

Will it Chart?

Identifying whether a song will be popular

Lance Wong, Chance Woods Consulting

Context

- An A&R division of a major music label wants to employ data science to discover and sign new talent
- As a preliminary search, they do not want to miss out on the “next big thing”, but are okay with finding songs and artists that seem promising (they will vet these out separately in a second pass)
- Like all music labels, they are concerned about music sales/streams and have established a benchmark of whether a song has potential to be on the Billboard Hot 100 as criteria for the “next big thing”

Assumptions/Caveats

- Evaluate whether a song is popular based on a song metadata and artist genre
 - Avoid number of followers an artist has to decrease bias towards already popular artists
- Binary Classification:
 - Class 0 - “Will not chart”
 - **Class 1 - “Will chart”** (main focus)
 - **Optimizing model for Class 1 recall**

Data Sources Leveraged

Songfacts



Songs in Songfacts from
1/1/10-5/6/18
($n \sim 16,000$)



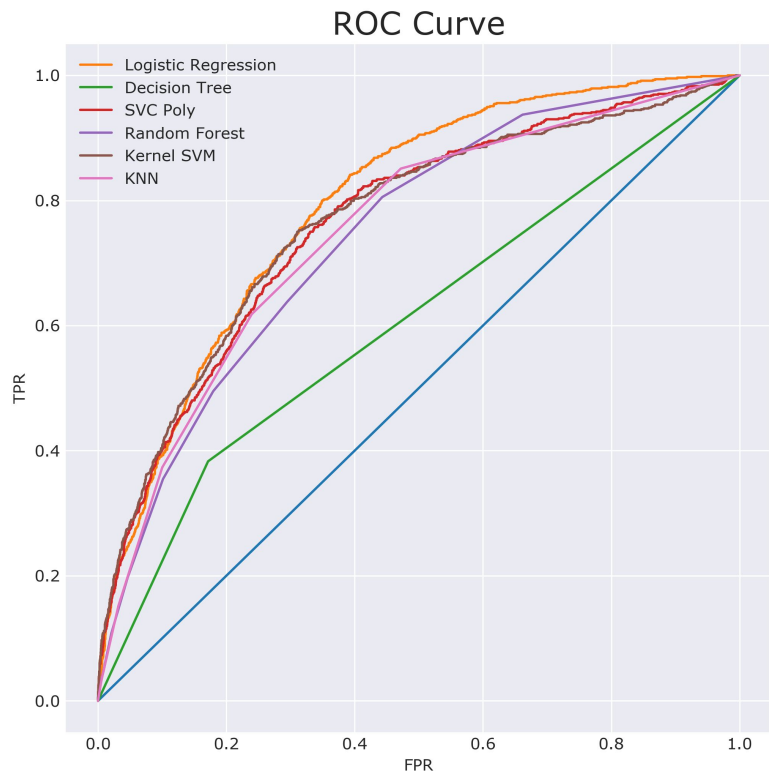
Spotify API
Song Metadata

billboard
THE HOT
100

Songs that appear on the
Billboard Hot 100
 $n \sim 3,000$

Imbalanced Classes (Class 0 : Class 1= 4:1)

Logistic Regression outperforms most models

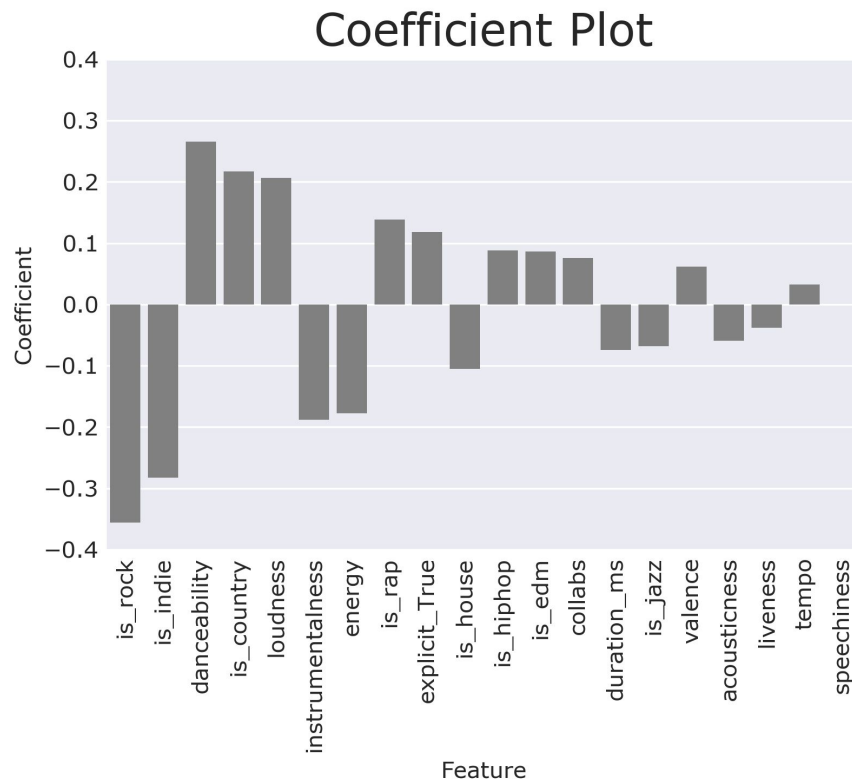


Model	Train Accuracy	Test Accuracy	Precision	Recall	AUC
Logistic Regression	0.809759	0.816412	0.609836	0.229064	0.792535
Decision Tree	1.000000	0.741005	0.360885	0.381773	0.604320
SVC Poly	0.828241	0.815919	0.610922	0.220443	0.763914
Random Forest	0.984804	0.797684	0.487535	0.216749	0.745506
Kernel SVM	0.825530	0.818876	0.644195	0.211823	0.765780
KNN	0.851980	0.794480	0.482484	0.373153	0.748962

GridSearchCV optimizes hyperparameters for recall

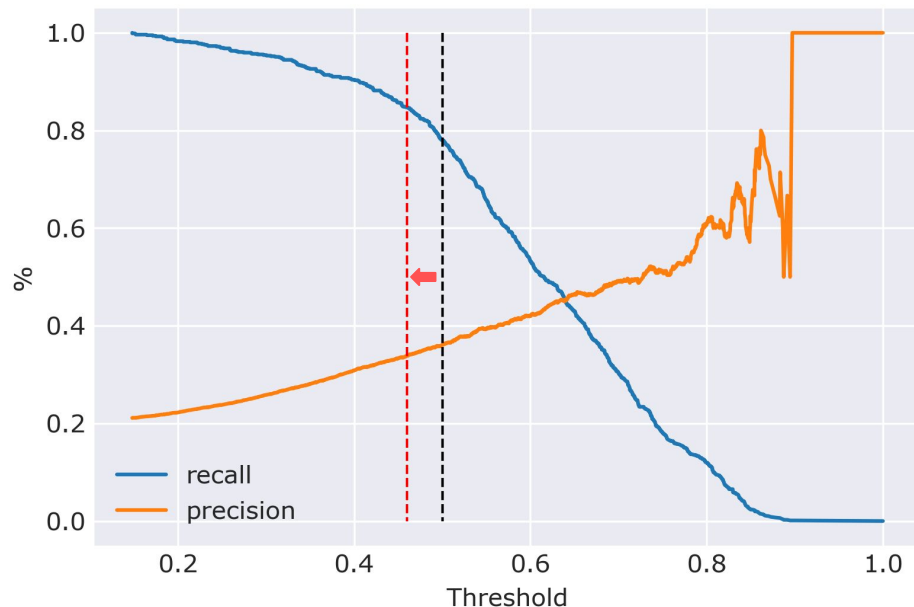
<u>Hyperparameters</u>	<u>Best Score Choice</u>
C	.001
Class Weight	"Balanced"
Penalty	"L2"

<u>Accuracy</u>	68%
<u>Class 1 Precision</u>	36%
<u>Class 1 Recall</u>	78%

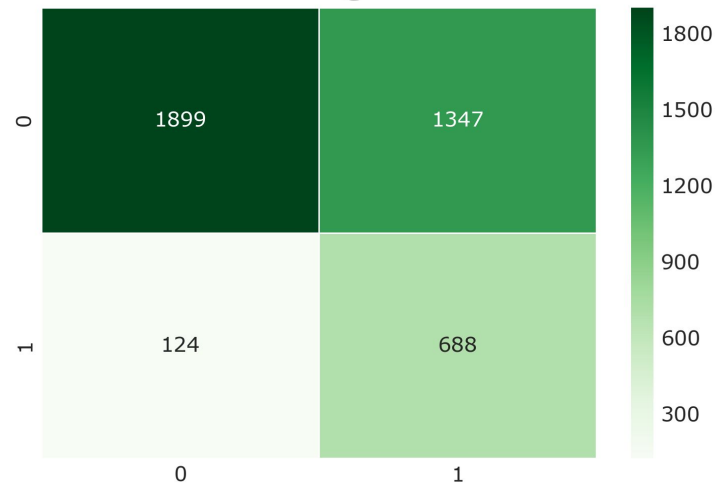


Lowering threshold to .46 improves recall dramatically

Precision Recall Curve



Confusion matrix @ .46 Threshold



Class 1 Precision	34% (-2%)
Class 1 Recall	85% (+7%)

Demo

Will it Chart?

Search for a song and see if it will appear on the Billboard Hot 100.

Thank you!
