Hit or Flop?

Predicting Track Success

Problem Statement

Marketing a track is an expensive process

Is there a way to tell which tracks will be successful?

Business Value

Decrease marketing budget

Increase the impact of hits

Increase profit

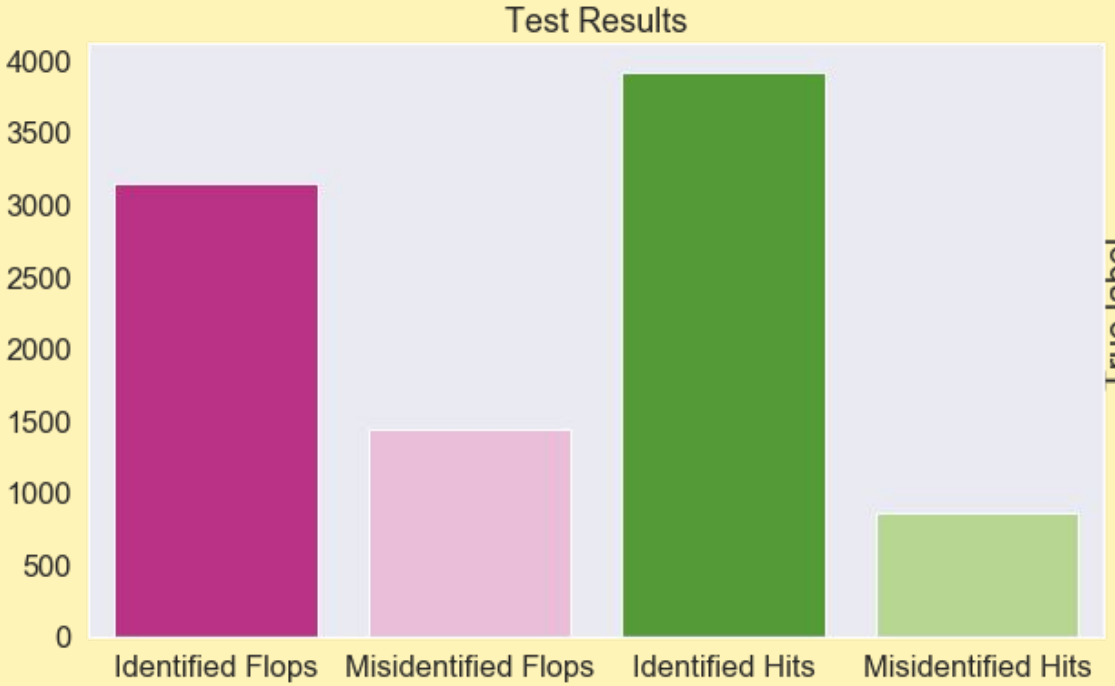
Methodology

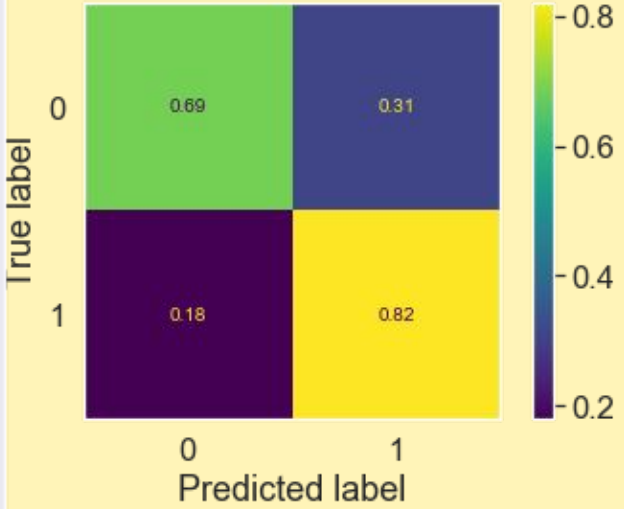
31,000+ Tracks obtained from Spotify.

