

Spotify Billboard Classifier

Course Project: DATA 1030 - Hands On Machine Learning

Sayan Samanta

Instructor: Dr. Andras Zsom

TA Advisor: Natalie Delworth

- 1 Introduction
 - Database Introduction

Introduction to Database

- The Database contains **acoustic features** of songs available on Spotify along with the date of release of the album which contains the track.
- The original database had songs from 1966-2019 that appeared on the Billboard top 20.
- Further data (songs not appearing in the billboard) were mined using the Python Spotify API - Spotipy.
- The auxillary database was based on top 10 tracks by artists similar to those in the billboard (but songs that didn't feature in the billboard)

| | acousticness | album_id | artist_id | billboard | \ | |
|---|--------------|------------------------|------------------------|-----------|---|--|
| 0 | 0.00784 | 2Auw0pTT6EcQdvHNimhLQI | 04gDigrS5kc9YwfZHWBETP | 1.0 | | |
| 1 | 0.05910 | 2Auw0pTT6EcQdvHNimhLQI | 04gDigrS5kc9YwfZHWBETP | 1.0 | | |
| 2 | 0.28400 | 2Auw0pTT6EcQdvHNimhLQI | 04gDigrS5kc9YwfZHWBETP | 1.0 | | |

| | danceability | date | duration_ms | energy | id | \ | |
|---|--------------|------------|-------------|--------|------------------------|---|--|
| 0 | 0.532 | 2015-05-15 | 238200.0 | 0.599 | 1m39XApXHXb2UkGItyyIU0 | | |
| 1 | 0.748 | 2015-05-15 | 235493.0 | 0.788 | 2iuZJX9X9P0GKaE93xcPjk | | |
| 2 | 0.501 | 2015-05-15 | 203453.0 | 0.704 | 13YQutqrAht5iX9H0ctlu5 | | |

| | instrumentalness | key | length | liveness | loudness | mode | speechiness | \ | |
|---|------------------|-----|--------|----------|----------|------|-------------|---|--|
| 0 | 0.0 | 6.0 | 12.0 | 0.1400 | -6.543 | 1.0 | 0.0333 | | |
| 1 | 0.0 | 1.0 | 12.0 | 0.0863 | -7.055 | 1.0 | 0.0334 | | |
| 2 | 0.0 | 3.0 | 12.0 | 0.1550 | -5.640 | 1.0 | 0.0326 | | |

| | tempo | time_signature | track_length | valence |
|---|---------|----------------|--------------|---------|
| 0 | 130.019 | 4.0 | 225984.25 | 0.113 |
| 1 | 120.076 | 4.0 | 225984.25 | 0.884 |
| 2 | 97.039 | 4.0 | 225984.25 | 0.270 |

Figure: Sample data with Features

