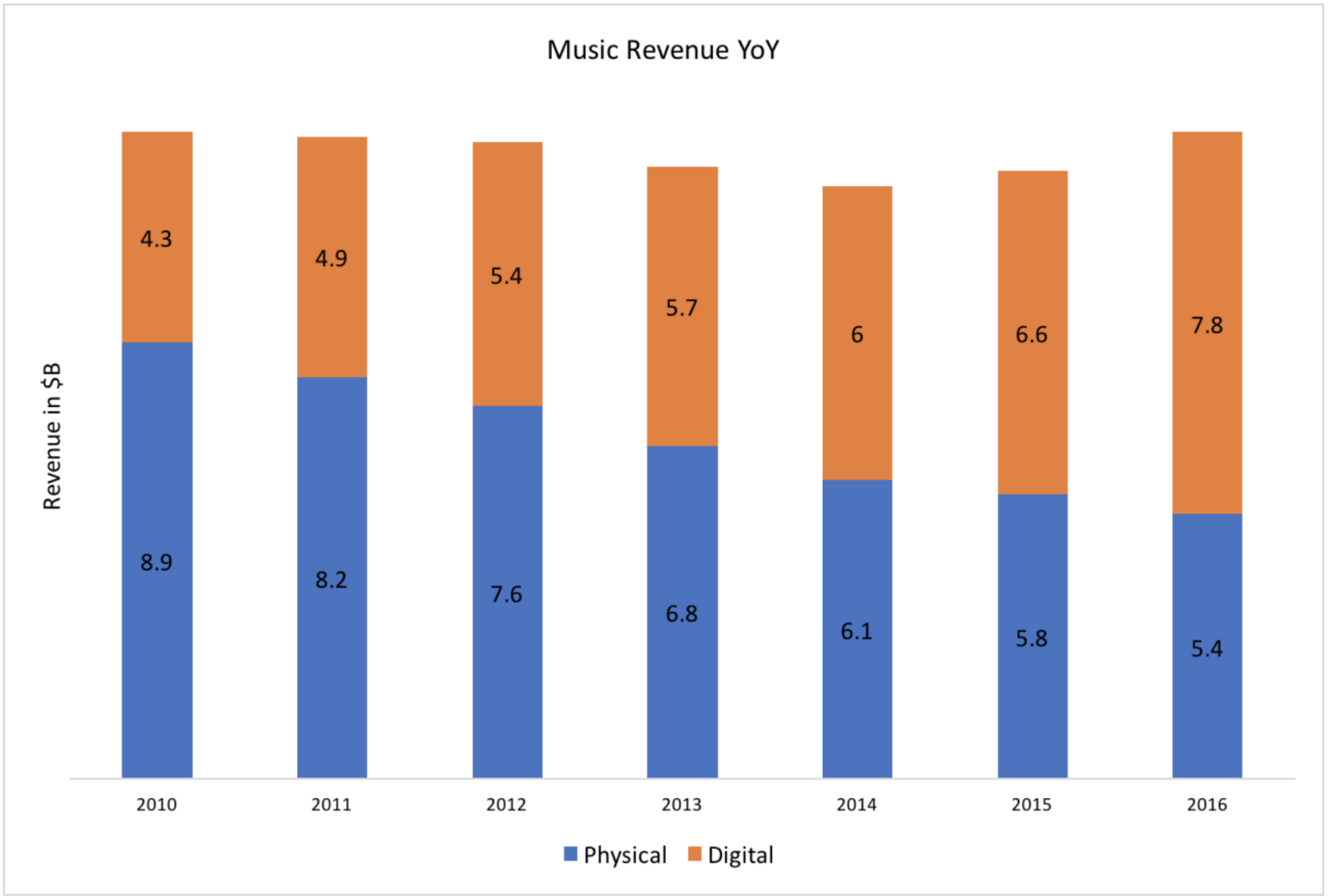
MUSIC Classification

Scale and Growth of Digital Music Industry

- YouTube has ~1.5B monthly active users and a revenue of 40B
- Digital Music revenue(subscription+ streaming+ Download) has grown 10% YoY
- Physical Music revenue has dropped 10% YoY


Music Revenue YoY
Revenue in $B
4.3
4.9
5.4
5.7
6
6.6
7.8
8.9
8.2
7.6
6.8
6.1
5.8
5.4
2010
2011
2012
2013
2014
2015
2016
Physical
Digital

