

CLIMBING THE CHARTS TOUR SHOHUM BOKER

AUG 22 - ONE NIGHT ONLY



The Line Up:

- Setting Up the Stage
- Opening Act
- The Main Event
- Encore
- Wrapping it up
- Backstage Autographs

Setting Up the Stage:

- Data Source: Kaggle, Billboard Hot 100
- Data: 1: 5, Hits to Duds ratio
- Metrics: Spotify
 - Fact-Based: Duration, Key, Tempo
 - Subjective: Danciblity, Valence, and Energy
 - Probability: Liveness, Acousticness, and Instrumentalness

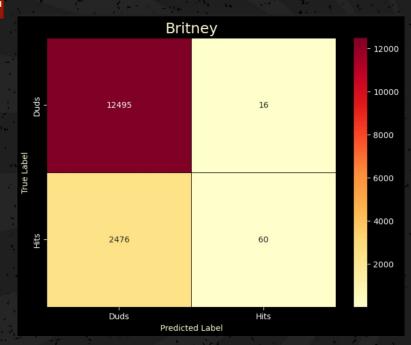
Opening Act - Business Questions:

Given many samples of various tracks, which going to be hits?

Objective: We are looking for top talent in a sea of mediocrity

Opening Act - Britney Model:

- The Britney model use a random forest.
- Really high precision, really low recall
- Most of the models hits are hits,
 but it gets a small amount of hits





Opening Act - Why Use it?

Great if you only need a few hits, but they better be hits

 So if you want filter through a thousands of samples to find to only a few quilty tracks



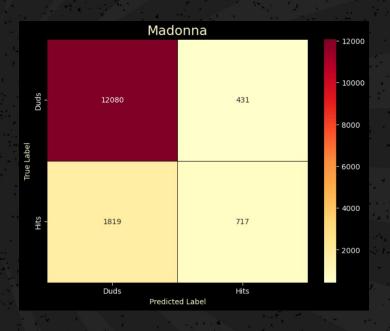


Main Act - Business Questions:

• Given a playlist with few hundred to a thousand tracks, how can we filter for hit songs?

• Objective: We need a model that can give us some hits

Main Act - Madonna Model



The Madonna model uses Catboost.

• Lower precision, but higher recall

Gets more hits to recommend to users

Main Act - What Makes a Hit



- Using the weights from a Logistic Regression as Swifty Stats.
- Put more vocals in track, but not full a spoke words piece

The people like dancing and nice loud track

Encore - What's next:

• Training the Madonna model for specific user's taste and then running it over many other playlists to filter it to their taste

Trying different genres that do not make to the Billboard Hot 100

Wrapping It Up - Recommendations:

 If you want make hit, vocals and loudness matter, little instrumentals parts

Training models on user's data and filtering already made playlist



Backstage Autographs - Recommendations:









