Music Genre Classification

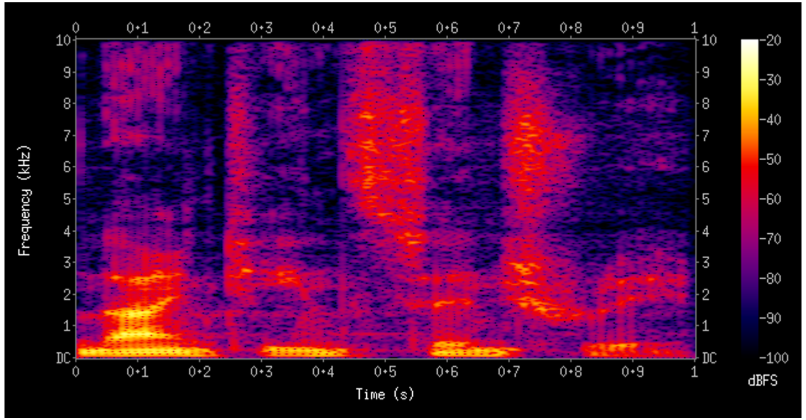
EEE498/591
Fall 2021
Naive Boyes: Dale, Zach, Adan, Jonah

Introduction

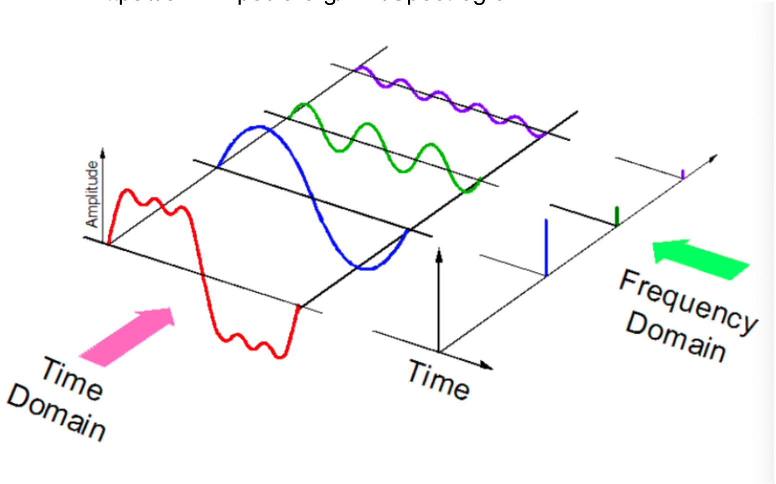
- Music streaming services like Spotify have over 70 million tracks available in 184 markets
- Spotify recommends tracks to its 172 million subscribers by classifying each track into certain music genres
- An incentive exists to accurately **classify** tracks and **make recommendations** to users and encourage subscription to the service
- We explore different ML algorithms and compare their performance in classifying tracks by measuring train/test accuracy and training time

Theory

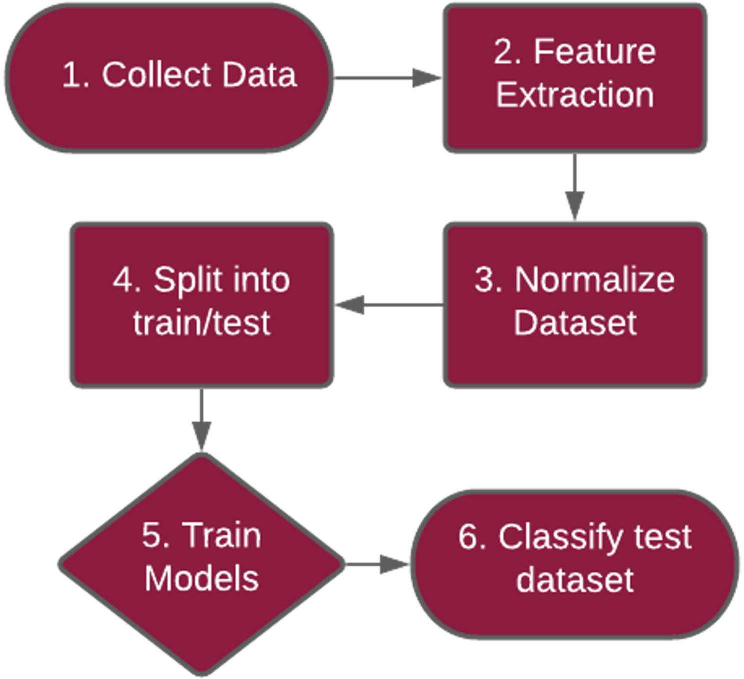
- Sound can be represented as audio signals with parameters including frequency, bandwidth, and decibels.
- These parameters are useful features for analysis
- Songs within a genre have similar patterns that allow ML algorithms to train and classify songs



