Predicting Album Cover Genres & Decades with Computer Vision

Katty Lein
General Assembly Capstone Project: August 2016

AGENDA

- Background
- Goal
- Process
- Dataset: Sources / EDA
- Model Building
- Findings
- Final Thoughts

BACKGROUND: WHY THIS PROJECT

BASED OFF OF PERSONAL AND PROFESSIONAL INTERESTS

Old, Weird Album Covers

Computer Vision Techniques

Xbox Kinect

Snapchat Filters

Facebook Tagging

GOAL: WHAT THIS PROJECT HOPES TO ACHIEVE

BE ABLE TO ACCURATELY PREDICT THE GENRE AND DECADE OF AN ALBUM COVER

Prototype for Website Application

Album Cover Decade/Genre Predictor

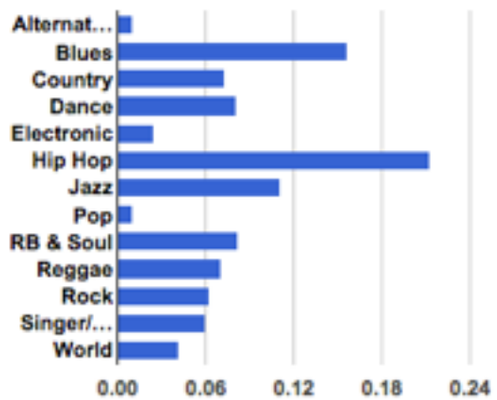
Sourcing from Wikipedia and the iTunes API, I built a model using Support Vector Machines to predict the genre and the decade of any album cover from the 1980s to the present year. This model will output probabilities for the following:

- *Decades*: 1980s to 2000, or 2000 to present Day
- *Genres* : Alternative, Blues, Country, Dance, Electronic, Hip Hop, Jazz, Pop, R&B / Soul, Reggae, Rock, Singer/Songwriter, and World

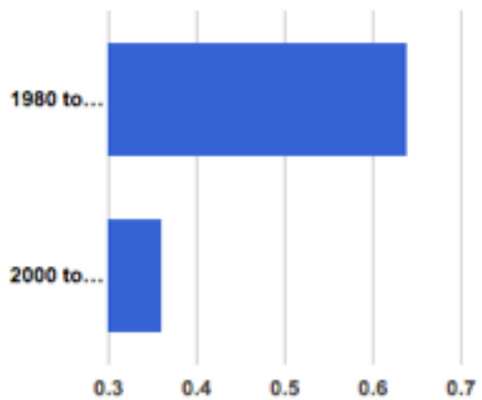
Upload Your Album Cover:

Choose File 51gZc3A+UAL.jpg

Predicted Probability of Genre



Predicted Probability of Decade



*Website currently only works on local. Please check git for more information

PROCESS: HOW THE MODEL WAS BUILT

DATA MINING
Wikipedia API + iTunes API

EDA
Seaborn Visualization
Python Image Blending

TARGET VARIABLE
Genre + Decade
Decade Only
Genre Only

FEATURE EXTRACTION
CV2 Key Point Feature Extraction
Neural Net Feature Extraction
ImageNet Deep Feature Extractor

MODEL BUILD PREP
Evenly Distributed Sample
Train-Test (80 - 20)
DB Scan Clustering
KMeans Clustering

MODEL BUILDING
XGBoost
Gradient Boosted Trees
Neural Networks
KNN
SVM

PROCESS: HOW THE FINAL MODEL WAS BUILT

DATA MINING
Wikipedia API + iTunes API

EDA
Seaborn Visualization
Python Image Blending

TARGET VARIABLE
Genre + Decade
Decade Only
Genre Only

FEATURE EXTRACTION
CV2 Key Point Feature Extraction
Neural Net Feature Extraction
ImageNet Deep Feature Extractor

MODEL BUILD PREP
Evenly Distributed Sample
Train-Test (80 - 20)
DB Scan Clustering
KMeans Clustering

MODEL BUILDING
XGBoost
Gradient Boosted Trees
Neural Networks
KNN
SVM

DATASET: SOURCES

USED WIKIPEDIA API FOR ARTISTS NAMES TO FEED INTO THE ITUNES API

Wikipedia API

Giants of jazz [edit]

Main article: Jazz royalty

Each section chronological by year of birth

Instrumentalists [edit]

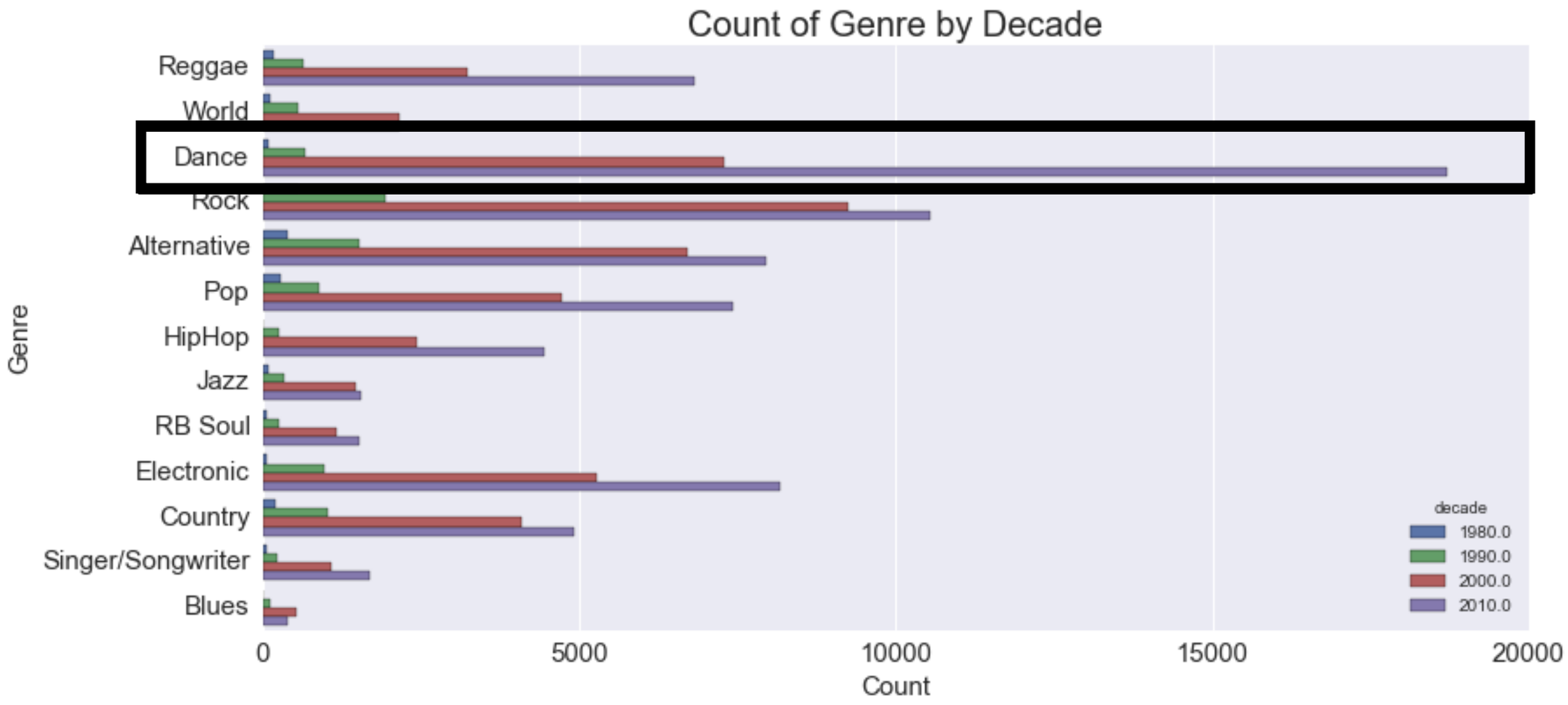
- Scott Joplin (1868–1917)
- Charles "Buddy" Bolden (1877–1931)
- Duke Ellington (1899–1974)
- Louis Armstrong (1901–1971)
- Joe Venuti (1903–1978)
- Earl Hines (1903–1983)
- Fats Waller (1904–1943)
- Count Basie (1904–1984)
- Stéphane Grappelli (1908–1997)
- Benny Goodman (1909–1986)
- Art Tatum (1909–1956)
- Sun Ra (1914–1993)
- Ward Kimball (1914–2002)
- Oliver Todd (1916-2001)
- Thelonious Monk (1917–1982)
- Dizzy Gillespie (1917–1993)