Measured features such as duration, energy, and danceability.

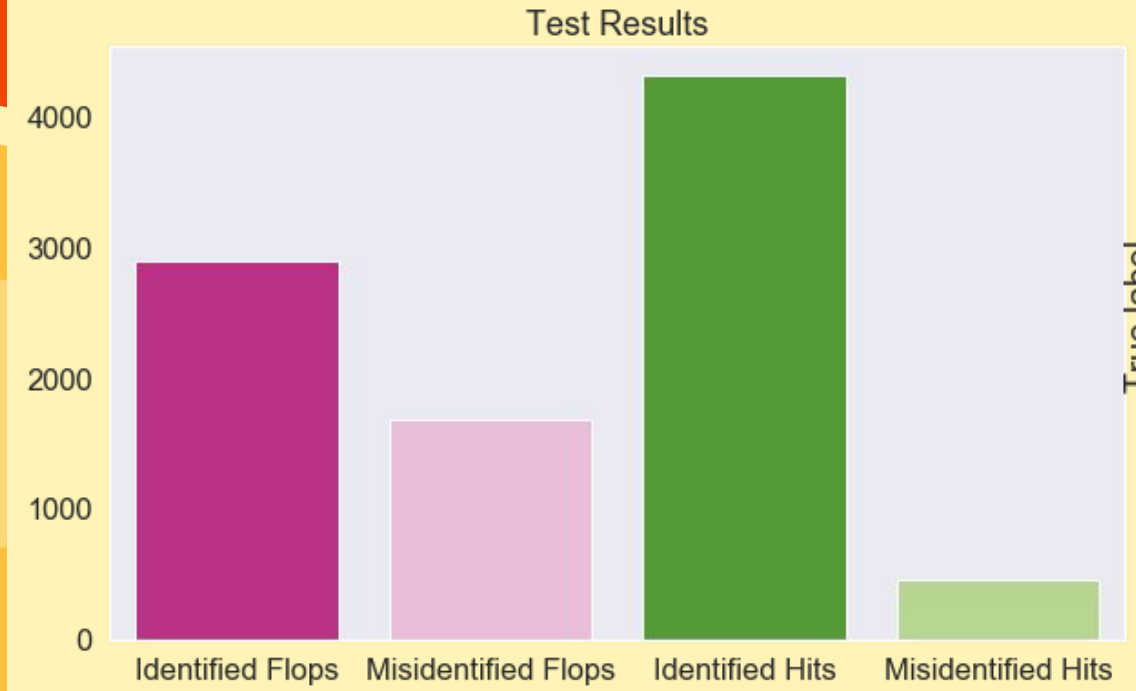
Built multiple classification models to test.

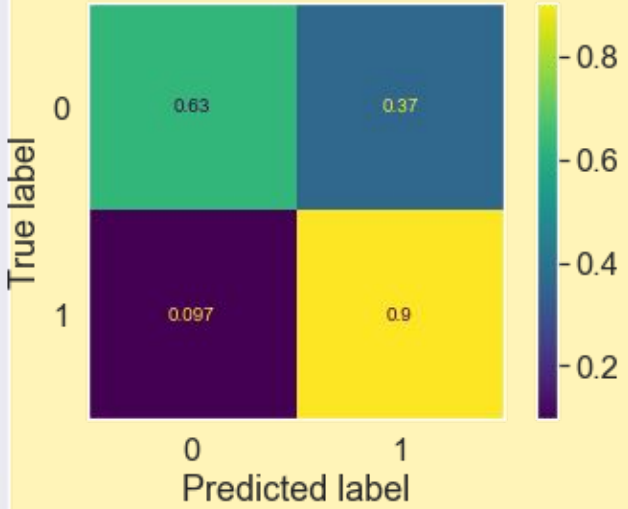
Final Models - Logistic Regression



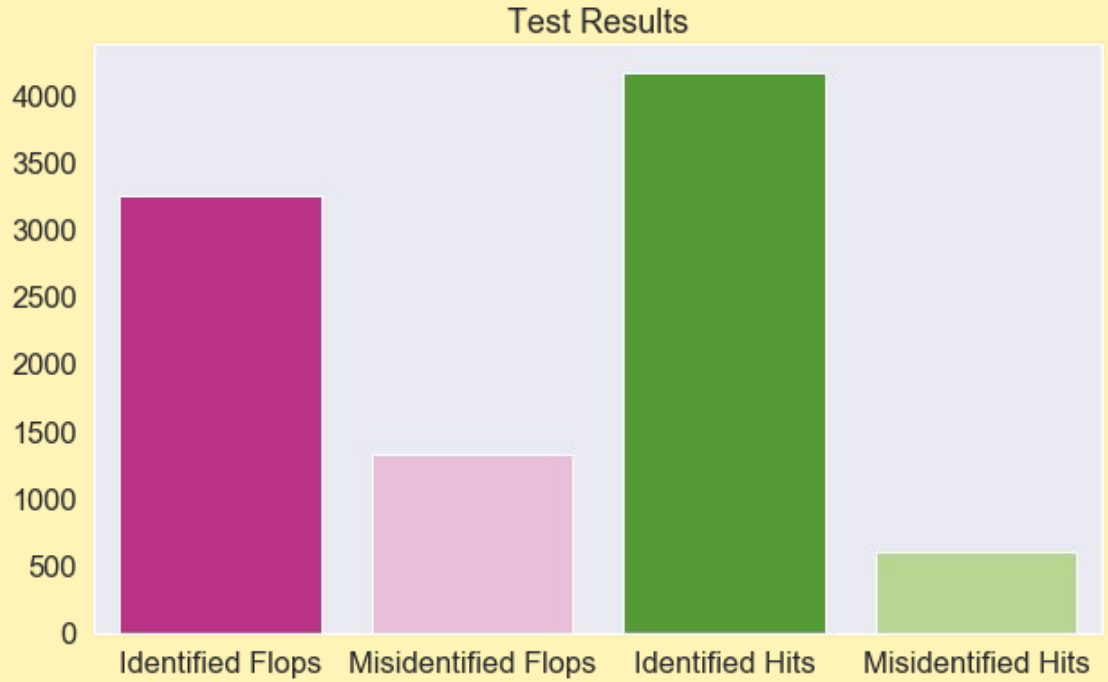


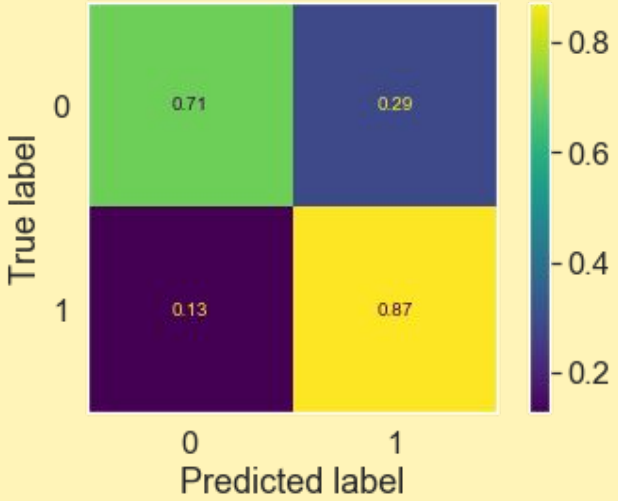
Final Models - K-Nearest Neighbors



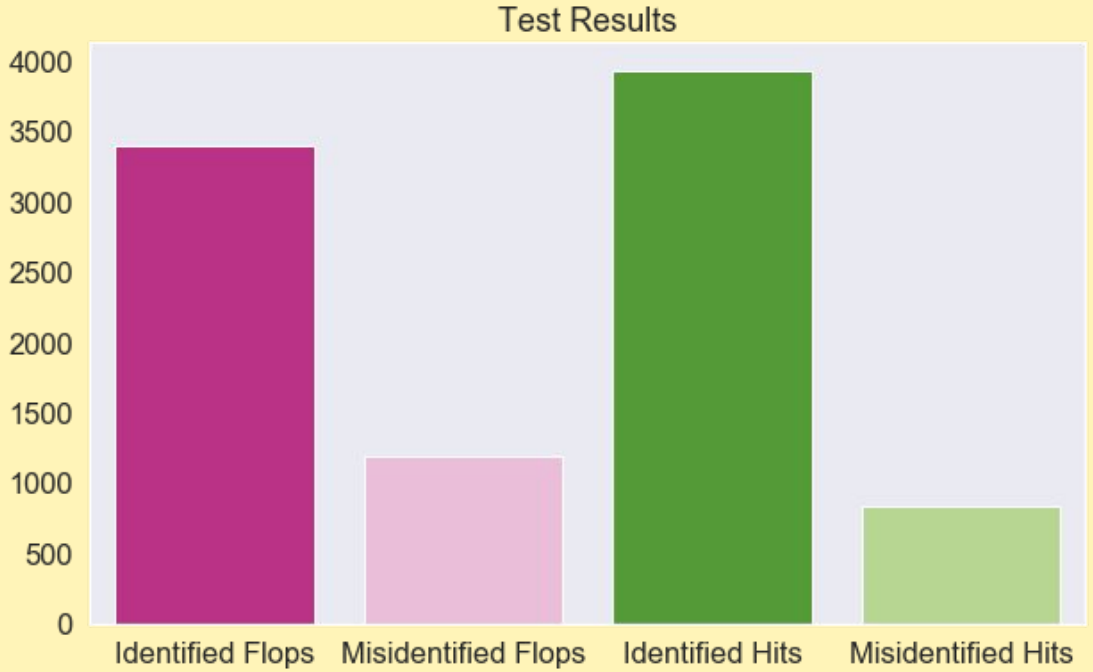


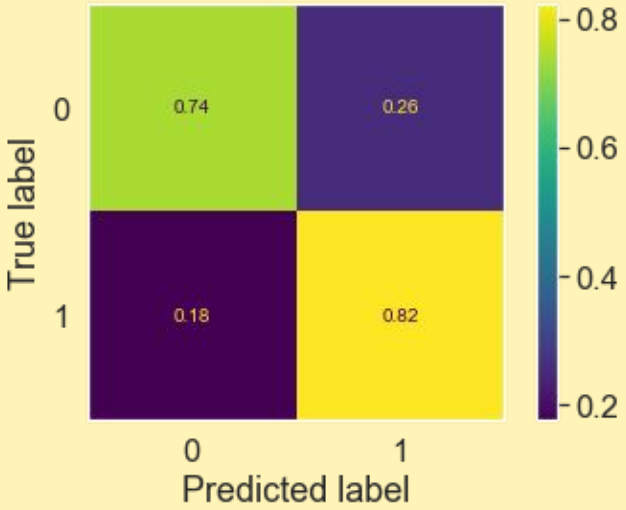
Final Models – Support Vector Machine



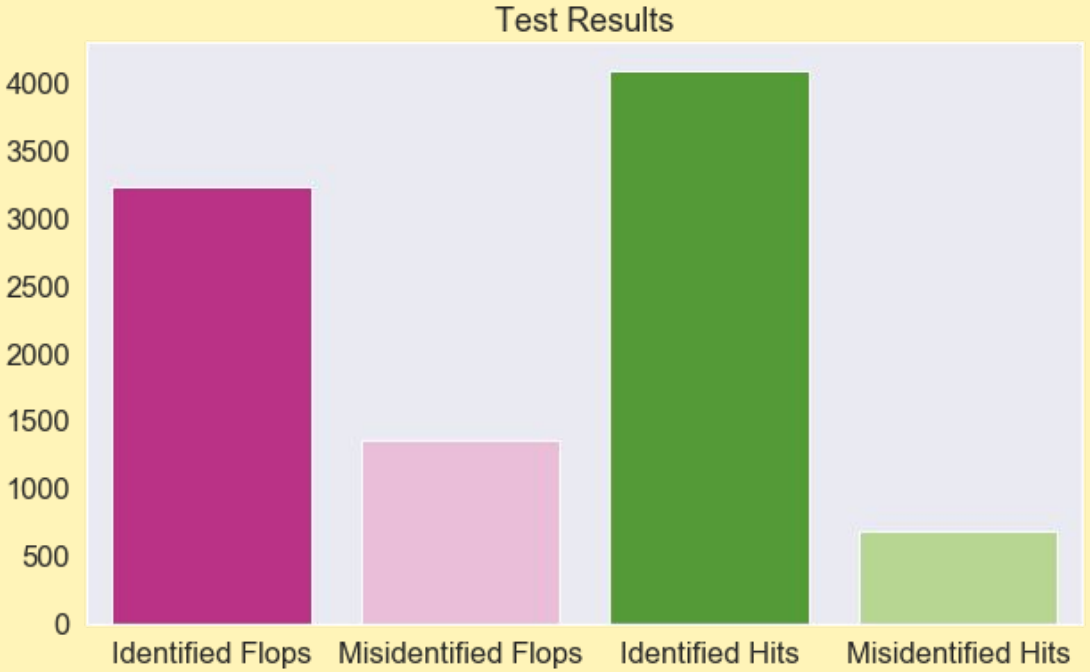