https://en.wikipedia.org/wiki/Spectrogram



Methodology



1. GTZAN dataset contains 1000 audio tracks each 30 seconds long. It contains 10 genres, each represented by 100 tracks.
2. Zero crossing rate, spectral centroid, spectral rolloff, MFCCs, Chroma frequencies, spectral bandwidth, and root mean square.
3. Standard scaler and PCA
4. 30% used as test dataset
5. KNN, SVM, Logistic Regression, Random Forest, MLP, Naive Bayes, Decision Tree, NN w/ dropout & L2 regularization
6. Accuracy metrics and confusion matrices

Results

- **Overfitting** is prevalent
- **20~50%** difference in training/testing

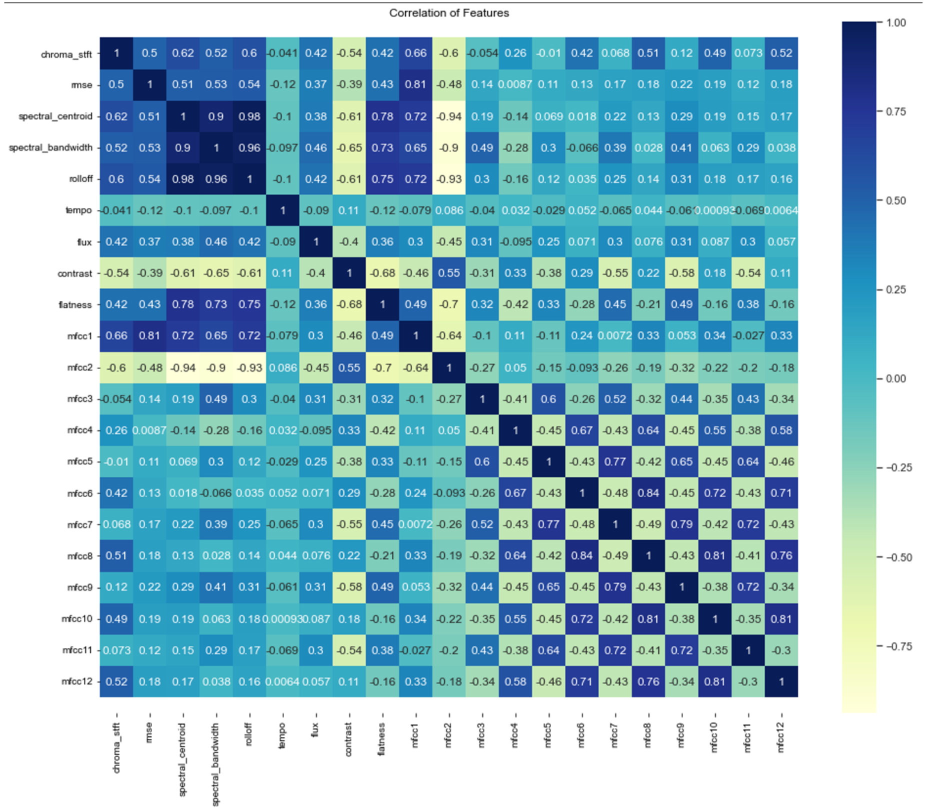
Algorithm	Test Accuracy	Train Accuracy	Training Time
KNN	64%	78%	1.39s
SVM	60%	86%	6.30s
Logistic Regression	64%	78%	.045s
Random Forest w/ parameter tuning	67%	100%	73.32s
MLP	59%	100%	.22s
Naive Bayes	57%	70%	.002s
Decision Tree	48%	100%	.008s
NN w/ DO	72%	93%	22.66s

Correlation Matrix

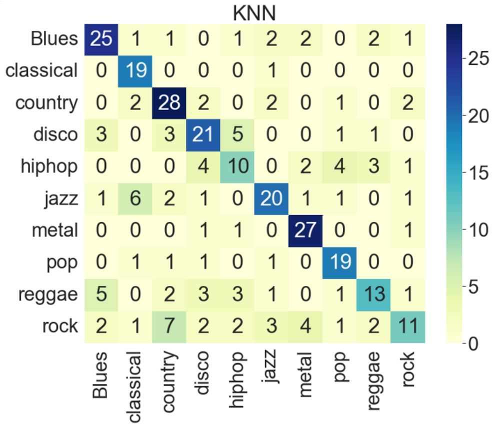
- Use this as feedback to determine if features are too dependent

