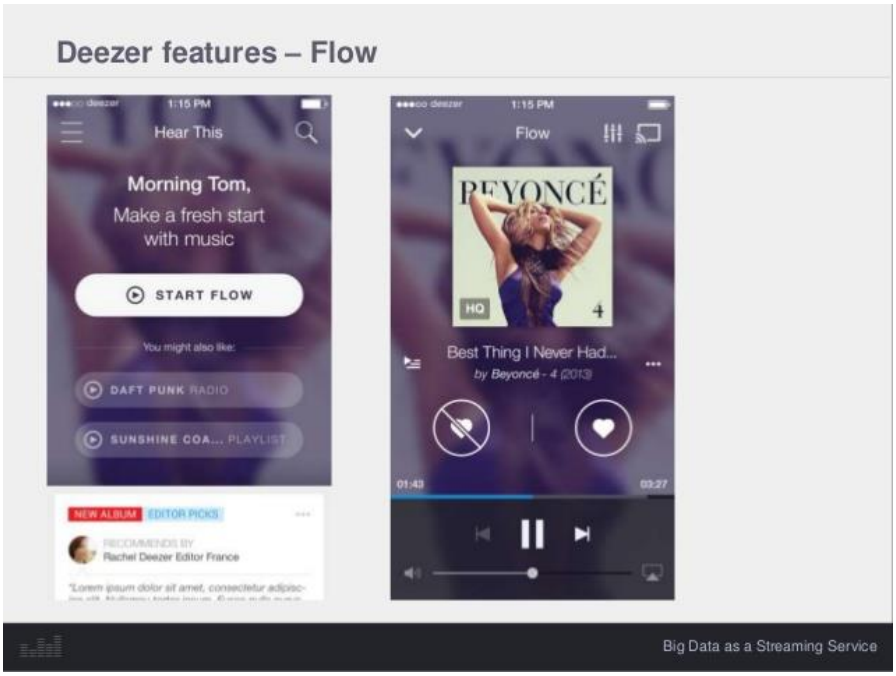
Prove that your university can qualify for the finals! Data Science Game Qualification phase: Music recommendation

Context

Getting the perfect music recommendation is a challenging task. Who has never dreamt of laying back and listening to some music, while having no buttons to press at all and still getting the perfect tune? But what defines this perfect tune?
Deezer is a music streaming app, also available on the web. It proposes more than 43 million tracks and is available in more than 180 countries, through a free limited service and a premium offer.

For this online challenge, Deezer wants you to look at Flow, its own music recommendation radio. The concept of Flow is simple: it uses collaborative filtering to provide a user with the music he wants to listen at the time he wants.... And if he does not want to listen to some specific tracks and skips songs by pressing the 'Next song' button, then the algorithm should detect it quickly. In this context, getting the first song recommendation right is really important.



Goal

The goal of this challenge is to predict whether the users of the test dataset listened to the first track Flow proposed them or not. Deezer considers that a track is "listened" if the user has listened to more than 30 seconds of it (is_listened =1). If the user presses the skip button to change the song before 30 seconds, then the track is not considered as being listened (is_listened = 0).

The test dataset consists in a list of the first recommended tracks on Flow for several users. Each row represents one user.

The train dataset was generated using the listening history of these Deezer users for one month. Each row represents one listened track. The list of distinct users in the train dataset matches exactly with the test dataset's one."

File descriptions

- **train.csv** - the training set
- **test.csv** - the test set
- **sample_submission_kaggle.csv** - a sample submission file in the correct format
- **extra_infos.json** - supplementary information about the songs

The datasets contain the following features (variable **is_listened** is only present in the train dataset as it is the variable you need to predict):

Data fields

- **media_id** - identifiant of the song listened by the user
- **album_id** - identifiant of the album of the song
- **media_duration** - duration of the song
- **user_gender** - gender of the user
- **user_id** - anonymized id of the user
- **context_type** - type of content where the song was listened: playlist, album ...
- **release_date** - release date of the song with the format YYYYMMDD
- **ts** - timestamp of the listening in UNIX time
- **platform_name** - type of os
- **platform_family** - type of device
- **user_age** - age of the user
- **listen_type** - if the songs was listened in a flow or not
- **artist_id** - identifiant of the artist of the song
- **genre_id** - identifiant of the genre of the song
- **is_listened** - 1 if the track was listened, 0 otherwise

You are also provided extra informations (**extra_infos.json**), linking each media_id, to a song title, album title and an artist name.

Evaluation

The evaluation metric is the ROC AUC:
https://en.wikipedia.org/wiki/Receiver_operating_characteristic

Submission Format

sample_id,is_listened
0,0.1
1,0.8
2,0.5
3,0.0
...
19917, 0.7

- When submitting a prediction, you have to change your account name with:
  Team Name - University - Country
- IMPORTANT: PLEASE UNDERSTAND THAT ONLY ONE KAGGLE ACCOUNT FROM EACH TEAM WILL BE INVITED TO ACCESS THE COMPETITION ! SO DO NOT TRY TO SHARE THE INVITATION OR INVITE OTHER TEAMMATES: THIS WON'T WORK!

Competition Rules

Joining the competition

As explained in the email you have received, each DSG team received only **one** invitation to join the competition: to simplify the process, we ask you to use only one Kaggle account per team. Before anything else, you need to create a Kaggle team for your own account and name it with your DSG team name. Contact us if you have any question on ADRESSE MAIL CHALLENGE.

External Data

You are allowed to use the Deezer API (you'll have to create a Deezer account) to get additional information if you need so. API documentation is available on the provided link.
No other external data source is allowed.

One account per team

You cannot sign up to Kaggle from multiple accounts and therefore you cannot submit from multiple accounts.

No private sharing outside teams

Privately sharing code or data outside of teams is not permitted. It's okay to share code if made available to all participants on the forums.

Submission Limits

You may submit a maximum of 4 entries per day.
You may select up to 2 final submissions for judging.

Qualification Rules

The competition will begin by an online qualifier round. The 20 best teams according to the private leaderboard will be invited to the final phase. These teams should match the following conditions:

- A university can be represented by at most one team. Only the best team from the online qualification round will qualify.
- A country can be represented by at most five teams. Only the five best from the online qualification phase will qualify for the final.
- To ensure that no team is cheating, each qualified team must provide the code reproducing their best submission .

Competition Timeline

Start Date: **15/04/2017**
Merger Deadline: **None**
Entry Deadline: **None**
End Date: **31/05/2017 12:00 AM UTC**