Music Consuption has become predominantly Digital

- Digital Music content accounts for >50% of music revenues
- For the first time, we have the Data to understand user behaviour
- Understanding user behaviour is a key to growth of music platforms

Music Recommendation

1. Collaborative Filtering of User behaviour

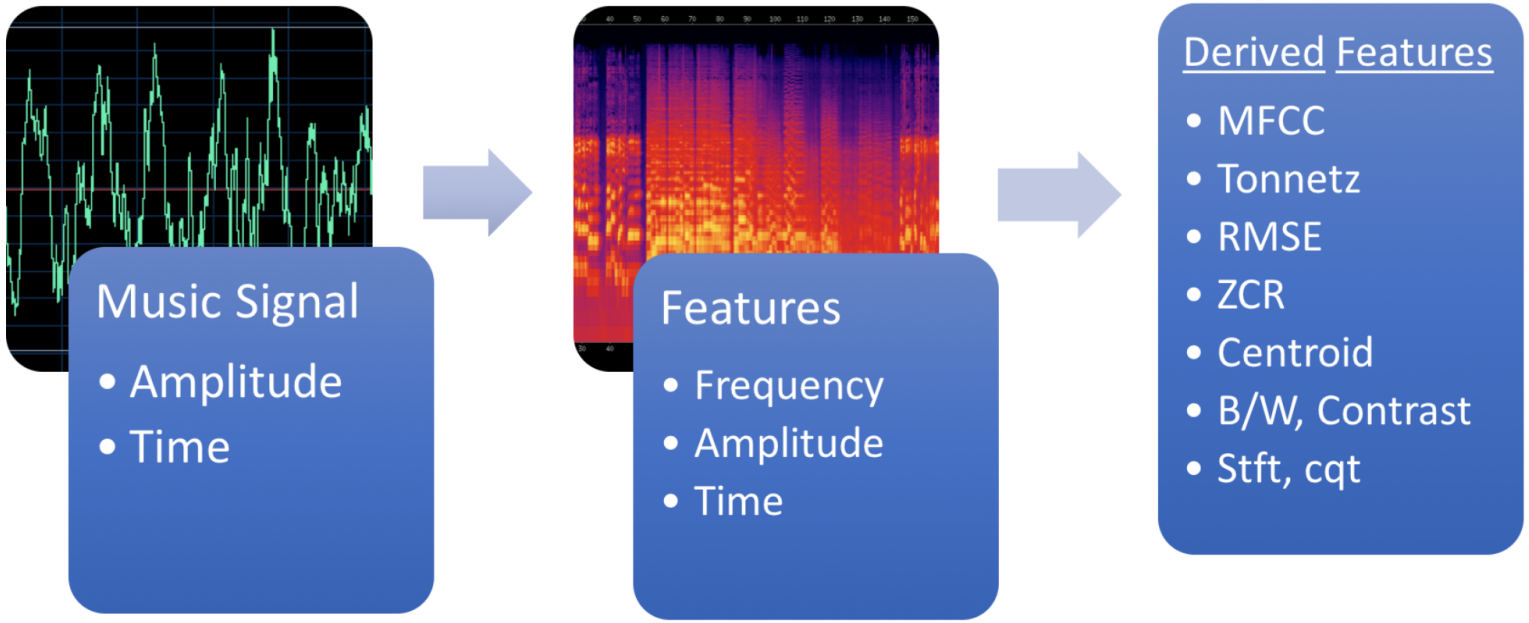
- Agnostic of Music content
- Accurate with large datasets

2. Understanding and modeling using Music content

- Understand the human perception of music
- Build recommendation engine on top of that
- **Music classification is the key**

