









Our Music Playlists

Netflix Data Science BootCamp Final Project Group 4

Adeh Hakobian, Lisa Hong, Justin Baytosh, Saad Iqbal, and Ashwini Athreya























| Source | Kaggle |
|--|--------|
| Rows | 586672 |
| Columns | 21 |
| Features Danceability, Loudness, Valence, Acoust Tempo, Energy, Popularity, etc | |











Library





Our Music Playlists





Our Goal

To predict the popularity of a song based off of a song's audio features

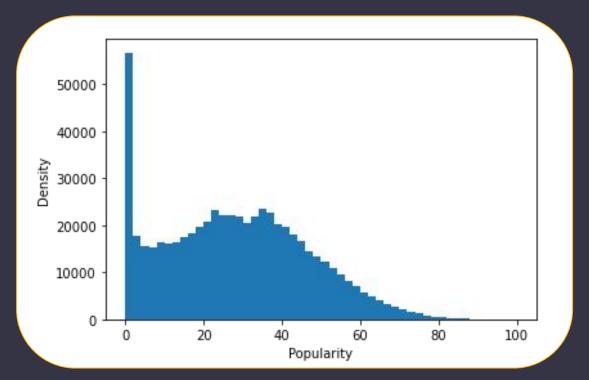
















Data Cleanup











Separated popularity into 5 bins







Resample

Imbalanced data





Better expose important relationships between input and target













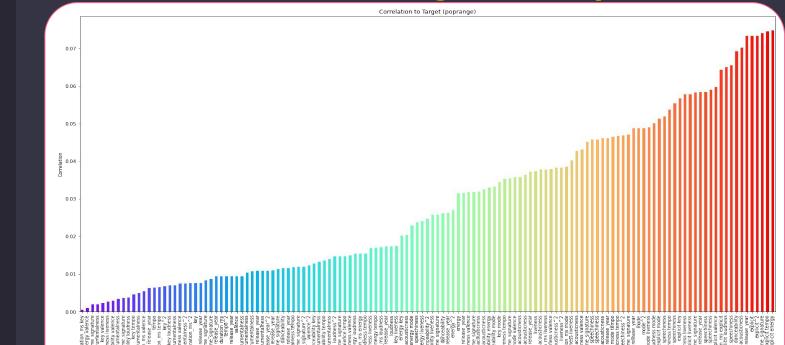
Imbalanced Support Vector Machine

0.516

Accuracy score

| | Precision | Recall |
|---|-----------|--------|
| 0 | 0.73 | 0.53 |
| 1 | 0.50 | 0.01 |
| 2 | 0.48 | 0.95 |
| 3 | 0.53 | 0.02 |
| 4 | 0.00 | 0.00 |

Correlation to Target + PolyFeats

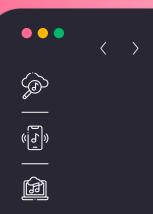












Random Forest Classifier

85%

Total Accuracy Of the Model









Results







0.8466

Accuracy score

| | Precision | Recall |
|---|-----------|--------|
| 0 | 0.89 | 0.94 |
| 1 | 0.77 | 0.79 |
| 2 | 0.70 | 0.59 |
| 3 | 0.85 | 0.91 |
| 4 | 0.99 | 1.00 |





Optimization of Classifier





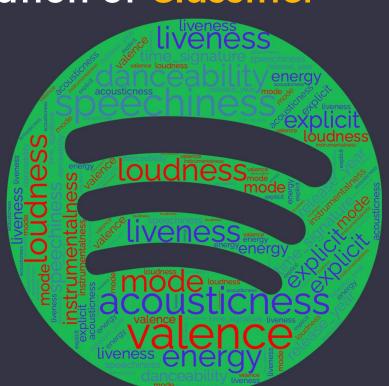




Sequential

Feature

Selection











Important Use Cases

Promotion

Help Potential Stars

Boosting songs that have potential for high virality

Creativity

Tailor-made

Spark creativity for artists to utilize current trends

Growth

Maximize Revenue

By increasing profits for songs that will increase subscribers











Conclusion

- High precision for most popular songs
- 85% overall accuracy
- Possibly improve accuracy with more computing power
- Promoting songs with most potential which can drive profits









Library





Our Music Playlists

Thank

You!





