

BLUEPRINT OF THE MODEL

Predicting Billboard Hits using Spotify Data

Blueprint follows the following pattern -

Data → Data Processing → EDA → TrainingModel →
Test Model & Evaluation → Model Prediction → Model Deployment

1) Data -

Dataset - songs_complete_data.csv

Format - .csv

2) Data Processing-

Data Cleaning - remove null values and duplicates if any , remove outlier, drop lyrics column.

Standardizing the data- If required

3) EDA -

Check for distribution of column TOP100

Plotting graphs - for Distribution of song release year wise

Plotting graphs - _for Distribution of Genre

Plotting graphs - for Distribution of 10 features

4) Training Model-

Segregating into dependent and independent - y(Mode) and the rest of the columns as X.

Splitting dataset - 75:25 ratio

Model Training - Logistic Regression,SVM,GDA, DecisionTree model,Neural Network with L2 regularization

Select Best Model - Logistic Regression,GDA

5) Test Model & Evaluation

TestData - 25% of the Dataset

Evaluation parameter - _Mean absolute error,Accuracy etc.

6) Model Prediction-

Prediction of song to hit top 100

Rectify model Performance technique- Like cross validating the model etc.

7) Model Deployment-

Deploy the model → Using Flask

→ Microsoft azure/ Heroku

Check functionality
