Music Genre Identification

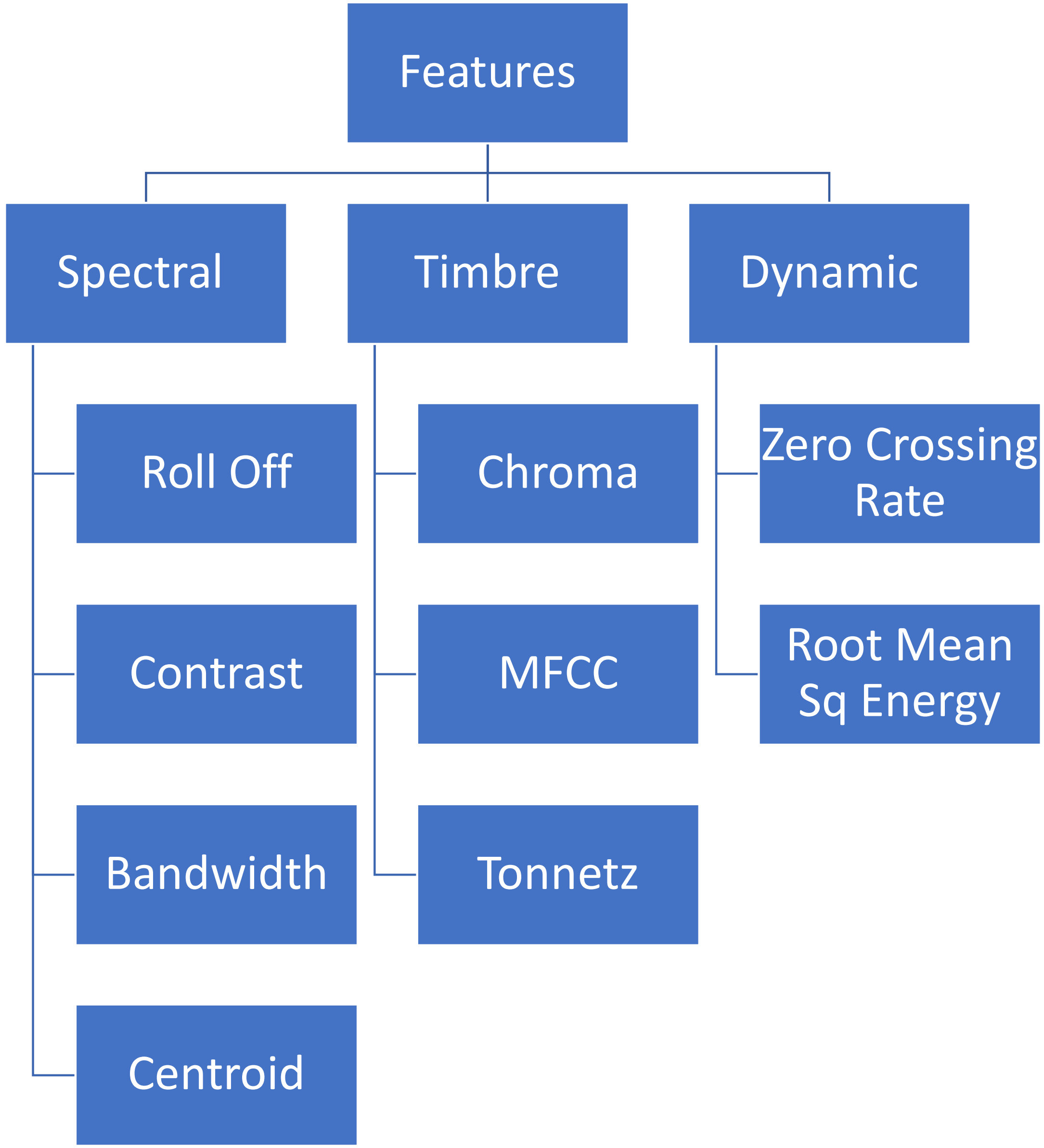
Dan O’Connor

Introduction

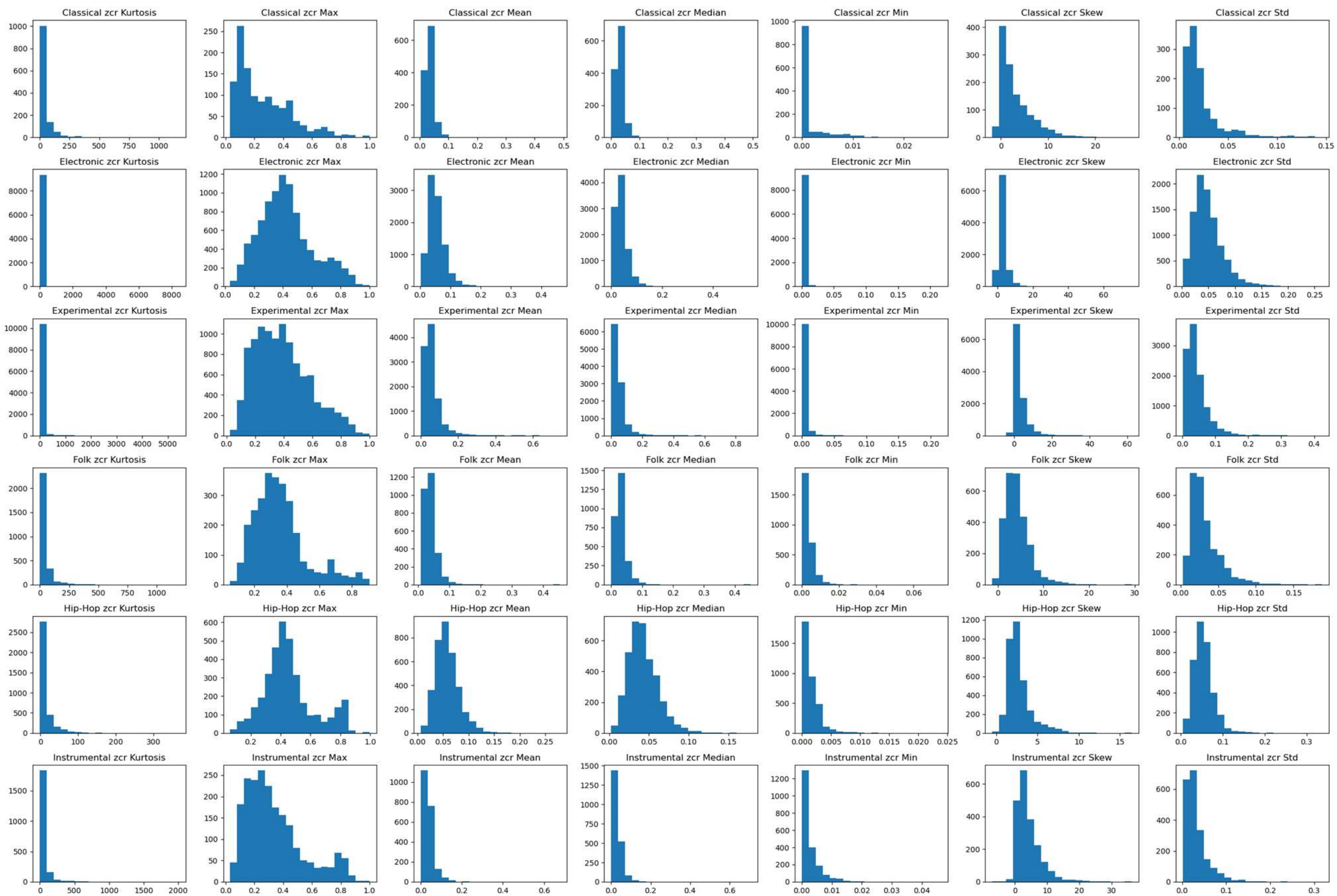
- **Problem statement**: Can musical genres be identified solely from acoustic features extracted from a 30 second snippet of the song?
- Useful for:
  - Large scale genre and subgenre classification
    - More efficient and accurate organization
  - Music recommender systems
    - Music libraries, radio stations, music streaming services

Data Collection

- Free Music Archive (FMA) is an online repository of royalty-free music
- Acoustic features come from a paper called “FMA: A Dataset For Music Analysis” (Defferrard et al., 2017)
- Feature summary:
  - **Spectral features**: captures different frequencies within the audio
  - **Timbre features**: captures unique sound qualities
  - **Dynamic features**: captures change in loudness

