PREDICTICTING BILLBOARD HOT 100 HITS USING SPOTIFY DATA

- Introduction

The Billboard Hot 100 is the music industry standard record chart in the United States for songs, published weekly by Billboard magazine. The Billboard Hot 100 Chart1 is still one of the most reliable ways to gauge a song's popularity. This project will be a walkthrough of simple machine learning techniques applied to predict the songs that will become Billboard Hot 100 hits.

- DataSet Description

Our dataset contains data from the following sources :

http://millionsongdataset.com/

https://www.billboard.com/charts/hot-100

Firstly, ten audio features will be extracted from the Spotify API. Spotify assigns each song a value between 0 and 1 for these features except loudness which is measured in decibels.

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

We will also create an Artistic Score metric, assigning a score of 1 if the artist previously had a Billboard Hit, and 0 otherwise.

- # Algorithms Used

Logistic Regression used to predict a data value based on prior observations of a data set.

Supervised Learning : Splitting the data into 75/25 train test ratio.

Decision Tree to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data(training data)

GDA for data classification

SVM or Support Vector Machine is a linear model for classification and regression problems.

Neural Networks

- # Approach

1. Collect raw data
2. Exploratory Data Analysis :It is an approach of analysing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. We will be performing initial investigations on data to discover patterns, spot anomalies i.e. data cleaning, and test hypothesis with help of summary statistics and graphical representations.
3. Data Modelling : Split data in 75/25 proportion to test it.