SDAIA T5 bootcamp

Classification Project

Spotify Classification Model

Final Report

_____

Basma Ghazi Abdullah Alduaiji

Mai Abdullah Fahad Aljuaid

28/10/2021

Contents

Table of tables

Introduction

Since music is a universal language, we developed a model that can predict how likely a song will be popular or not based on some song's features. Moreover, this model will increase music industry revenue and allow the music producers and artists to create songs that will be more likely to be popular.

Design

The dataset in this project is taken from the Kaggle website. This data set presents the complete status of the songs and the popularity range.

Algorithm

The steps to analyse the data set are loading data from Kaggle. In addition, exploring data by using all the functions like info and describe. Then cleaning data by removing null values and duplicates as well as, removing any irrelevant data. After that, plot the graphs using seaborn and matplotlib modules from python. Finally, find the correlation between features and fit the classification models in order to find the best model that fits our data and also generalize on unseen data. Finally, deploy the model.

Used data set

The dataset that will be used in this project was found in Kaggle. The target of this project is to predict whether a song will be popular or not. In addition, a set of features will be included in building predictive models such as danceability, duration, energy, liveness, loudness, speediness, tempo and valence as shown in table 1.

Table 1 Spotify Songs Dataset

Acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic.
Danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
Duration ms	Duration of the track in millisecond

Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
Instrumental ness	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context.
Key	The estimated overall key of the track.
Liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
Loudness	The overall loudness of a track in decibels (dB)
Mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived.
Speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
Tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
Time signature	An estimated overall time signature of a track.
Valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

In addition, The Individual sample from the songs data set are 130K rows. Moreover, the expected characteristics used to predict the song popularity are the attributes from song data such as Danceability, Acousticness and Duration of the song. Furthermore, the predicted target is to predict if the song will be popular or not according to a specific song status.

.

Used tools

The company will utilize the tools provided as a module in pythons such as Pandas, Matplotlib, Seaborn NumPy, Sklearn and Flask. The python libraries support multiple data analysis and cleaning methods to ensure the data is clean and ready to visualize. Moreover, Flask is a framework to connect between model and web interface.

Conclusion

This project demonstrated the possibility of predicting music popularity. Moreover, Decision Trees provided the best predictions on the validation model, with an Accuracy score of 0.75 / 0.74 Recall.

Moving forward, we would like to explore how additional features such as artist location or release date can influence a song's popularity.