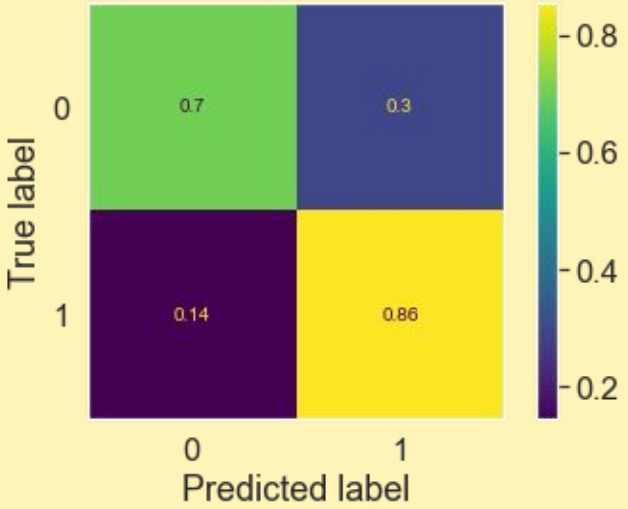
Final Models - Random Forest





Final Models - Adaboost





Best Features

Instrumentalness

Acousticness

Recommendations

Use SVM Model

Use Random Forest when needed

Look for optimal features

Future Work

Testing model on genres of music

Include artist statistics

