



p4k.ai

Predicting Pitchfork rating from  
Spotify music characteristics

Brian Rice

# Background

## Pitchfork

- Music publication company
- Spotlights, reviews, and ratings for new albums, tracks, and reissues
- >1.5M unique visitors per month

# Motivation

But...

- Biased
- Little transparency
- Pretentious

→ Let's shed some light on their scoring process!

# Demo



# Data

Data source:

*P4KxSpotify: A Dataset of Pitchfork Music Reviews and Spotify Musical Features*

- ~18K records/reviews
- Pitchfork reviews scraped from website
- Spotify features queried from API
- Cleaned and stored in RDS



Column	Sample value
artist	Run the Jewels
album	Run the Jewels 2
recordlabel	Mass Appeal
releaseyear	2014
score	9.0
reviewauthor	Ian Cohen
genre	Rap
reviewdate	October 29 2014
key	4.91667
acousticness	0.09457
danceability	0.63983
energy	0.65425
instrumentalness	0.04700
liveness	0.27186
loudness	-7.84217
speechiness	0.23649
valence	0.36117
tempo	123.15392

# Model

## Goal:

- Predict score based on music characteristics

## Model:

- Gradient-boosted tree used to minimize validation set error

## Evaluation:

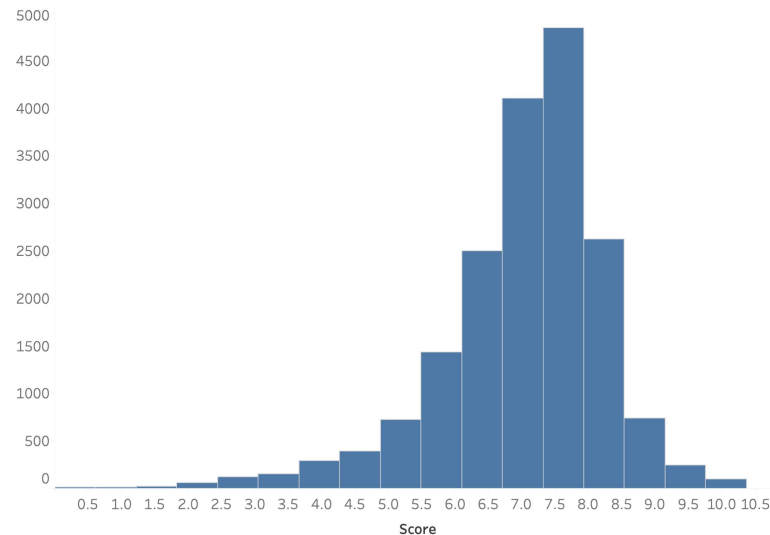
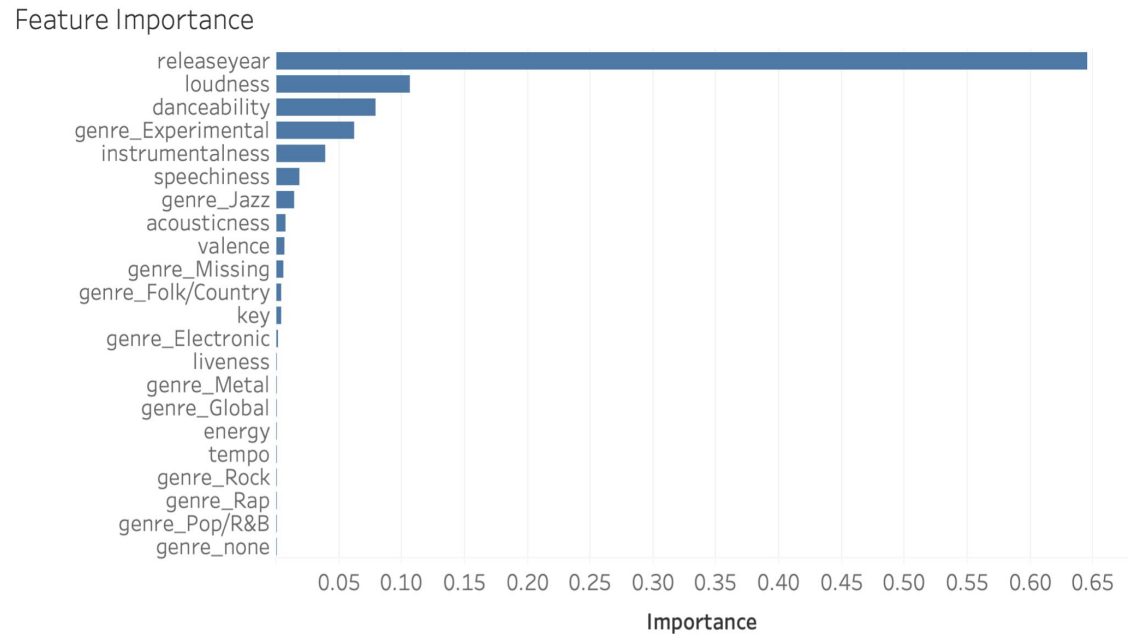
- Target RMSE: 1.00
- Holdout performance: 1.21

# Insights

Key features:

- Release year
- Loudness
- Danceability

½ of scores are 7s!



# Thank you!

Connect with me:

