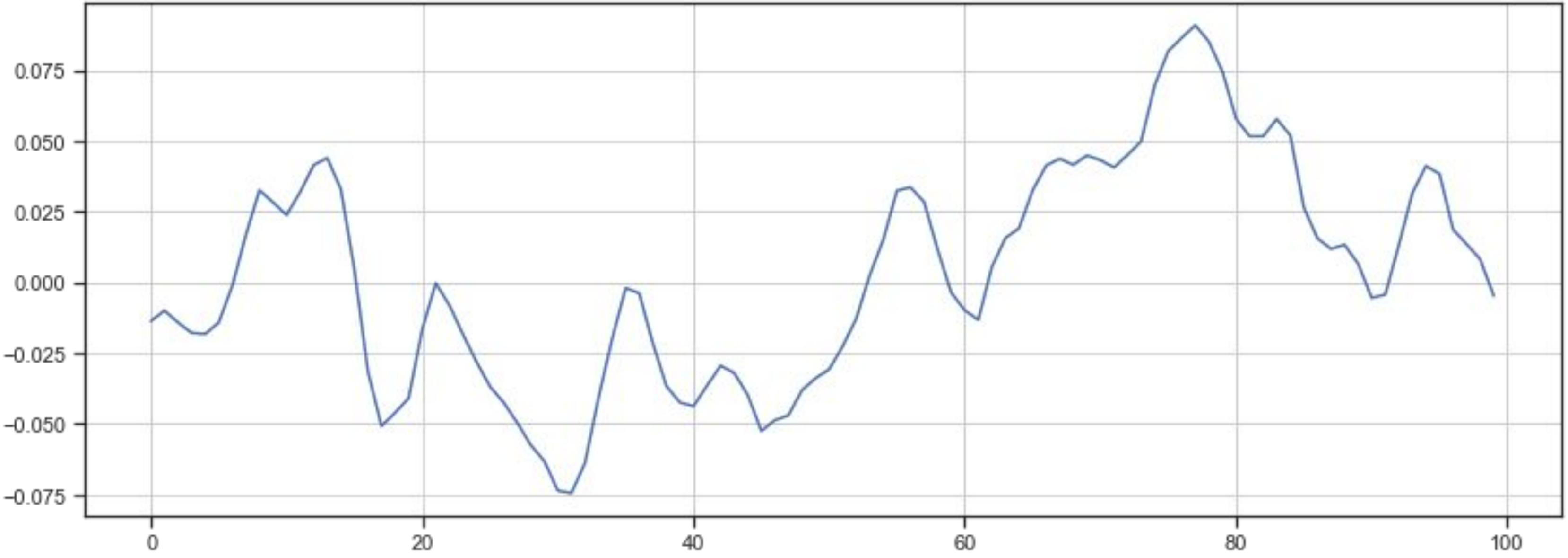

Input spectrogram
Hz
Time
Components
Activations
Reconstructed spectrogram
+0 dB
-20 dB
-40 dB
-60 dB
-80 dB

Waveforms

- Waveforms are a visualization of a sound's amplitude envelope
- Amplitude
  - Fluctuation of a wave from its' mean value
  - In this case, the extent to which air particles are displaced
  - "Loudness"
- Envelope
  - How the amplitude changes over time

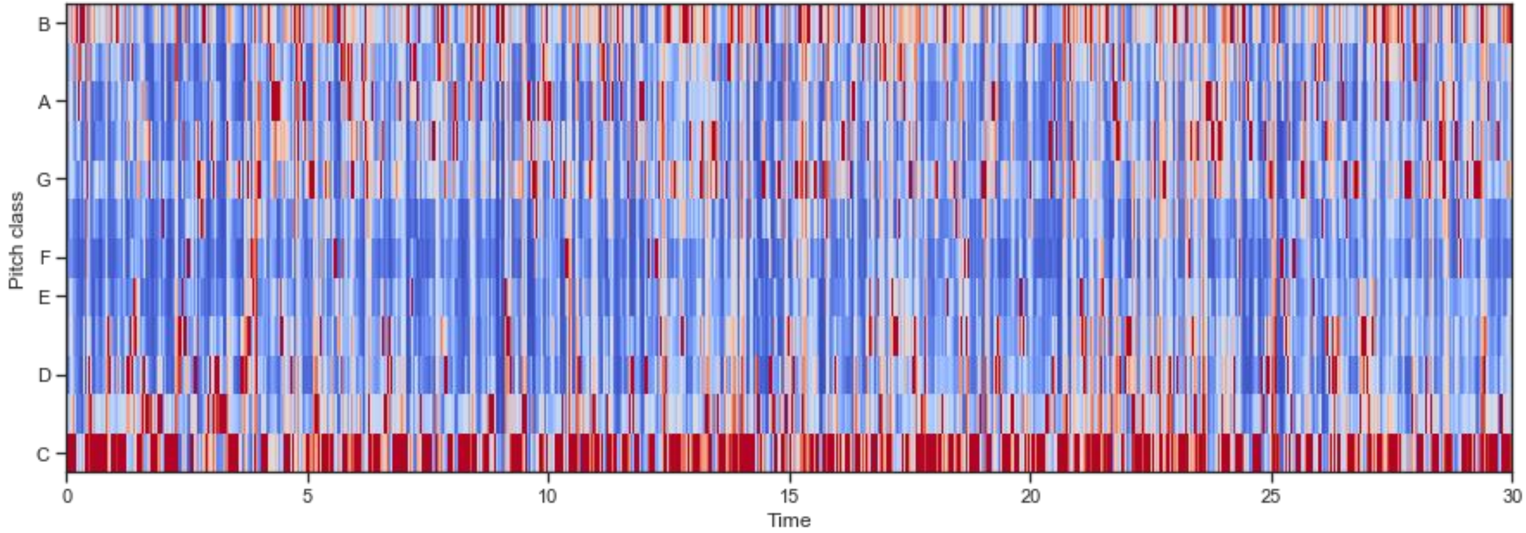


Zero-Crossing Rate



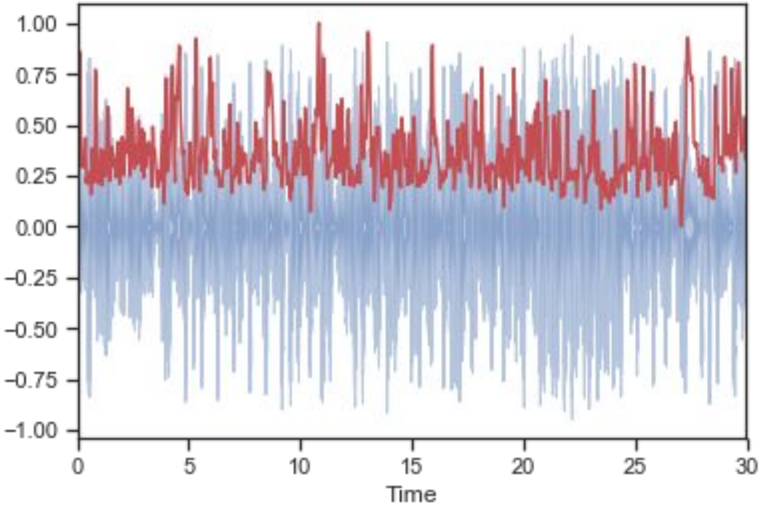
- The rate at which a signal changes from positive to negative
  - Can aid in describing music with a percussive focus