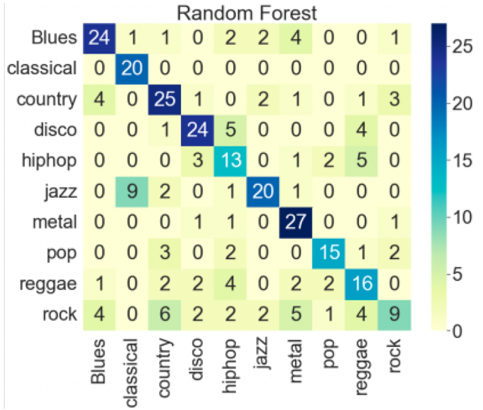
- Implement PCA

- Drop MFCCs to reduce complexity

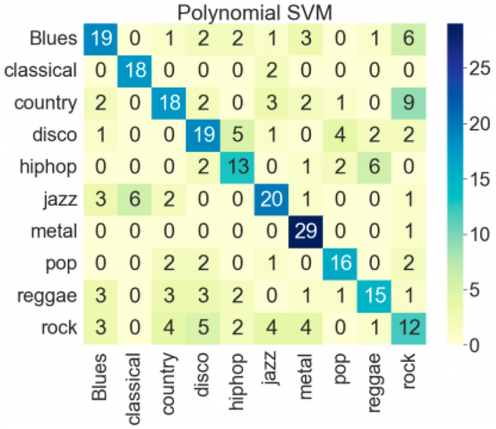


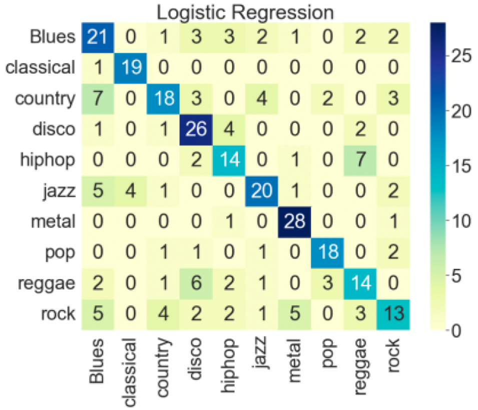
Model Confusion Matrices

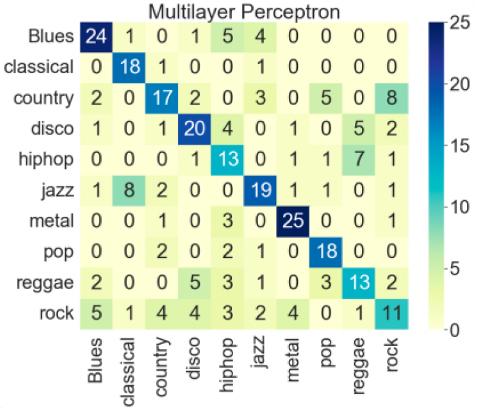




- Visualization of performance
- Misclassification of similar genres

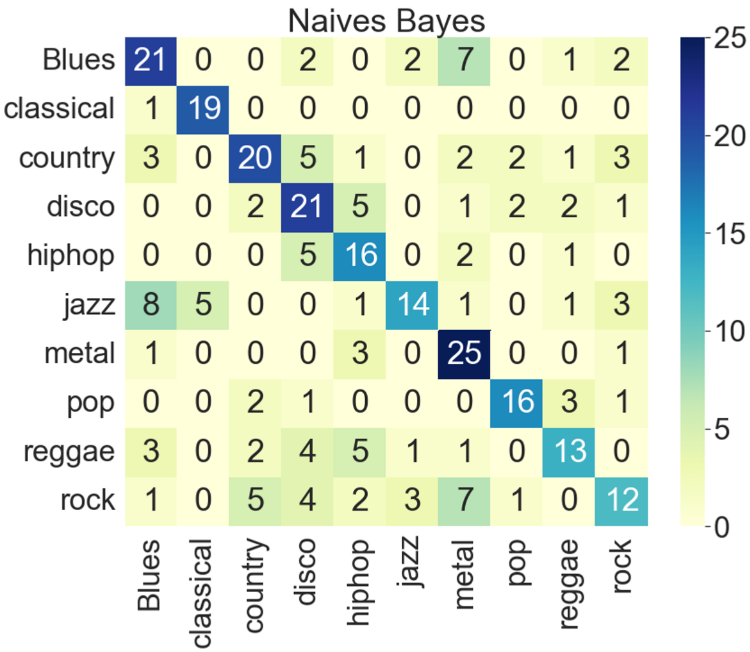


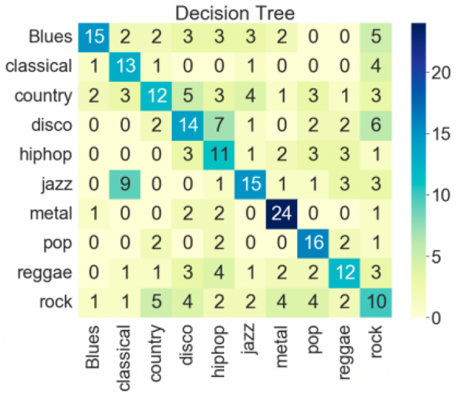




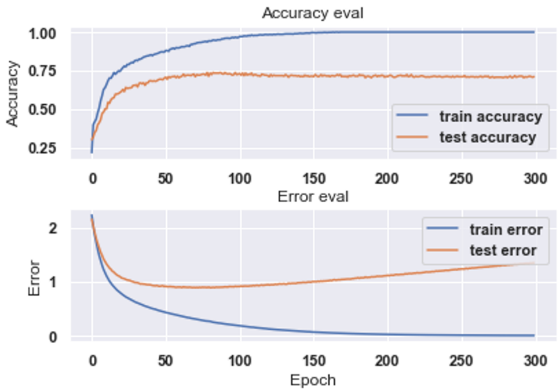
Worst Classifiers

- Genre correlation adds to error
- Naive bayes relies on independency = prediction error
- Improve by adding more specific genre classifications



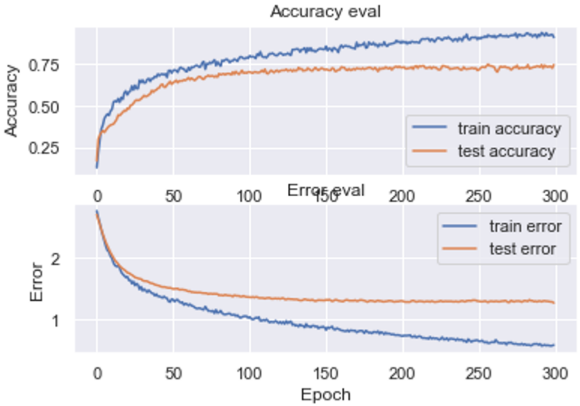