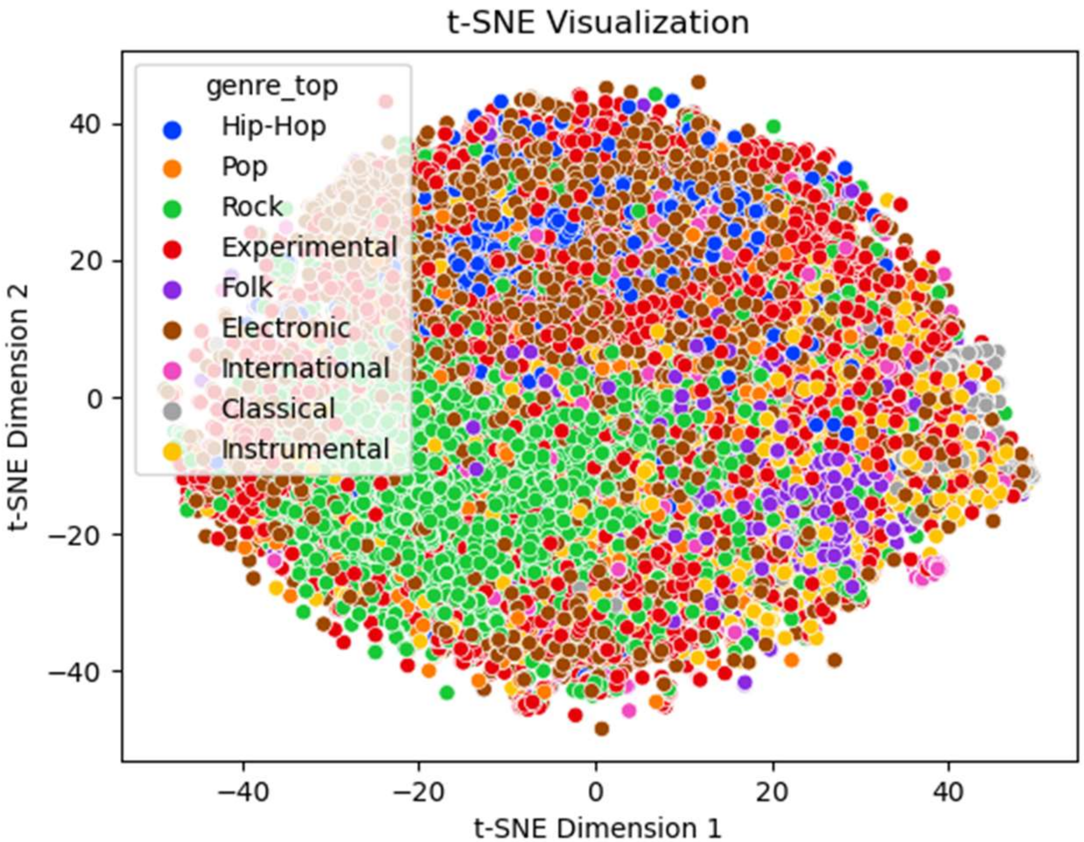


Features
Spectral
Timbre
Dynamic
Roll Off
Contrast
Bandwidth
Centroid
Chroma
MFCC
Tonnetz
Zero Crossing Rate
Root Mean Sq Energy


Classical zcr Kurtosis
Classical zcr Max
Classical zcr Mean
Classical zcr Median
Classical zcr Min
Classical zcr Skew
Classical zcr Std
Electronic zcr Kurtosis
Electronic zcr Max
Electronic zcr Mean
Electronic zcr Median
Electronic zcr Min
Electronic zcr Skew
Electronic zcr Std
Experimental zcr Kurtosis
Experimental zcr Max
Experimental zcr Mean
Experimental zcr Median
Experimental zcr Min
Experimental zcr Skew
Experimental zcr Std
Folk zcr Kurtosis
Folk zcr Max
Folk zcr Mean
Folk zcr Median
Folk zcr Min
Folk zcr Skew
Folk zcr Std
Hip-Hop zcr Kurtosis
Hip-Hop zcr Max
Hip-Hop zcr Mean
Hip-Hop zcr Median
Hip-Hop zcr Min
Hip-Hop zcr Skew
Hip-Hop zcr Std
Instrumental zcr Kurtosis
Instrumental zcr Max
Instrumental zcr Mean
Instrumental zcr Median
Instrumental zcr Min
Instrumental zcr Skew
Instrumental zcr Std

EDA cont.