iTunes API


```
{"wrapperType":"collection", "collectionType":"Album", "artistId":4038534,
"collectionId":318145456, "amgArtistId":29861, "artistName":"Admiral Bailey",
"collectionName":"The Best of Admiral Bailey", "collectionCensoredName":"The
Best of Admiral Bailey", "artistViewUrl":"https://itunes.apple.com/us/artist/
admiral-bailey/id4038534?uo=4", "collectionViewUrl":"https://itunes.apple.com/
us/album/the-best-of-admiral-bailey/id318145456?uo=4", "artworkUrl60":"http://
is5.mzstatic.com/image/thumb/Music/v4/03/5f/
94/035f9477-5386-5470-5a40-82ade1d61959/source/60x60bb.jpg",
"artworkUrl100":"http://is5.mzstatic.com/image/thumb/Music/v4/03/5f/
94/035f9477-5386-5470-5a40-82ade1d61959/source/100x100bb.jpg",
"collectionPrice":11.99, "collectionExplicitness":"notExplicit", "trackCount":
20, "copyright":"℗ 1987 King Jammys", "country":"USA", "currency":"USD",
"releaseDate":"1987-06-09T07:00:00Z", "primaryGenreName":"Reggae"},
{"wrapperType":"collection", "collectionType":"Album", "artistId":4038534,
"collectionId":512983830, "amgArtistId":29861, "artistName":"Admiral Bailey",
"collectionName":"Big Belly", "collectionCensoredName":"Big Belly",
"artistViewUrl":"https://itunes.apple.com/us/artist/admiral-bailey/id4038534?
uo=4", "collectionViewUrl":"https://itunes.apple.com/us/album/big-belly/
id512983830?uo=4", "artworkUrl60":"http://is3.mzstatic.com/image/thumb/Music/
v4/09/05/fd/0905fd80-3674-1649-f774-2ede909cfbc5/source/60x60bb.jpg",
"artworkUrl100":"http://is3.mzstatic.com/image/thumb/Music/v4/09/05/fd/
0905fd80-3674-1649-f774-2ede909cfbc5/source/100x100bb.jpg", "collectionPrice":
7.99, "collectionExplicitness":"notExplicit", "trackCount":14, "copyright":"℗
2012 Jammys", "country":"USA", "currency":"USD",
"releaseDate":"1988-01-01T08:00:00Z", "primaryGenreName":"Reggae"},
{"wrapperType":"collection", "collectionType":"Album", "artistId":4038534,
"collectionId":512976769, "amgArtistId":29861, "artistName":"Admiral Bailey",
"collectionName":"Best of Admiral Bailey", "collectionCensoredName":"Best of
Admiral Bailey", "artistViewUrl":"https://itunes.apple.com/us/artist/admiral-
bailey/id4038534?uo=4", "collectionViewUrl":"https://itunes.apple.com/us/album/
best-of-admiral-bailey/id512976769?uo=4", "artworkUrl60":"http://
is4.mzstatic.com/image/thumb/Music/v4/37/c7/75/37c775ac-1a70-
ae09-7ec7-3ae42db01774/source/60x60bb.jpg", "artworkUrl100":"http://
is4.mzstatic.com/image/thumb/Music/v4/37/c7/75/37c775ac-1a70-
ae09-7ec7-3ae42db01774/source/100x100bb.jpg", "collectionPrice":7.99,
"collectionExplicitness":"notExplicit", "trackCount":20, "copyright":"℗ 2012
Jammys", "country":"USA", "currency":"USD",
"releaseDate":"1987-06-09T07:00:00Z", "primaryGenreName":"Reggae"},
```

