Topic: HitPredict: Predicting Billboard Hits Using Spotify Data

❖ Overview:

1. We approach the "Hit Song Science" problem, aiming to predict which songs will become Billboard Hits.
2. Published weekly, first publication :August 4,1958.
3. "Radioactive", by Imagine Dragons (A Major, 136 BPM).
4. To reach the Billboard HOT 100, based on the song's audio features we implement model to predict whether it is a song or not a song.
5. The goal of this project is to see if a song's audio characteristics and lyrics can determine a song's popularity.

❖ Data Collection:

1.) **Billboard Hits:**

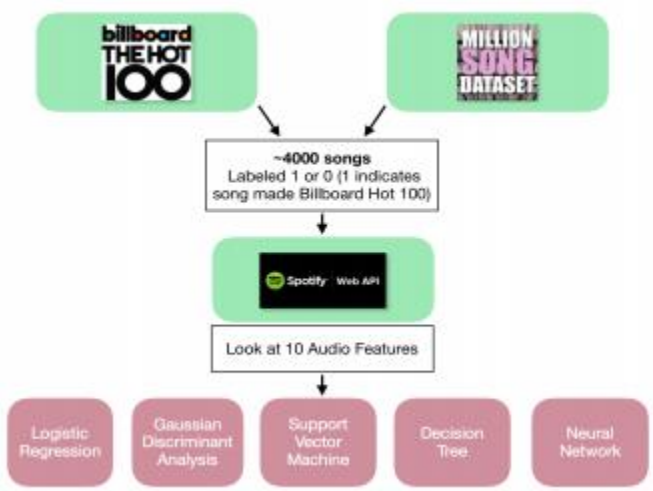
- All unique songs featured on "Billboard Hot 100"
- 1990- 2018
- Billboard API Library
- Dataset: › Artist name, song title, other misc. features

2.) **Million Song Dataset:**

- Million Song Dataset ,a free dataset maintained by labROSA at Columbia University and EchoNest.
- This was narrowed down to songs released between 1990 and 2018.

3.) **-4000 steps:**

- To balance the dataset between positive (hits) and negative (non-hits).
- We removed two thirds of the songs collected from the Billboard Hot 100.
- At the end, removed overlapping songs to form a dataset of approximately 4,000 songs.



❖ **Algorithms:**

1.) Supervised Learning:
2.) Logistic Regression & GDA
3.) Decision Tree
4.) Neural Network

❖ **Features :**

- Ten audio features were extracted from the Spotify API.
- The Artist Score metric was created, assigning a score of 1 to a song if the artist previously had a Billboard hit, and 0 otherwise.

Audio Features	
Danceability	Loudness
Liveness	Tempo
Instrumentalness	Valence
Speechiness	Energy
Acousticness	Artist Score

Fig: . Audio features extracted from Spotify's API.

- Result:
    - We used accuracy, precision and recall on the training and validation sets to evaluate the performance of each algorithm.