- t-SNE:

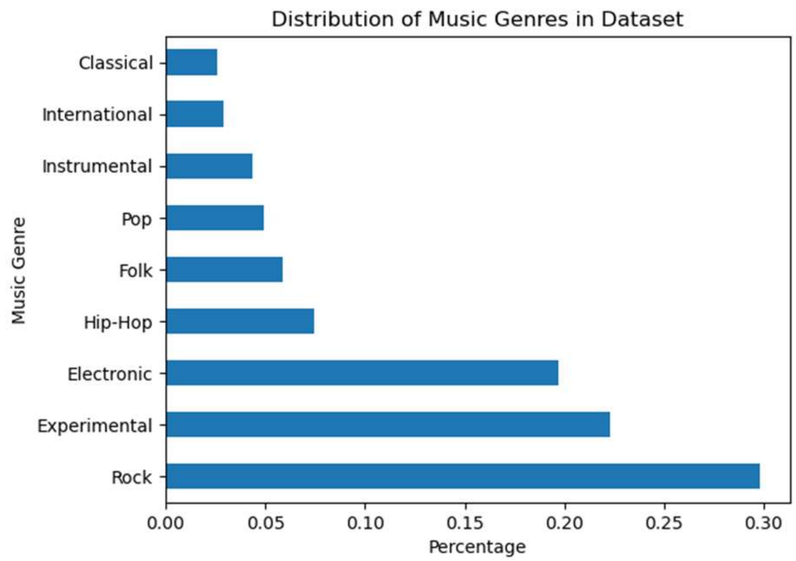


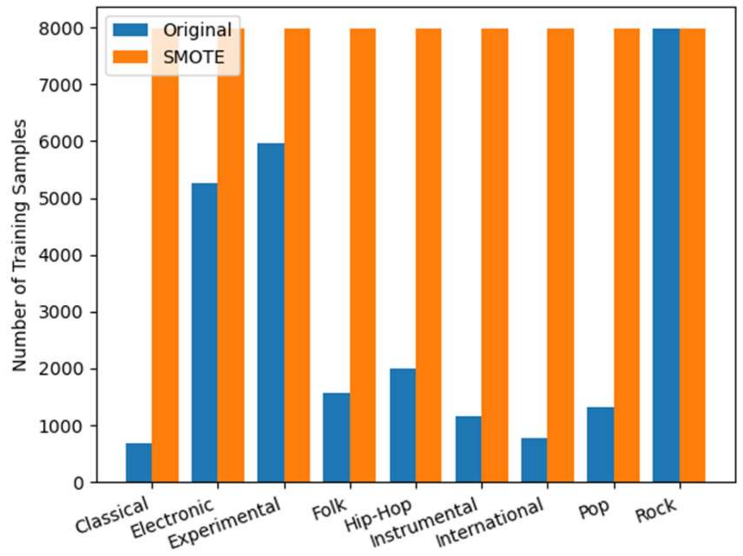
Modeling Hyperparameters

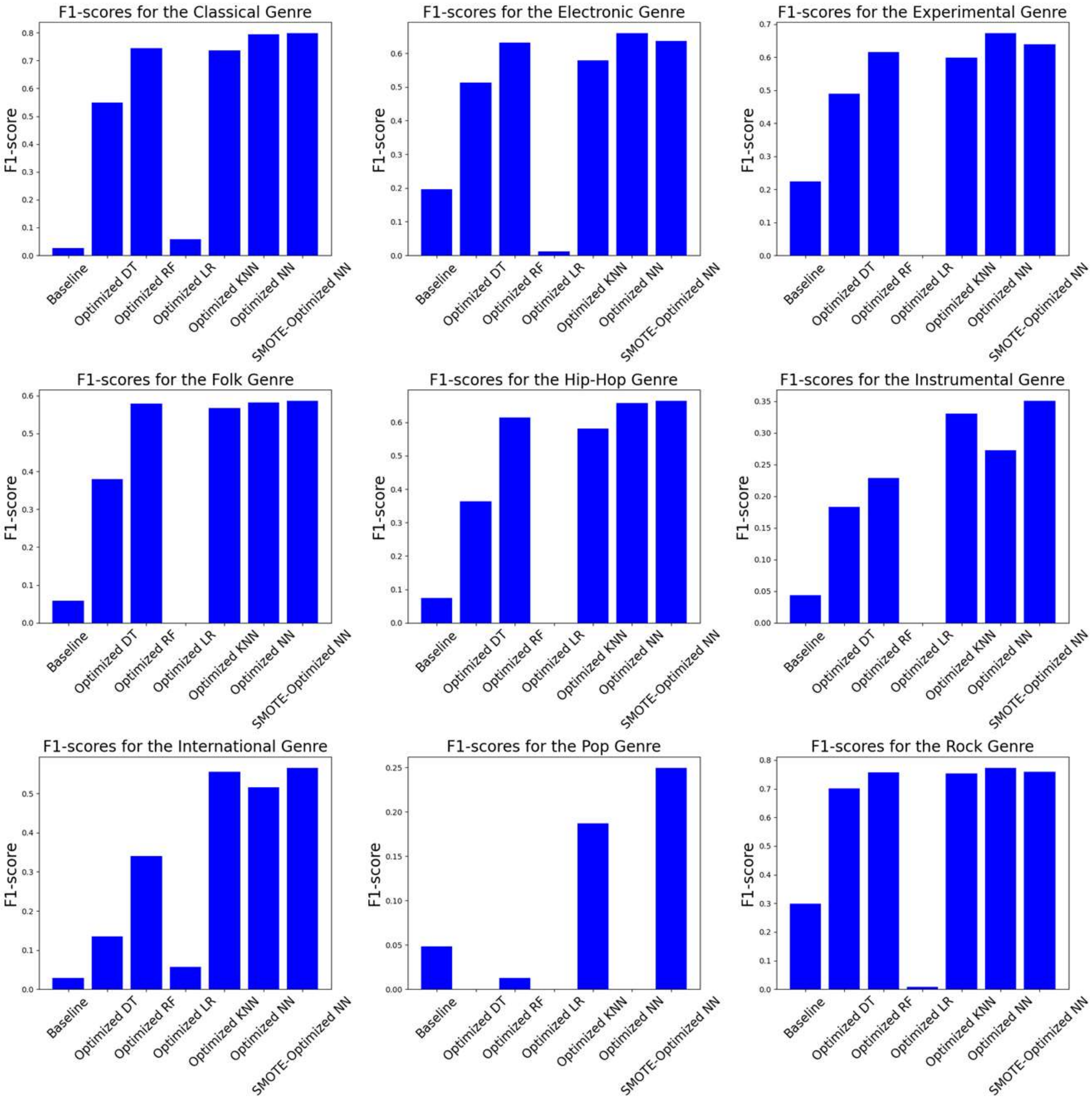
- Decision Tree
  - Max depth = 7, min samples/leaf =11 , gini
- Random Forest
  - Max depth = 15, num estimators = 300
- Logistic Regression
  - Standard scaler, C=0.01, L2
- KNN
  - Standard scaler, num neighbors = 7, weights = distance
- Neural Network
  - 3 hidden layers, Adam, relu, 50% dropout/layer