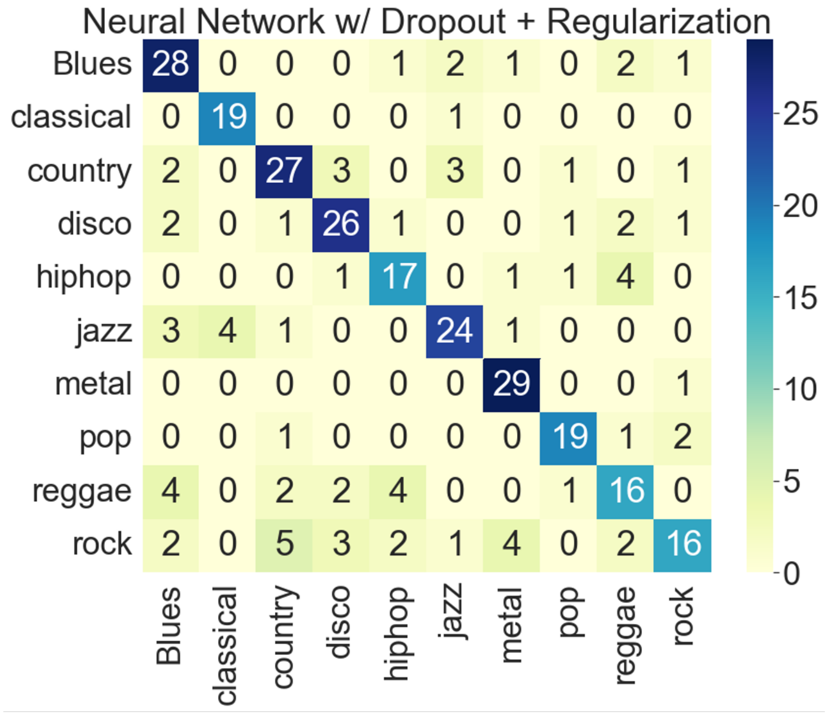
Neural Network Evaluation



Overfitting:
100% Training
~65% Testing

Dropout + Regularization:
92% Training
72% Testing





Conclusion

- Classification models were prone to overfitting due to a relatively small amount of data with high variance
- Out of all the classification algorithms, the Neural Network proved best with ***69% Test Accuracy*** and ***92% Training Accuracy***
  - Adding in Regularization, Dropout, and PCA, overfitting was reduced
- We can improve the accuracy by supplementing our model with **more data** or by **implementing CNN** with extracted spectrogram images