Chromagram



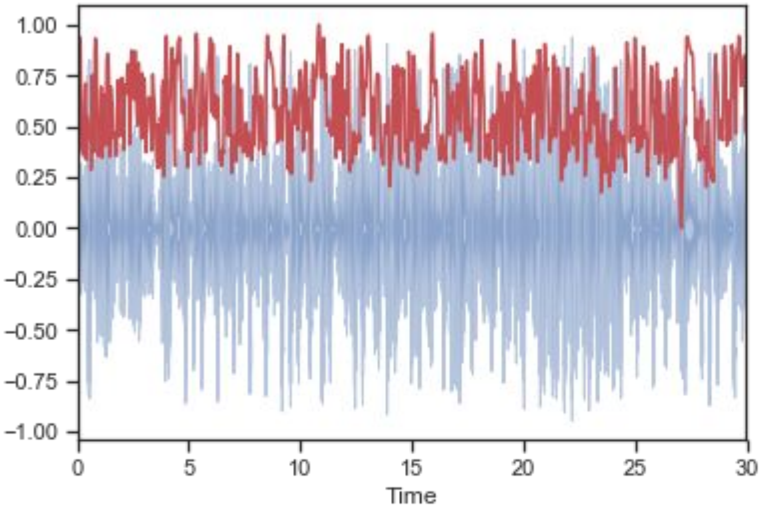
- Divides spectrogram into 12 bins based on absolute frequency
  - Relationship between notes; insight into melody & harmony

Spectral Centroids



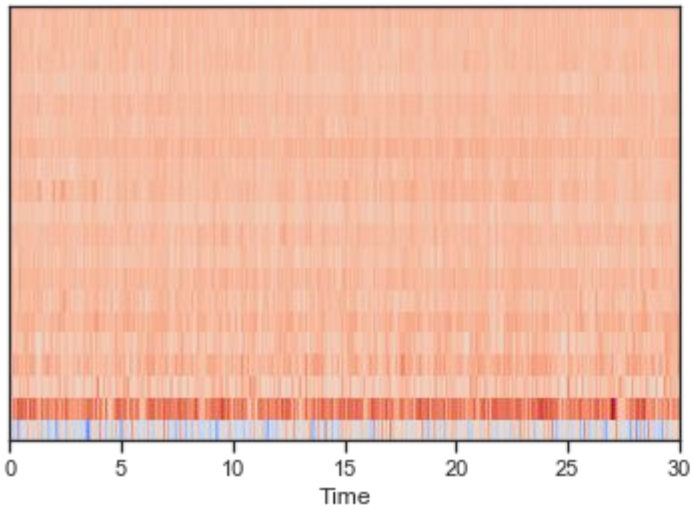
- “Center of spectral mass”
- Weighted mean of frequencies present in a recording
- Density of higher frequencies
    - “Brightness”

Spectral Rolloff



- The shape of the signal
- The frequency below which a specified percentage of total spectral energy lies

Mel-Frequency Cepstral Coefficients



- 10 - 20 “snapshots” of the overall shape of the spectral envelope
    - I used 20 in my data
- Models qualities of the human voice

Modeling

Neural Network

- Topology
  - Input
    - Dense; 256 nodes
  - Hidden (Two Layers)
    - Dense; 128 nodes
    - Dense; 64 nodes
  - Output
    - 10 nodes
      - One for each class label
- Difficult to interpret
  - “Black box”

Random Forest Classifier

- Selects a random subset of features for each split in the tree
- Better interpretability
- Few parameters to tune
- Best Parameters
  - Maximum depth: None
  - Min. Samples Split: 2
  - Number of estimators: 80

Extra Trees Classifier

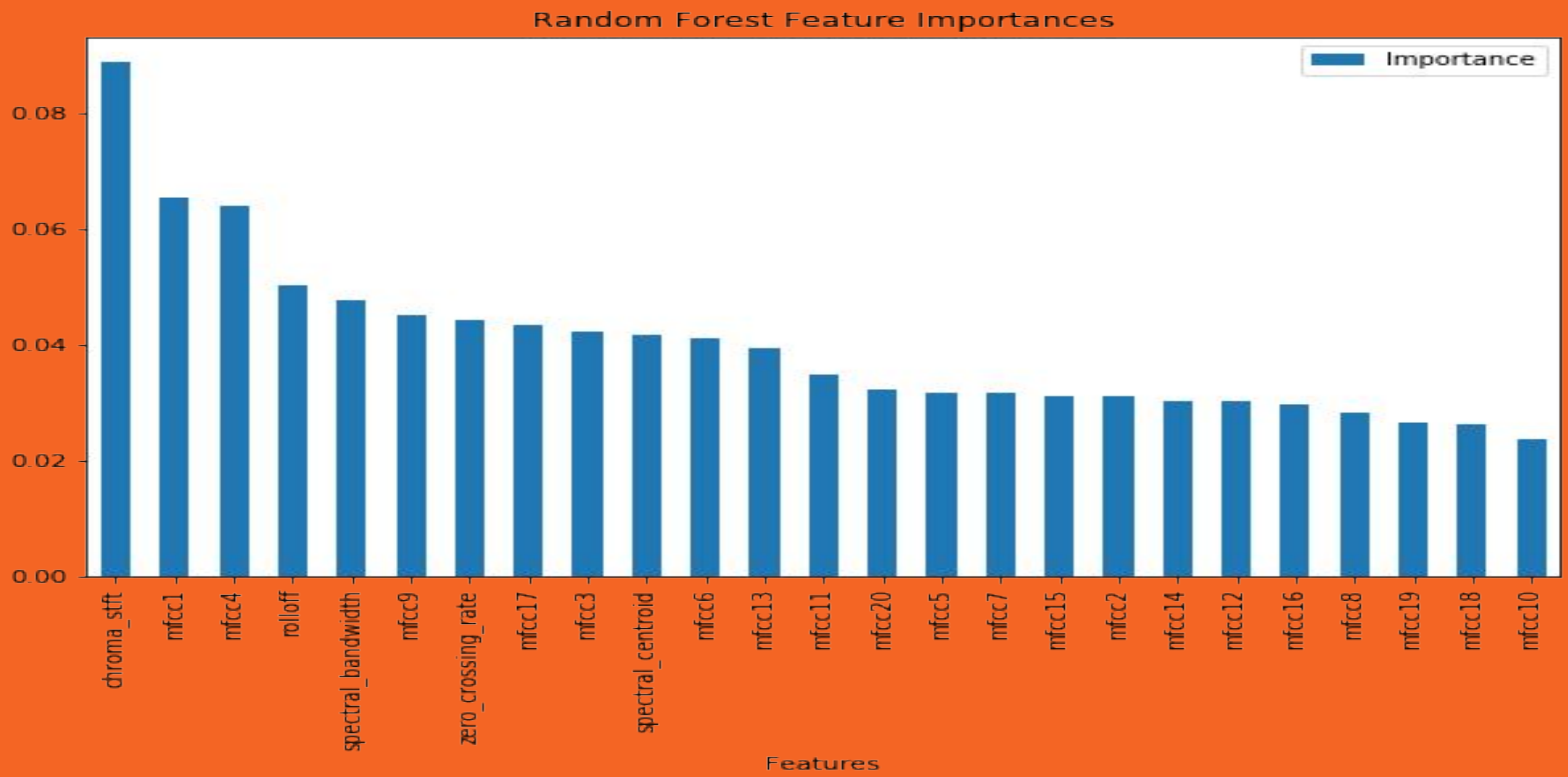
- Instead of computing optimal feature/split combination, a random value is assigned for each split
- Helps to further de-correlate our trees
- Best Parameters
    - Max Depth: None
    - Min Samples Split: 2
    - Number of Estimators: 40