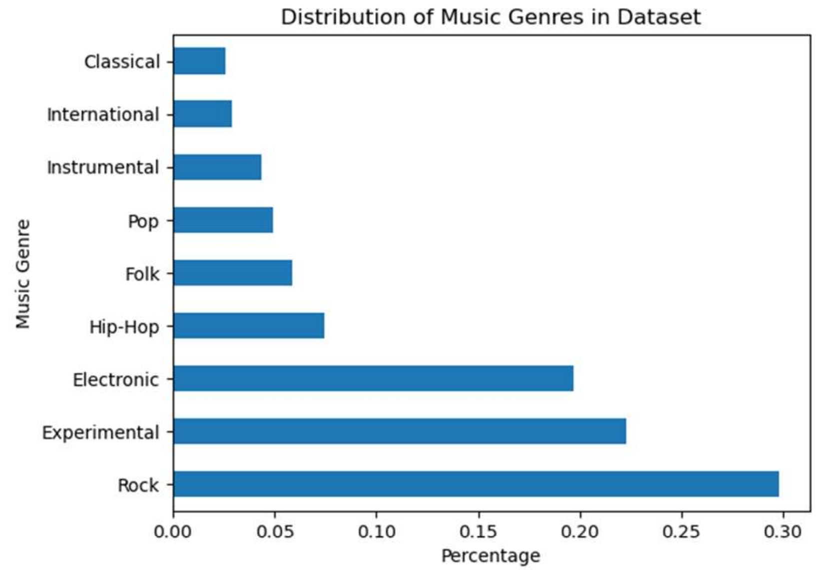
Imbalanced Dataset

- Synthetic Minority Over-Sampling Technique (SMOTE)
  - Applied to NN



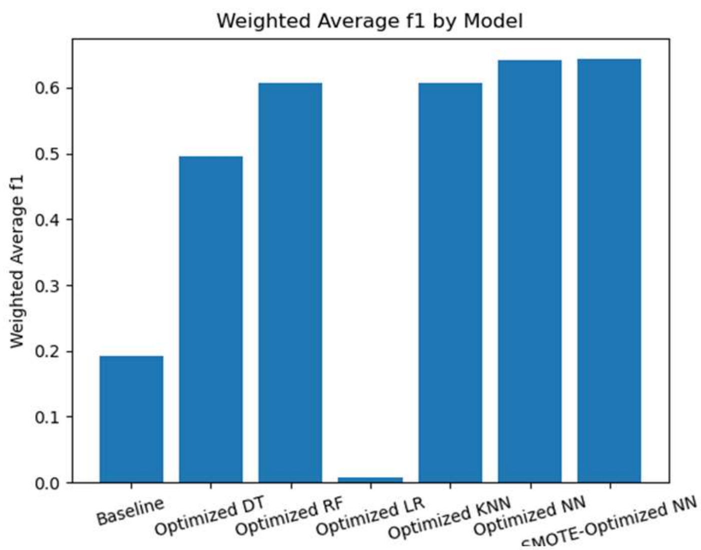


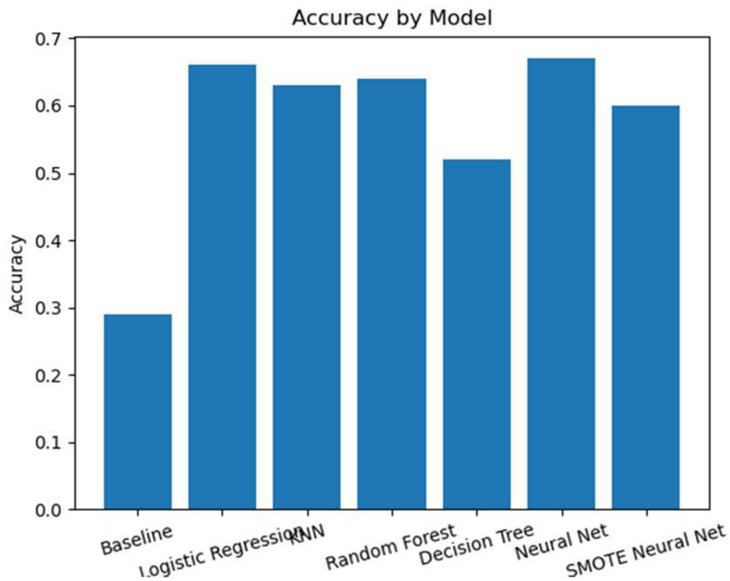

F1-scores for the Classical Genre
F1-scores for the Electronic Genre
F1-scores for the Experimental Genre
F1-scores for the Folk Genre
F1-scores for the Hip-Hop Genre
F1-scores for the Instrumental Genre
F1-scores for the International Genre
F1-scores for the Pop Genre
F1-scores for the Rock Genre
F1-score
Baseline
Optimized DT
Optimized RF
Optimized LR
Optimized KNN
Optimized NN
SMOTE-Optimized NN


Distribution of Music Genres in Dataset
Music Genre
Classical
International
Instrumental
Pop
Folk
Hip-Hop
Electronic
Experimental
Rock
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Percentage