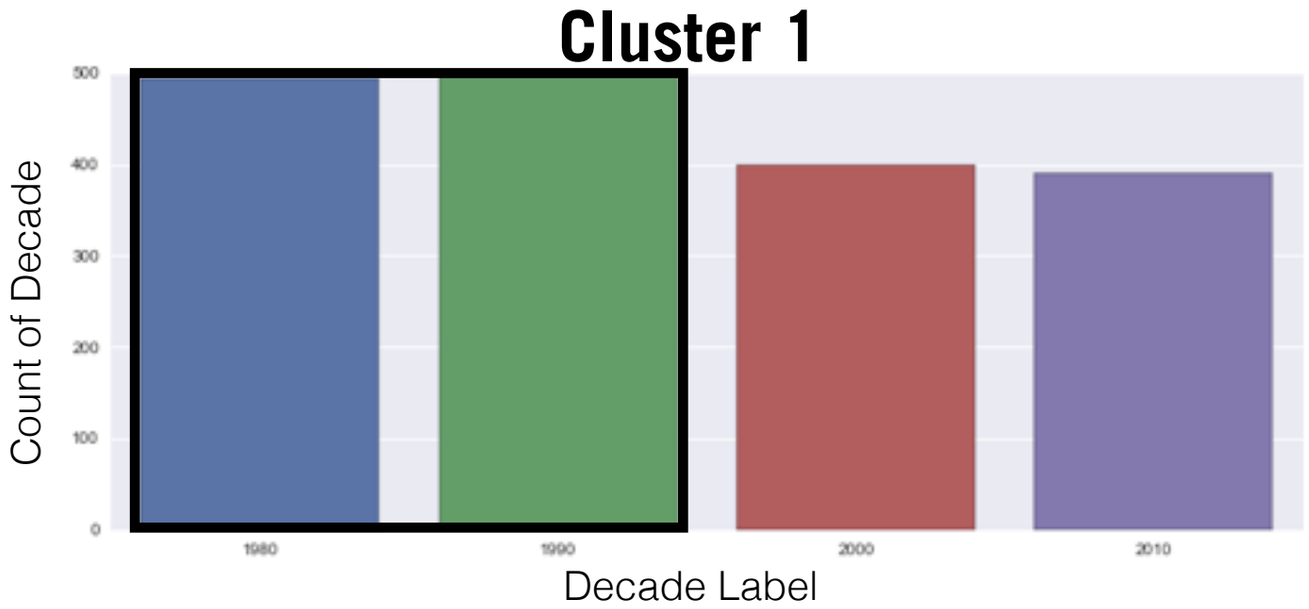
DATASET: EDA

SOME DECADES AND GENRES WERE OVER-REPRESENTED COMPARED TO OTHERS

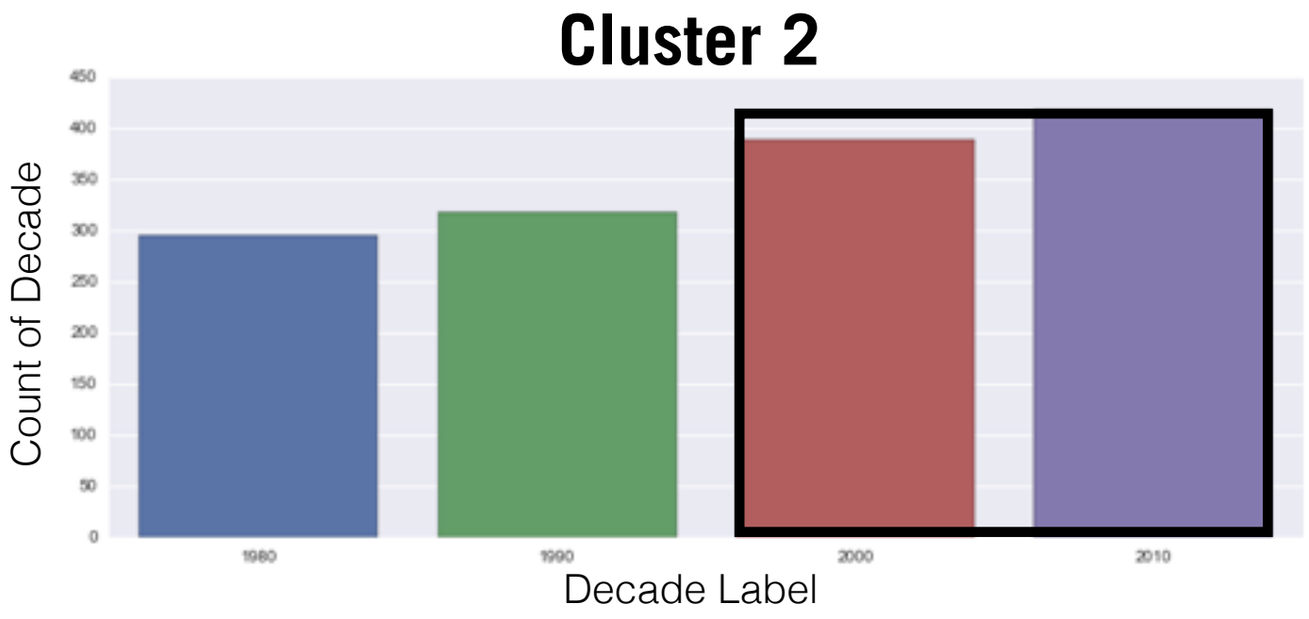


DATASET: EDA

THROUGH CLUSTERING, THE DECADES FELL INTO TWO GROUPS OF PRE 2000 AND POST 2000



Cluster 1 had an over-representation of album covers from the 1980s and 1990s

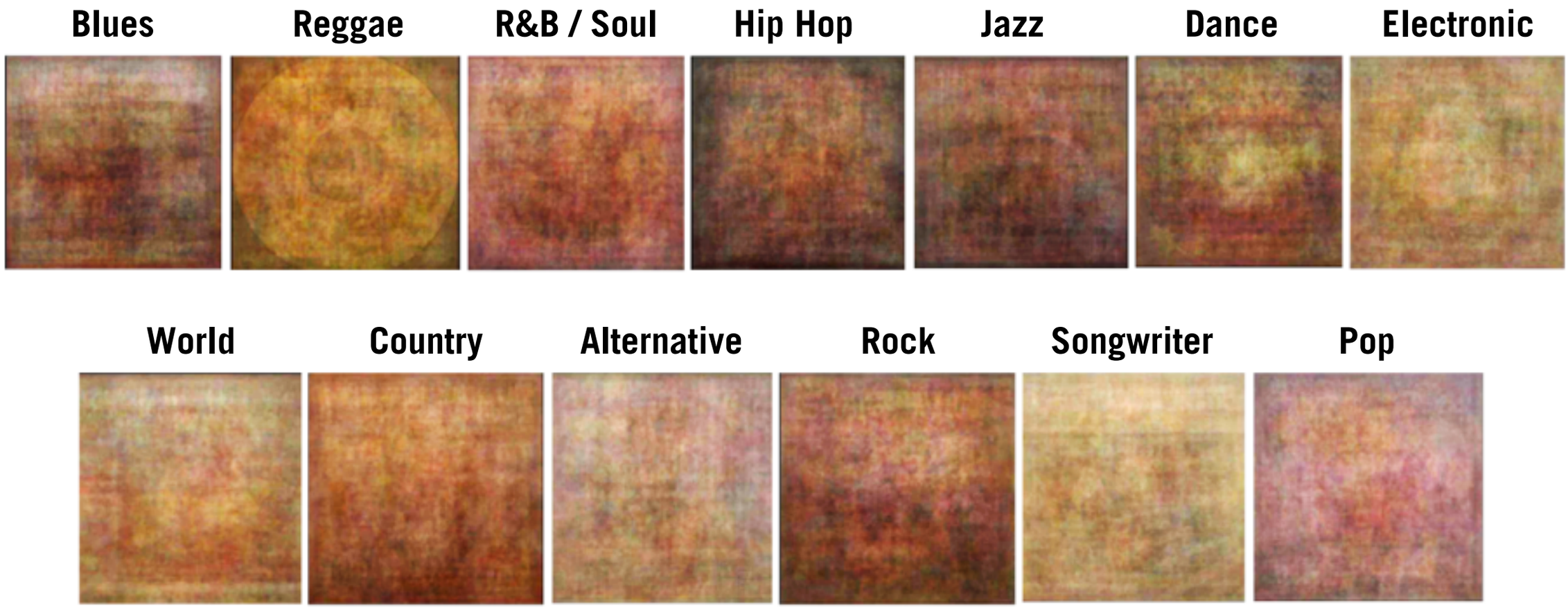


Cluster 2 had an over-representation of album covers from the 2000s and 2010s

Clusters were determined by KMeans analysis with k=2 on train set only. Directional only, since silhouette score was .03

DATASET: EDA

QUALITATIVELY, GENRES SHOWED DIFFERENCES IN COMPOSITION



250 randomly sampled album covers per genre. For illustrative purposes only.

DATASET: EDA

DIFFERENCES WERE ALSO APPARENT BY DECADE

1980s - 2000

2000 - Present

250 randomly sampled album covers per time period. For illustrative purposes only.

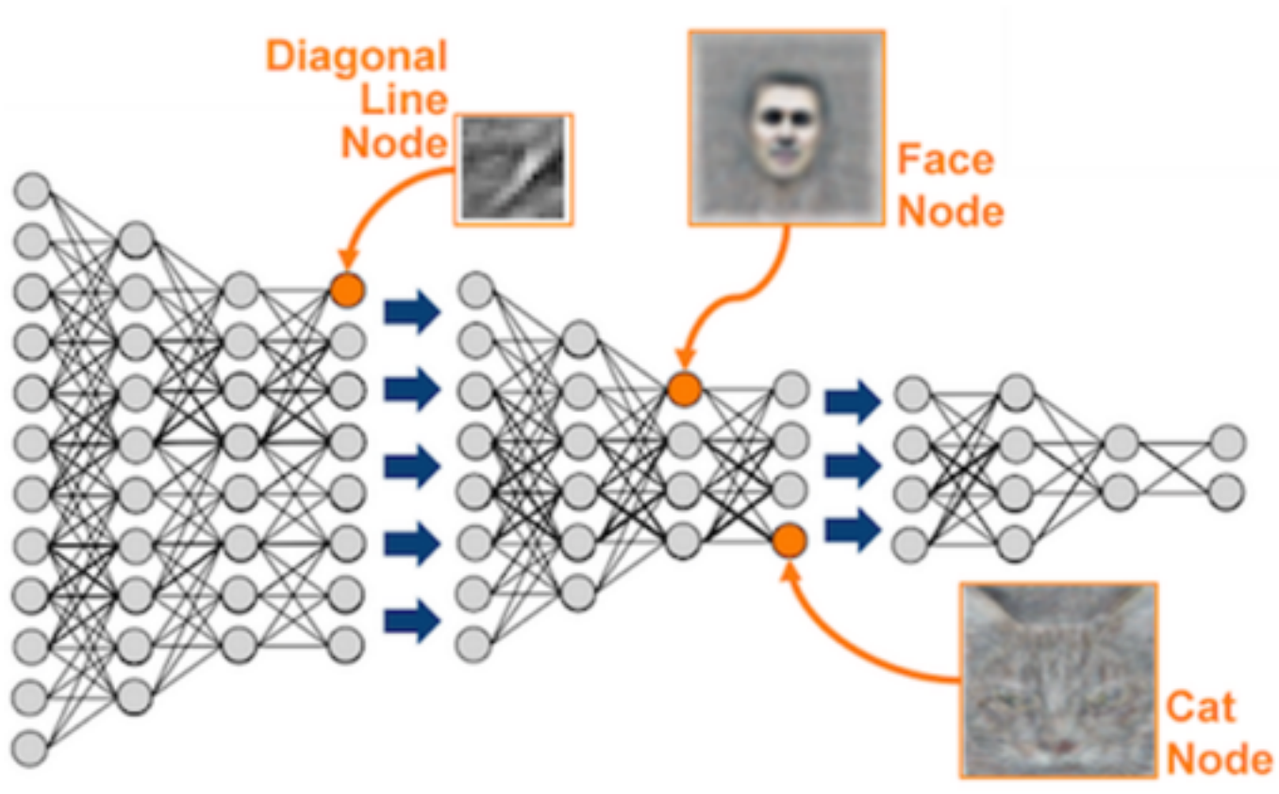
MODEL BUILDING: THE FINAL 2 MODELS

FINAL MODELS WERE BUILT USING 2,000 PER DECADE GROUP AND 800 PER GENRE

	GENRE	DECADE
TARGET	13 Genres (Reggae, World, Dance, Rock, Alternative, Pop, Hip Hop, Jazz, R&B / Soul, Electronic, Country, Blues, Singer/Songwriter)	2 Decade Groups (1980s to 2000s, 2000s to Present)
FEATURES	ImageNet Deep Feature Extractor	ImageNet Deep Feature Extractor
MODEL	One vs Rest Support Vector Machine Polynomial Kernel	One vs Rest Support Vector Machine Polynomial Kernel

MODEL BUILDING: FEATURE EXTRACTION

UTILIZED GRAPHLAB'S BUILT IN NEURAL NETWORK FEATURE EXTRACTOR



- ImageNet is a convolutional neural network trained on ~1 million images and 1,000 classes
- This trained model can be used to extract features from other image data sets

More information on Graphlab's ImageNet found here:
http://on-demand.gputechconf.com/gtc/2015/presentation/S5630-Piotr-Teterwak.pdf