Findings

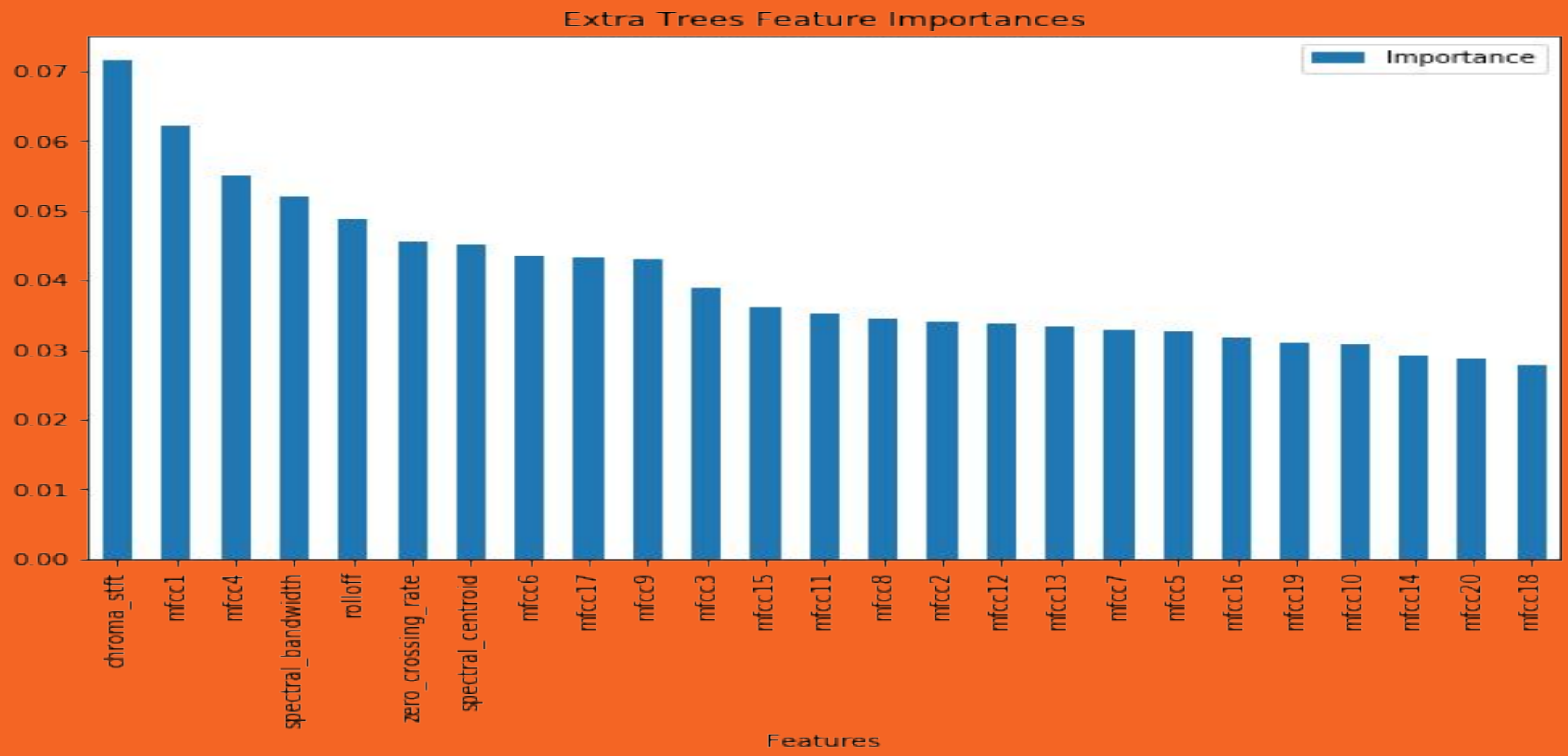
Metrics

- Feature importance
  - Hierarchy of features crucial to the classification algorithm
- Precision
  - Percentage of correctly predicted positive values
- Recall
  - How many positive values did we correctly predict?
- F1
  - Harmonic average of Precision & Recall

Random Forest Feature Importances



Extra Trees Feature Importances



Best Grouping

Model	Label	Precision	Recall	F1
Neural Network	Classical [1]	.93	1	.96
Random Forest Classifier	Classical [1]	.93	.1	.96
Extra Trees Classifier	Classical [1]	.76	.89	.82

Worst Grouping

Model	Label	Precision	Recall	F1
Neural Network	Hip-Hop [9]	.34	.48	.40
Random Forest Classifier	Rock [9]	.44	.33	.38
Extra Trees Classifier	Rock [9]	.39	.5	.44

Musical Interlude

Further Steps

- More data!
  - Larger volume of recordings
  - Full-length recordings for structural analysis
- Tuning Models
  - Experimenting with NN topology
- Building Blocks for more complex projects
  - Sample identification
  - Instrument identification
  - Recommender systems

Questions?