Model Performance

- Two main metrics:
  - Accuracy
  - Weighted F1
- Best model
  - NN

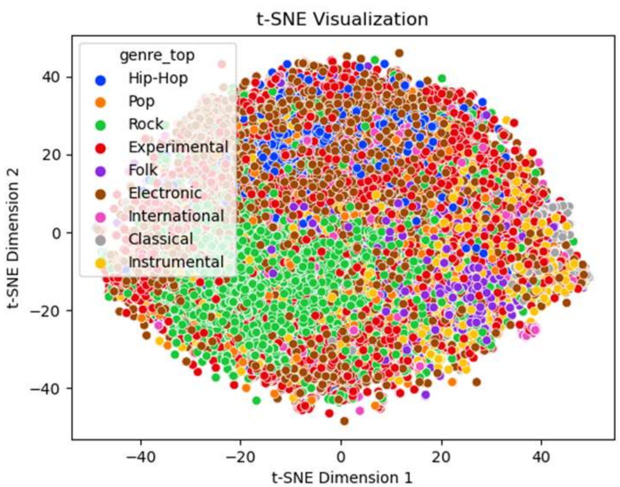




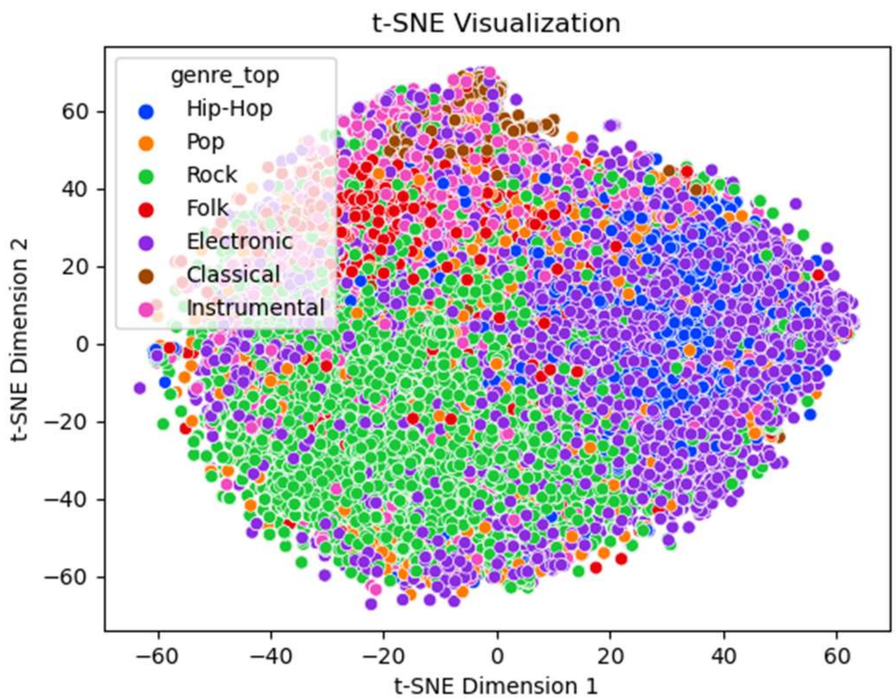
Continuation of Project

- Fix logistic regression
- Further tune NN
- Work with less genres

9 Genre t-SNE