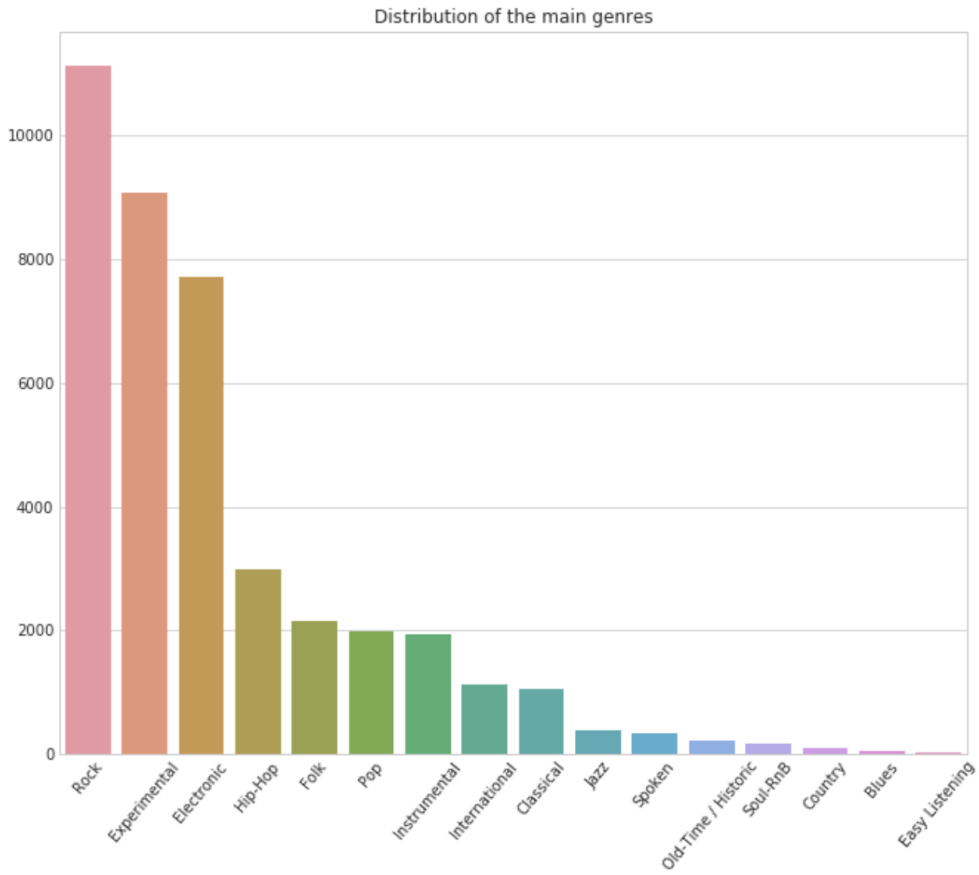
Goal: Music Classification

Features

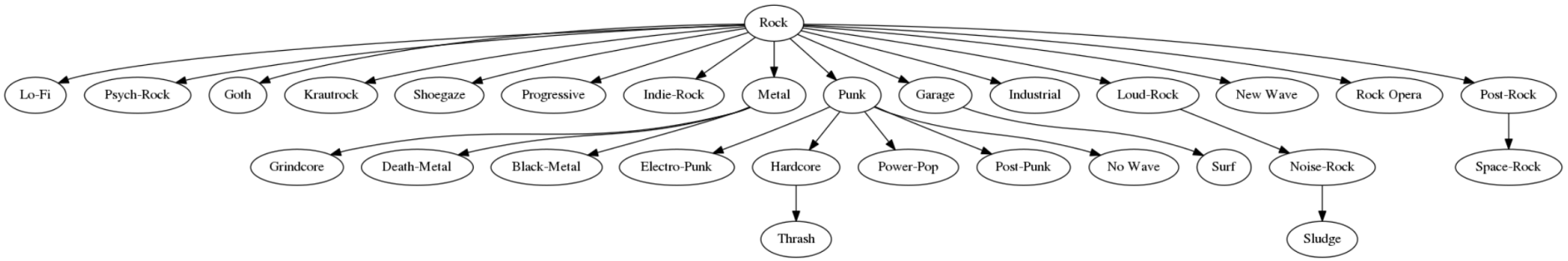


- 7 sub_features -> Mean, Median, Min, Max, Std.Dev, Skew and Kurtosis

Dataset: 91213 songs, 163 genres , 15 main genres, 522 features