MODEL BUILDING: FEATURE EXTRACTION

DEEP FEATURE EXTRACTION WAS APPLIED TO THE ALBUM COVERS DATASET

INPUT

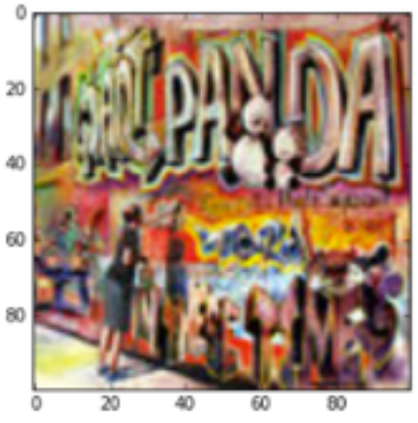


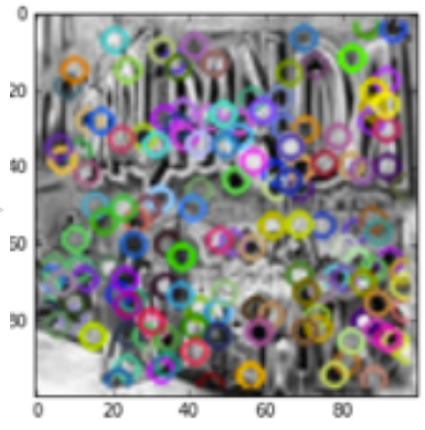
Image is read in…

PROCESSING

```
[[[255, 255, 255], [250, 250,
[[[39, 27, 85], [20, 21, 41],
[[[128, 202, 244], [125, 199,
[[[67, 28, 26], [54, 15, 13],
[[[155, 186, 185], [156, 207,
```

…converted into pixels…

PROCESSING



…fed into ImageNet deep feature extractor, to identify key points…

FINAL FEATURES

4090	4091	4092	4093	4094	4095
0.000000	0.0	0.000000	0.0	3.36760	0.0
0.000000	0.0	0.066012	0.0	0.00000	0.0
0.000000	0.0	1.279530	0.0	1.60196	0.0
0.244365	0.0	0.000000	0.0	0.00000	0.0
0.000000	0.0	0.000000	0.0	1.16809	0.0

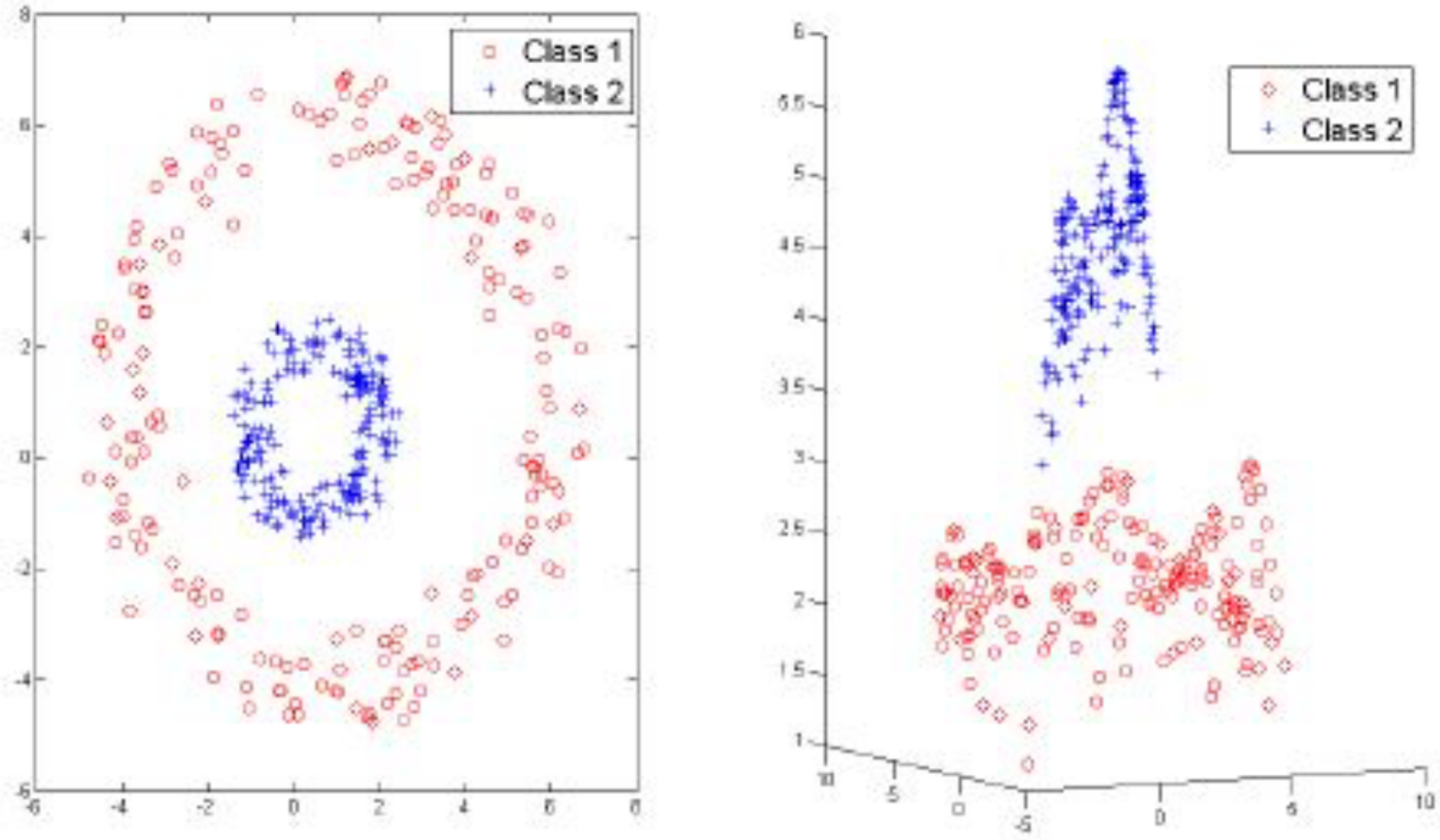
…resulting in 4,096 individual features

Example used for illustrative purposes only - this utilizes SIFT features of CV2.