Train models on more recent music

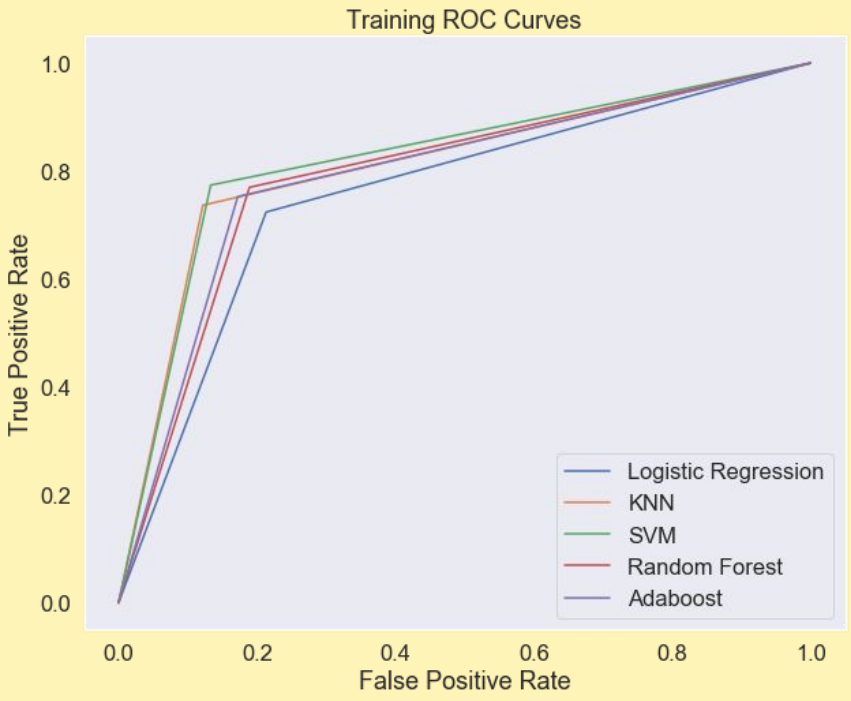
Thank You!

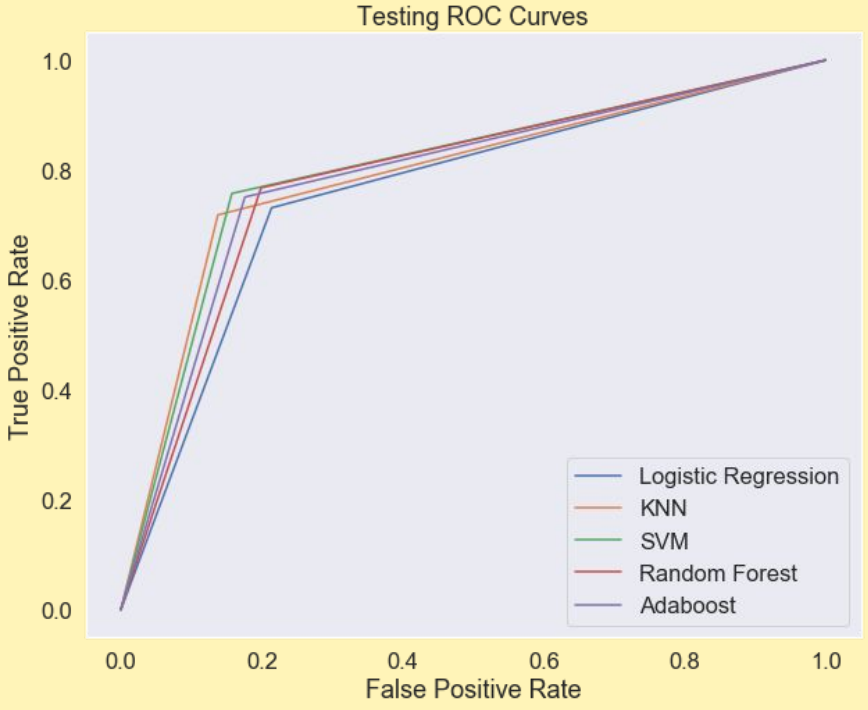
Appendix - Stats

- Trained five models:
  - Logistic Regression - Train AUC .755, Test AUC .759
  - K-nearest neighbors - Train AUC .807, Test AUC .790
  - SVM - Train AUC .820, Test AUC .800
  - Random Forest - Train AUC .790, Test AUC .784
  - Adaboost - Train AUC .790, Test AUC .787

- Attempted to use PCA, but accuracy suffered
  - Logistic Regression - Train acc .695, Test acc .699
  - K-nearest neighbors - Train acc .804, Test acc .720
  - SVM - Train acc .760, Test acc .751
  - Random Forest - Train acc .999, Test acc .756
  - Adaboost - Train acc .726, Test acc .732