Genre and Sub-Genre



Methodology

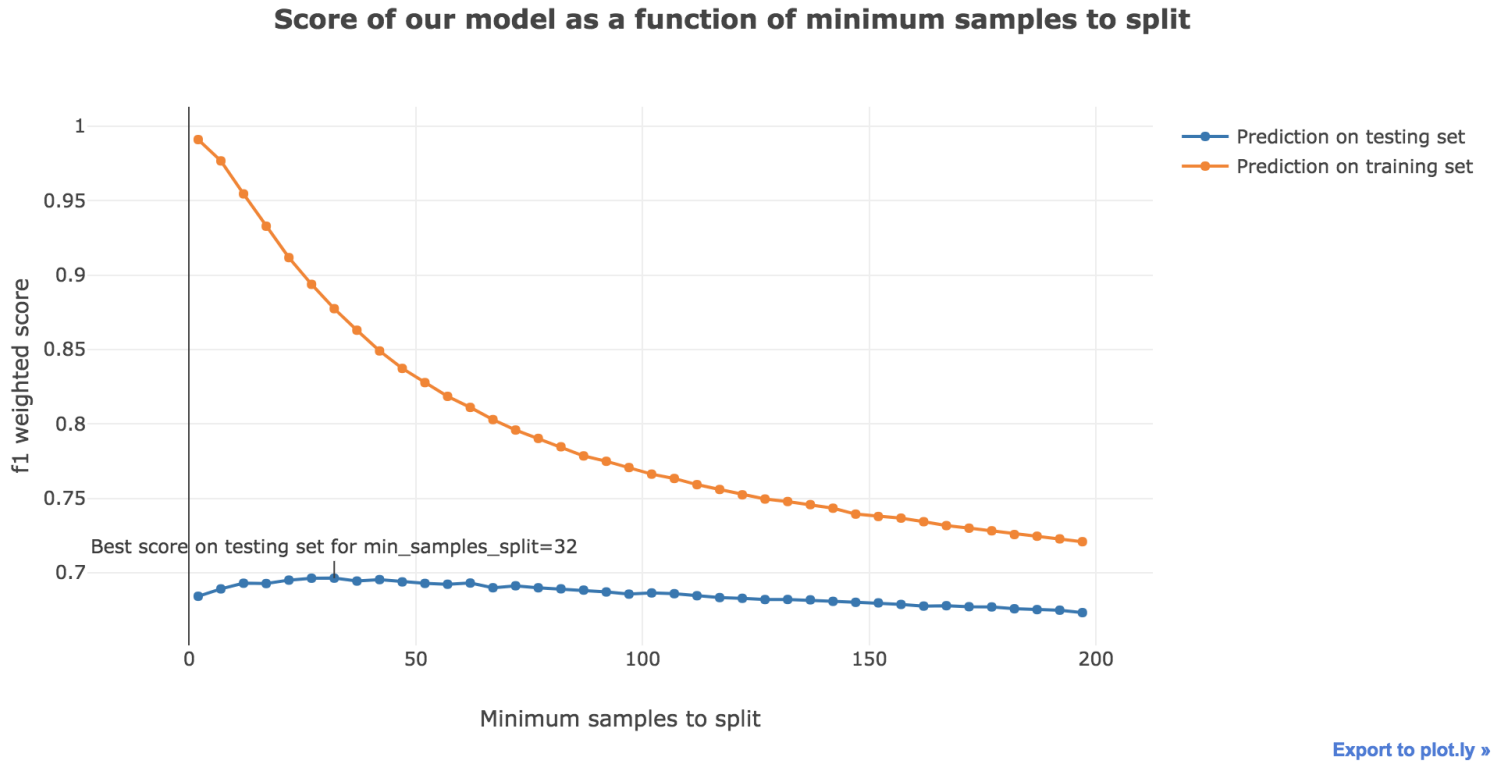
We have used five approaches to predict Genres:

- 1.1 KNN
- 1.2 Random Forest
- 1.3 Gradient Boosting
- 1.4 Neural Network
- 2.0 Multilabel Classification