MODEL BUILDING: THE MODEL

POLYNOMIAL SVC WAS CHOSEN FOR ITS ABILITY TO SPATIALLY CLASSIFY

Illustrative Example

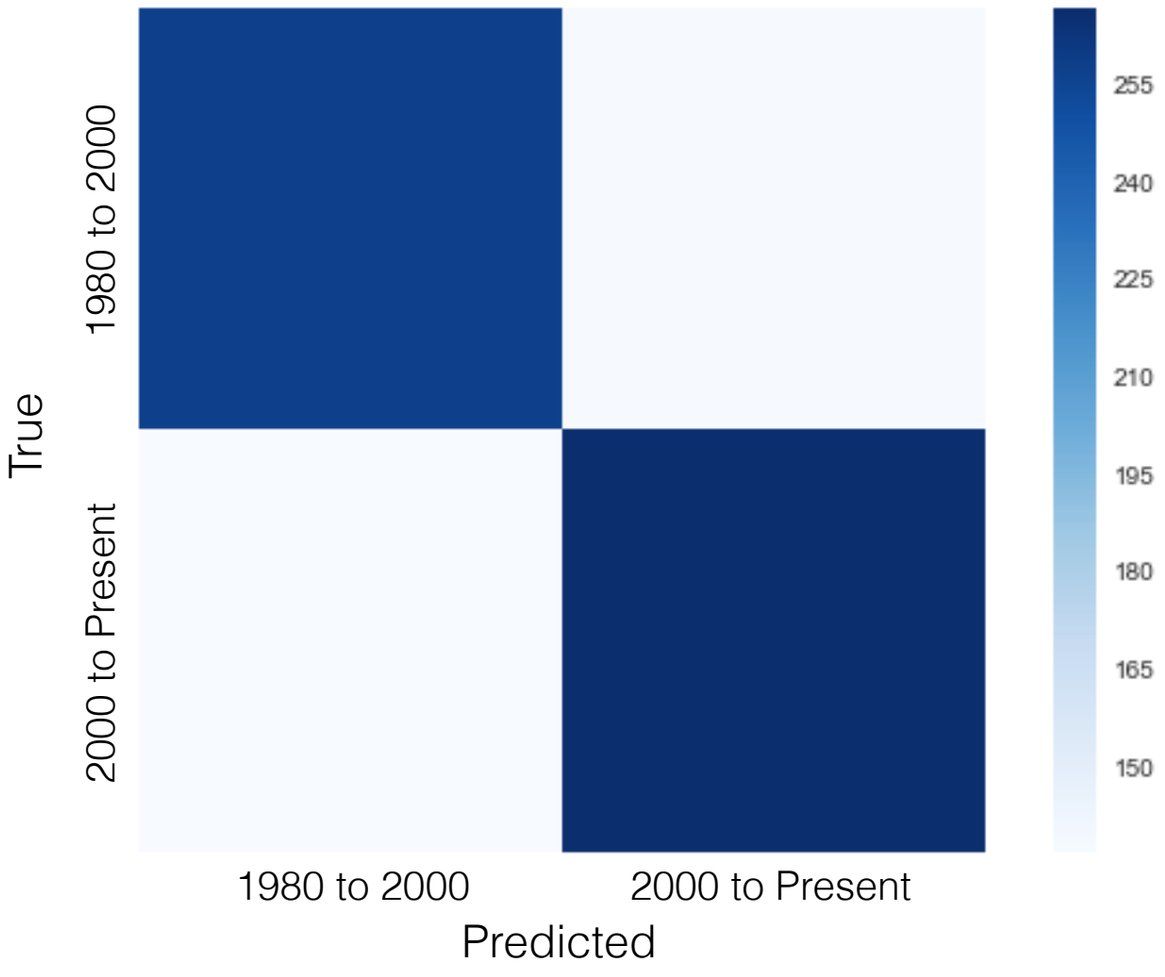


Example used for illustrative purposes only

FINDINGS: DECADE MODEL

RELATIVELY SUCCESSFUL, BEATING OUT BASELINE

Confusion Matrix



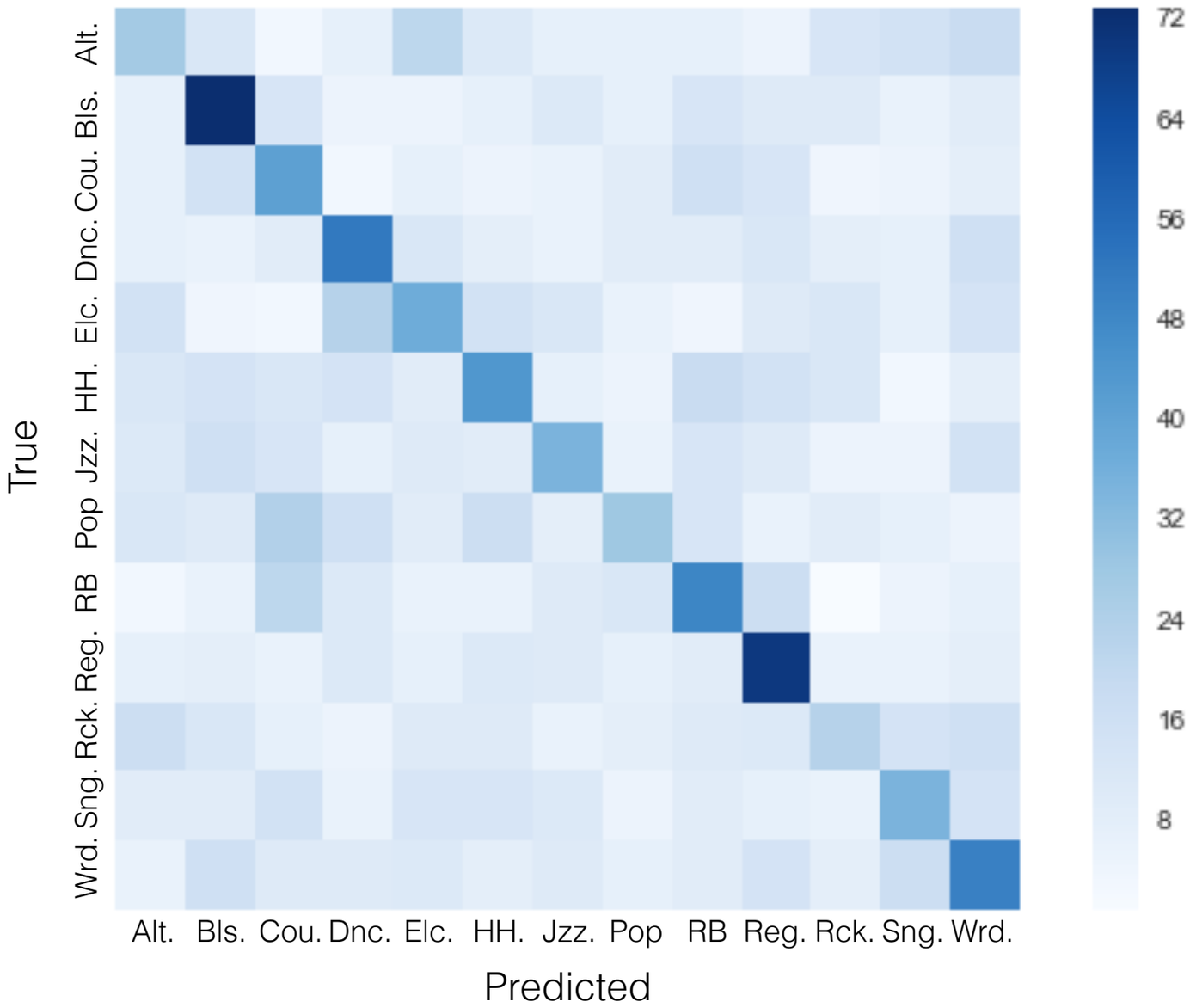
Scores

Baseline:	.50
Average Accuracy:	**.64**
Average F-1:	**.66**

FINDINGS: GENRE MODEL

GENERALLY SUCCESSFUL, 3X BETTER THAN BASELINE

Confusion Matrix

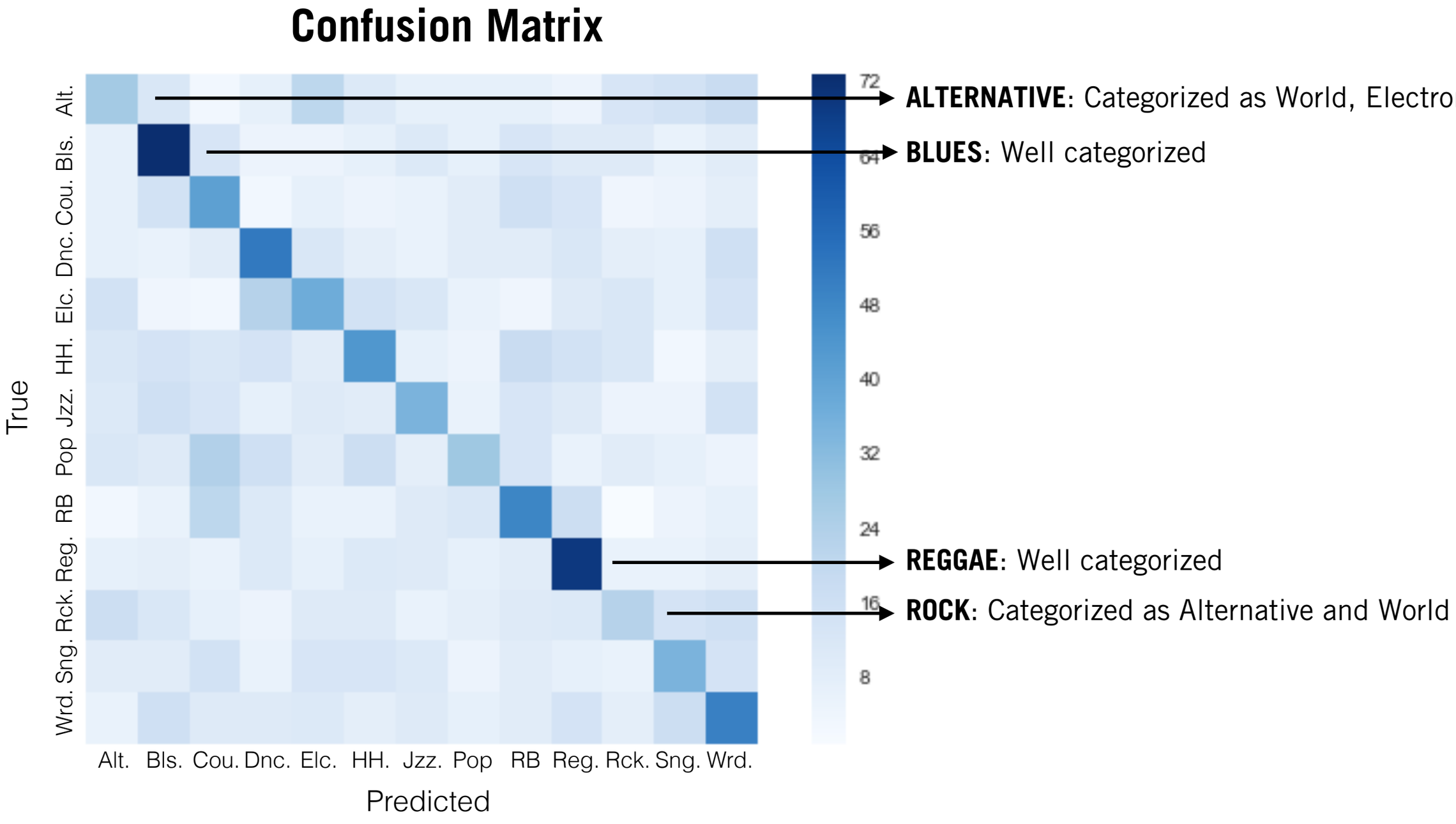


Scores

Baseline:	.08
Average Accuracy:	**.27**
Average F-1:	**.27**

FINDINGS: GENRE MODEL DEEP DIVE

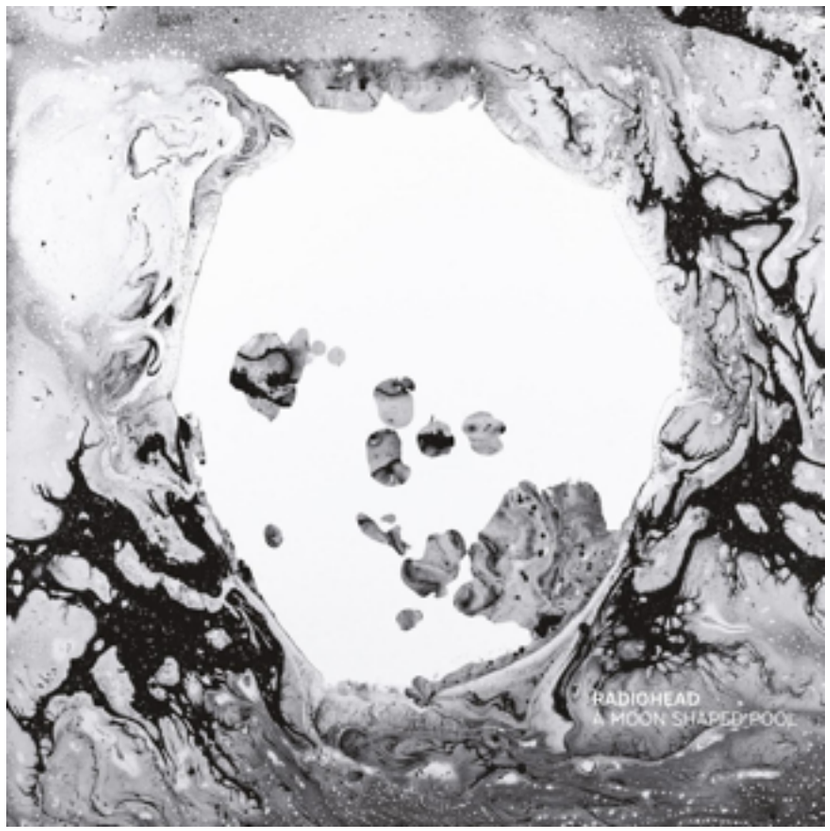
CERTAIN GENRES ARE MORE LIKELY TO BLEND INTO OTHERS



FINDINGS: SAMPLE TEST WITH RADIOHEAD

MODEL PERFORMED WELL WITH EXTREME TIME PERIOD

LABELS: 2016, Alternative



PREDICTED GENRES

Alternative:	**.162**
Singer/Songwriter:	.130
Jazz:	.109
Rock:	.103
Electronic:	.102
Hip Hop:	.076
Country:	.075
Blues:	.048
Pop:	.046