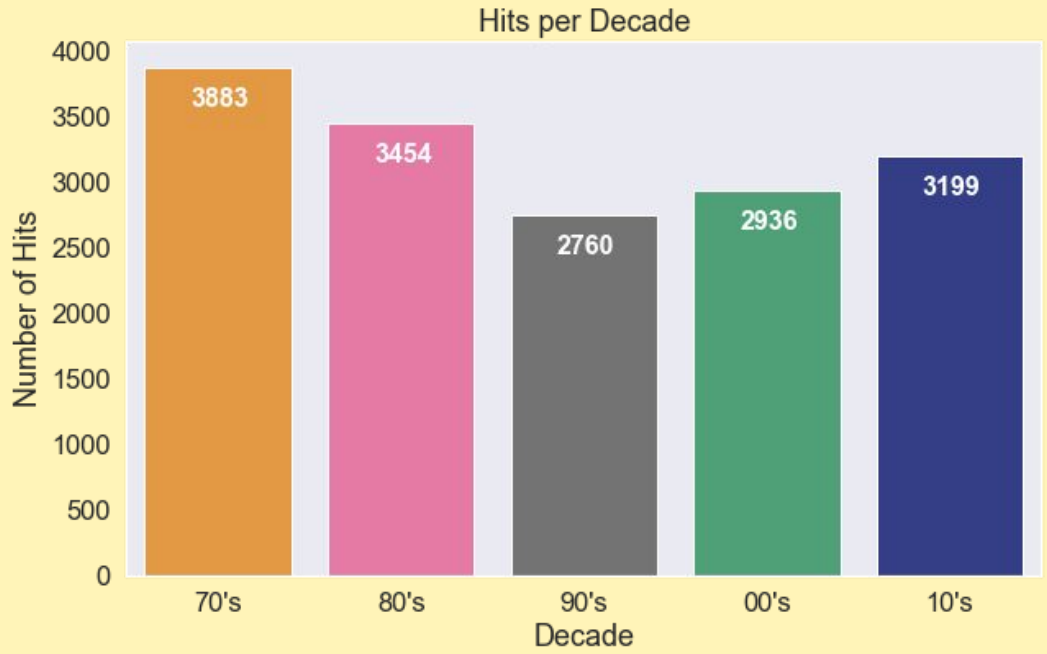
Appendix – ROC Curves





Appendix – Hit Distribution

Distribution of hits by decade



Appendix – Features

SVM Best Features

Weight	Feature
0.0857 ± 0.0039	instrumentalness
0.0527 ± 0.0063	acousticness
0.0498 ± 0.0043	danceability
0.0359 ± 0.0082	energy
0.0263 ± 0.0067	duration_ms
0.0191 ± 0.0047	valence
0.0190 ± 0.0019	loudness
0.0098 ± 0.0034	speechiness
0.0052 ± 0.0008	liveness
0.0047 ± 0.0046	sections
0.0027 ± 0.0008	key
0.0025 ± 0.0012	tempo
0.0020 ± 0.0022	mode
0.0010 ± 0.0026	chorus_hit
0.0010 ± 0.0012	time_signature_4
0.0006 ± 0.0007	time_signature_3
0.0001 ± 0.0003	time_signature_1
0.0000 ± 0.0003	time_signature_5

Random Forest Best Features

Weight	Feature
0.1030 ± 0.0080	instrumentalness
0.0495 ± 0.0024	acousticness
0.0321 ± 0.0023	duration_ms
0.0272 ± 0.0050	danceability
0.0077 ± 0.0012	energy
0.0061 ± 0.0022	speechiness
0.0036 ± 0.0012	loudness
0.0023 ± 0.0023	valence
0.0011 ± 0.0009	time_signature_4
0.0009 ± 0.0005	liveness
0.0007 ± 0.0009	tempo
0.0007 ± 0.0013	mode
0.0006 ± 0.0015	sections
0.0004 ± 0.0003	time_signature_3
0.0001 ± 0.0003	chorus_hit
0 ± 0.0000	time_signature_5
0 ± 0.0000	time_signature_1
-0.0001 ± 0.0004	key