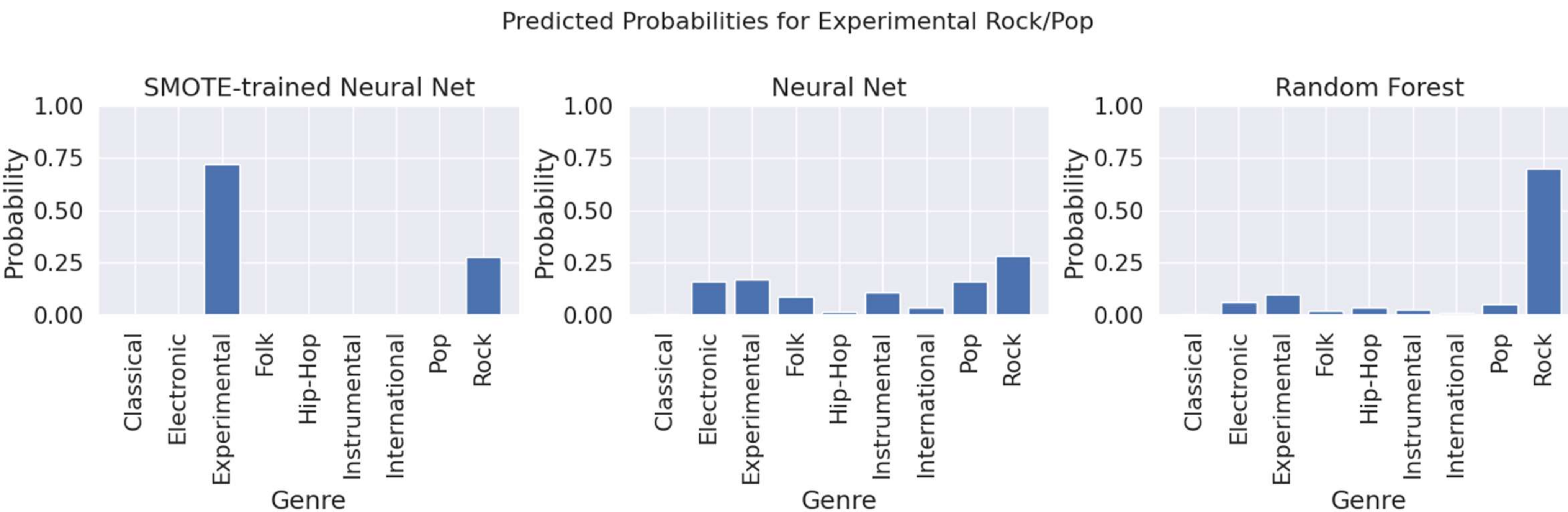


7 Genre t-SNE



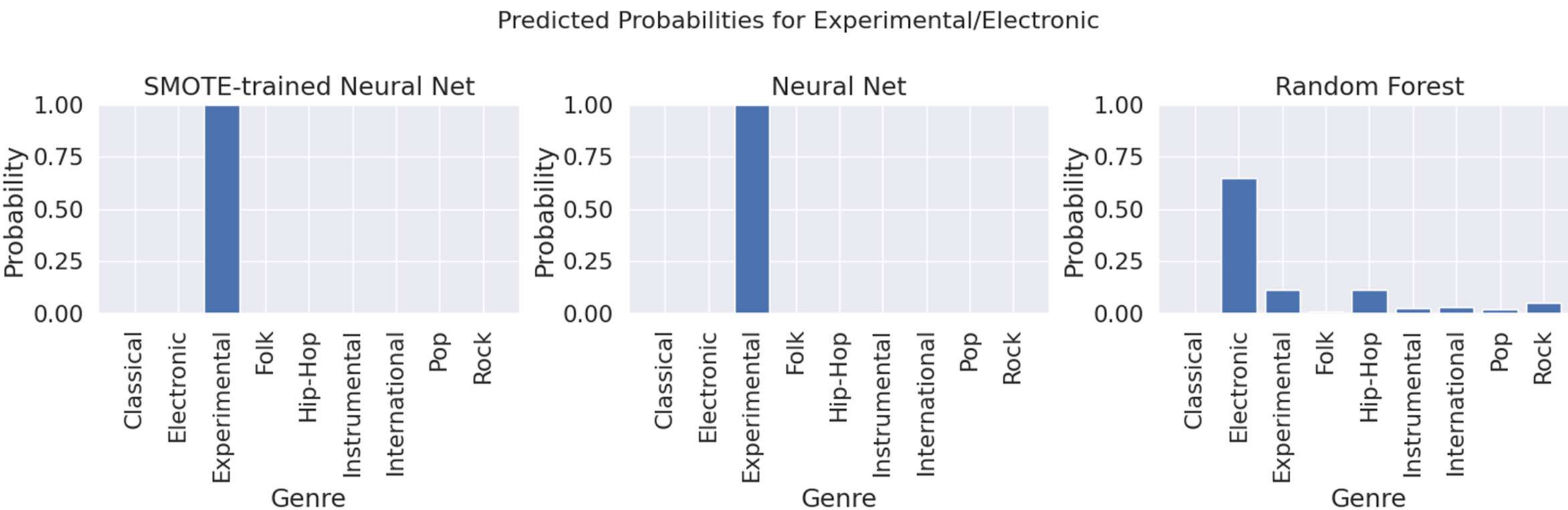
Unseen Sample 1

- Referents by Bird Names
  - Experimental Rock and Pop



Unseen Sample 2

- Frenetic (ft. Audiologist) by Nonima
  - Electronic and Experimental



Thank you!

- Questions?

Confusion Matrix: NN and SNN