R&B / Soul:	.045
World:	.045
Dance:	.033
Reggae:	.026

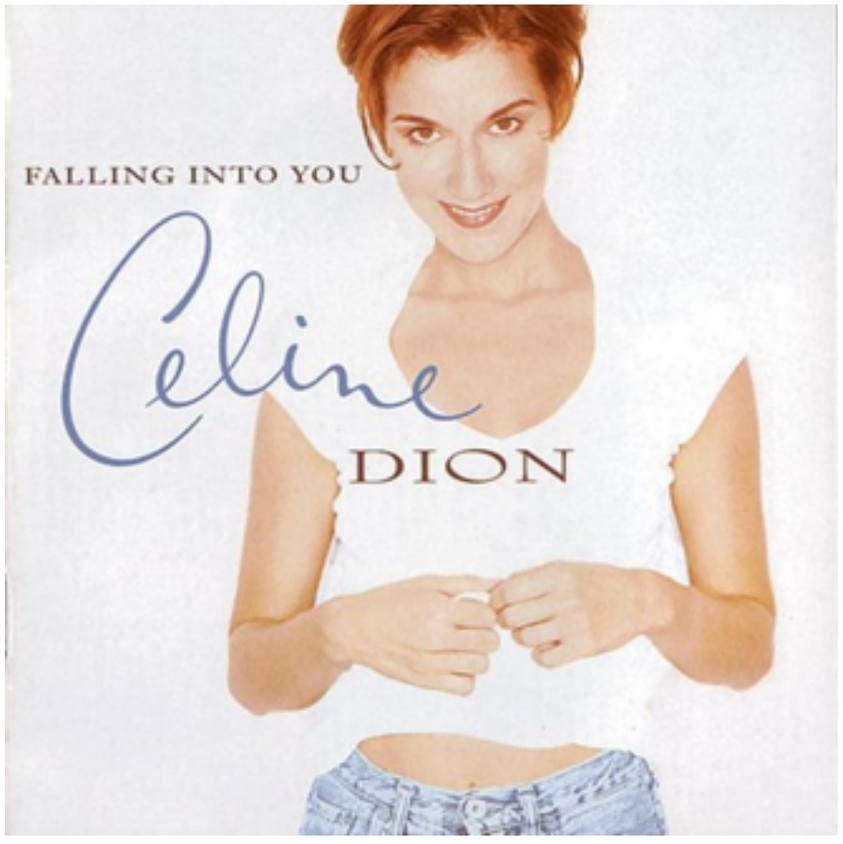
PREDICTED DECADES

2000 - Present:	**.66**
1980 - 2000:	.33

FINDINGS: SAMPLE TEST WITH CELINE DION

WITH A TIME PERIOD CLOSE TO THE CUSP, THE MODEL DID NOT PERFORM AS WELL

LABELS: 1996, Pop



PREDICTED GENRES

Pop:	**.126**
Singer/Songwriter:	.094
Alternative:	.091
Hip Hop:	.080
Rock:	.080
World:	.080
Dance:	.076
Electronic:	.071
Jazz:	.062
Country:	.061
Reggae:	.050
Blues:	.041
R&B / Soul:	.010

PREDICTED DECADES

2000 - Present:	**.56**
1980 - 2000:	.44

FINDINGS: SAMPLE TEST WITH EDDIE MURPHY

MODEL PERFORMED WELL WITH ANOTHER EXTREME TIME PERIOD, BUT NOT SO WELL WITH GENRE

LABELS: 1985, Pop

PREDICTED GENRES

R&B / Soul:	**.225**
Jazz:	.093
Dance:	.092
Alternative:	.066
Blues:	.078
Electronic:	.078
Singer/Songwriter:	.065
Pop:	.061
Hip Hop:	.057
Reggae:	.050
Rock:	.047
World:	.046
Country:	.040

PREDICTED DECADES

1980 - 2000:	**.80**
2000 - Present:	.20

FINDINGS: SAMPLE TEST WITH DOGE

DIRECTIONALLY MISLABELLED?

LABELS: 2013, ???

PREDICTED GENRES

Alternative:	**.120**
Rock:	.089
Singer/Songwriter:	.088
Pop:	.085
Jazz:	.080
Electronic:	.077
Hip Hop:	.074
R&B / Soul:	.073
Dance:	.070
Country:	.064
Blues:	.063
World:	.063
Reggae:	.056

PREDICTED DECADES

1980 - 2000:	**.61**
2000 - Present:	.39

FINAL THOUGHTS

- First attempt at learning neural nets and image classification models was relatively successful but could be improved upon

- Genres are relative - ways to amend this could be to increase sample so that target variable could be genre + decade vs individual genres, and individual decade

- Although deep feature extraction yielded results that were better than baseline, features were difficult to engineer and optimize - could potentially add in other features in the future (text extraction, etc.)

- For detailed information, including website development using Flask, please see git: https://github.com/yoyoyokatty/DSI-projects/tree/master/Capstone_AlbumCovers