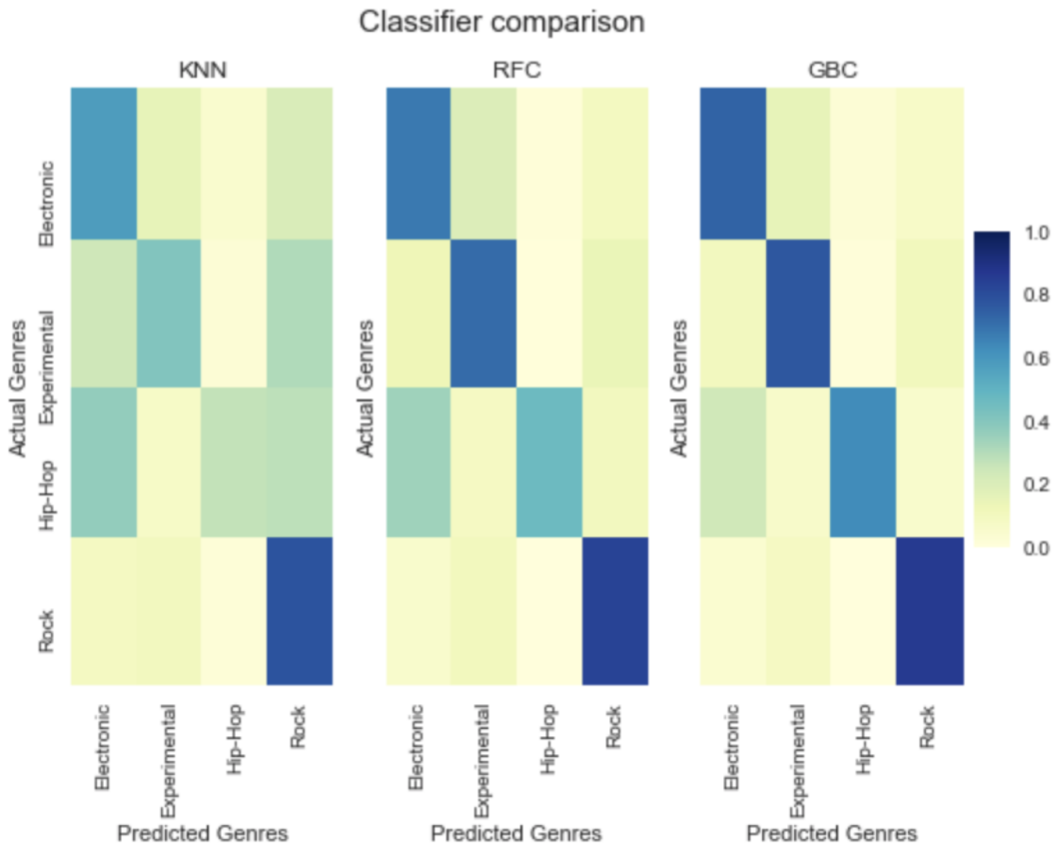
KNN, Random Forest, Gradient Boost

In [7]:
```
fig = Figure(data=data, layout=layout)
plotly.offline.iplot(fig)
```

In [5]:
```
fig = Figure(data=data, layout=layout)
plotly.offline.iplot(fig)
```

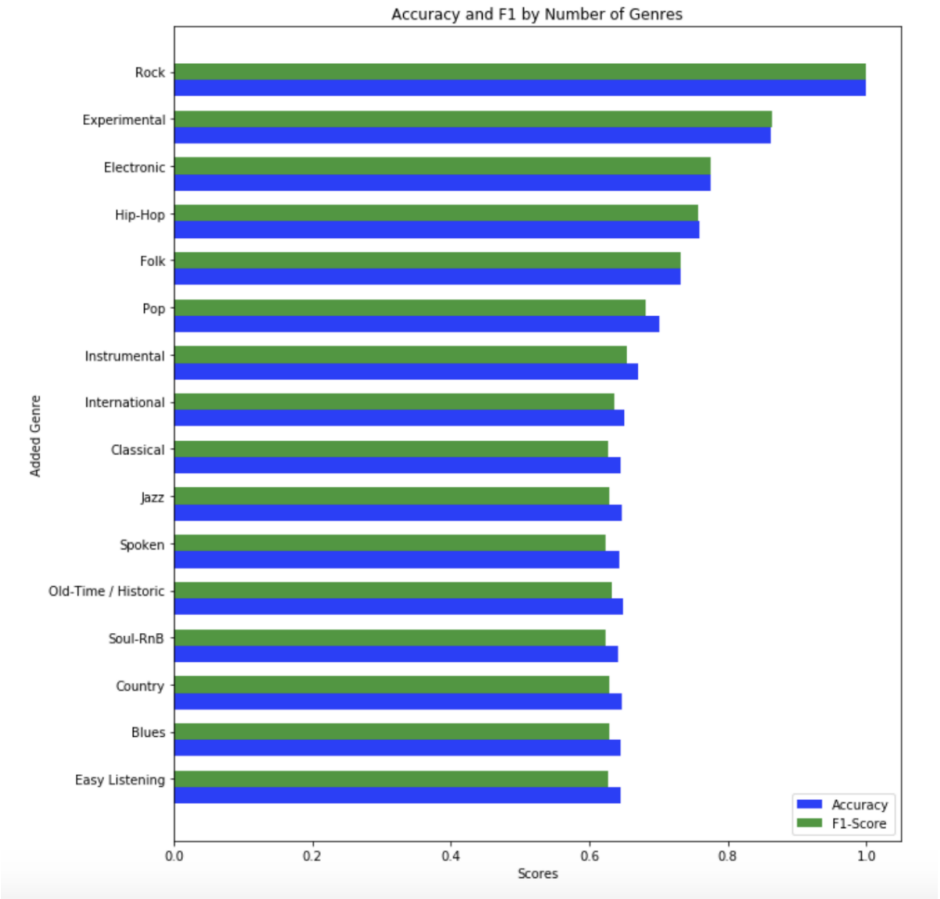


Results: KNN, Random Forest, Gradient Boost



Neural Networks

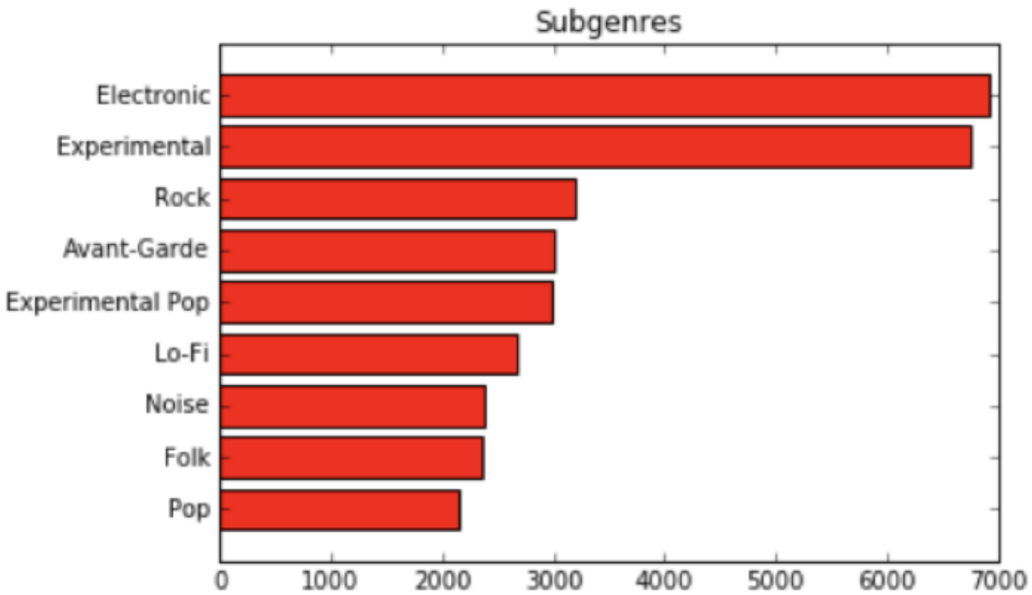
Accuracy


Accuracy and F1 by Number of Genres
Rock
Experimental
Electronic
Hip-Hop
Folk
Pop
Instrumental
International
Classical
Jazz
Spoken
Old-Time / Historic
Soul-RnB
Country
Blues
Easy Listening
Added Genre
Accuracy
F1-Score
0.0
0.2
0.4
0.6
0.8
1.0
Scores

Multilabel classification

Multilabel classification

- One vs Rest classification
- Chain classification



Chain Classification

The order in which the y's are predicted impacts overall accuracy of the prediction. We tried three different approaches:

- sorting the y's by increasing number of combinations they appear in
- sorting the y's by decreasing number of combinations they appear in
- randomly sorting the y's

There is no significant difference in accuracy between these three options on our data!

Results

	F1_score
KNN_OneVsRest	0.46
KNN_Chain	0.46
RF_OneVsRest	0.31
RF_Chain	0.33
NN_OneVsRest	0.41
NN_Chain	0.44