# Spotify Sequential Skip Prediction Challenge

Predict if users will skip or listen to the music they're streamed

By **Spotify**

Completed

| 28623 | 526 | 686 |
|-------|-----|-----|
| Views | Participants | Submissions |

♥ 74    FOLLOW

Overview    Leaderboard    Discussion    Dataset

## Overview

> Update 08 Jan 2019: The submission system is now live on EasyChair . We hope to see many of you submit reports on your work for this challenge.

> Update 07 Jan 2019: The final results are now available. We will be contacting the teams to confirm that their code is open sourced and can be verified, but a provisional congratulations to the winning teams, and thank you all for participating in this challenge. We look forward to reading and hearing about your insights.

> Update 04 Jan 2019: Good luck to all contestants in the final hours of the challenge! Once the submission period has concluded we will begin the final leaderboard evaluations. Additionally, we are in the process of finalizing the paper submission system for the WSDM Cup workshop, the deadline will be January 11, 2019. In the meantime, see the Call for Papers section here on the challenge overview page for information. Please note that submitting a paper is mandatory in order to be considered for the winning leaderboard positions, and in order to be eligible for the prizes.

> Update 12 Dec 2018: We would like to make several announcements: (1) We are happy to share that Google have kindly offered to sponsor coupons for google cloud compute resources for participants of this challenge. Please see the 'Google Sponsored Computational Resources' section of the overview page for further details. (2) We have released the call for papers for the WSDM Cup Workshop day. Please see the 'Rules' and 'Call for Papers' sections of the overview page for further details. (3) We are now providing the training set split into 10 files to make it easier for participants with slow connections to download the training set. Please see the Training_Set_Split_Download.txt file under the Dataset tab for the download links. (4) There was some ambiguity in the description of the challenge metric which has now been clarified, see the 'Evaluation' section of the overview page for further details. Please note that the metric is unchanged we have simply clarified the terminology.

> Update 20 Nov 2018: Unfortunately we have had to make some changes to the challenge dataset. More specifically, we have had to remove some features from the track features table (the updated Dataset Description file outlines the new track features schema). Please note that the other parts of the dataset all remain unchanged, except the track features table in the mini version of the dataset which was changed correspondingly. We apologize for any inconvenience caused by this change. If your work on the challenge is affected, we would appreciate if you email us at wsdm-cup-2019@spotify.com so that we can better understand any potential impact on participants.

Spotify is an online music streaming service with over 190 million active users interacting with a library of over 40 million tracks. A central challenge for Spotify is to recommend the right music

to each user. While there is a large related body of work on recommender systems, there is very little work, or data, describing how users sequentially interact with the streamed content they are presented with. In particular within music, the question of if, and when, a user skips a track is an important implicit feedback signal.

We release this dataset and challenge in the hope of spurring research on this important and understudied problem in streaming. Our challenge focuses on the task of session-based sequential skip prediction, i.e. predicting whether users will skip tracks, given their immediately preceding interactions in their listening session.

The organization of this challenge is a joint effort of Spotify , WSDM , and CrowdAI .

## Dataset

The public part of the dataset consists of roughly 130 million listening sessions with associated user interactions on the Spotify service. In addition to the public part of the dataset, approximately 30 million listening sessions are used for the challenge leaderboard. For these leaderboard sessions the participant is provided all the user interaction features for the first half of the session, but only the track id's for the second half. In total, users interacted with almost 4 million tracks during these sessions, and the dataset includes acoustic features and metadata for all of these tracks.

If you use this dataset in an academic publication, please cite the following paper:

@inproceedings{brost2019music, title={The Music Streaming Sessions Dataset}, author={Brost, Brian and Mehrotra, Rishabh and Jehan, Tristan}, booktitle={Proceedings of the 2019 Web Conference}, year={2019}, organization={ACM} }

## Challenge

The task is to predict whether individual tracks encountered in a listening session will be skipped by a particular user. In order to do this, complete information about the first half of a user's listening session is provided, while the prediction is to be carried out on the second half. Participants have access to metadata, as well as acoustic descriptors, for all the tracks encountered in listening sessions.

The output of a prediction is a binary variable for each track in the second half of the session indicating if it was skipped or not, with a 1 indicating that the track skipped, and a 0 indicating that the track was not skipped. For this challenge we use the skip_2 field of the session logs as our ground truth.

There will be a workshop at WSDM where selected or top performing teams will be invited to present their work on this challenge. The paper submission deadline will be January 11, 2019, and the workshop will be held on February 15, 2019, as part of WSDM in Melbourne, Australia

## How to generate submissions

The test set sessions are always split between two files. Each session is partly contained in a prehistory file, and a corresponding input file. The full interaction feature set for the first half of the session is contained in the prehistory file, and the track id's for which you need to make a prediction are contained in the input file. For each test set session a row of 1's and 0's of the same length as the input part of the session must then be generated. Sample submissions are contained in the Sample_Submissions.tar.gz file under the Dataset tab, and code for generating a random submission is contained in the Starter Kit .

## Partners



## Evaluation

Accurate skip prediction can enable us to avoid recommending a potential track to the user, based on the user's immediately preceding interactions. At a given moment in time, it is therefore most important to predict if the next immediate track is going to be skipped, but it would also be useful to predict if the tracks further into the session will be skipped. This motivates our use of

Mean Average Accuracy as the primary metric for the challenge, with the average accuracy defined by

$$AA = \frac{\sum_{i=1}^{T} A(i)L(i)}{T}$$

where :
- T is the number of tracks to be predicted for the given session
- A(i) is the accuracy at position i of the sequence
- L(i) is the boolean indicator for if the i'th prediction was correct.

We will use the accuracy at predicting the first interaction in the second half of the session as a tie breaking secondary metric.

## Rules

The competition rules are given below, the dataset terms of use and challenge terms and conditions can be found on the Dataset page.

1. Participants must agree to the dataset terms of use and challenge terms and conditions in order to participate in this challenge.
2. Each [participant] account shall represent only one (1) team.
3. Team mergers are allowed at any time prior to the two (2) week period before the competition end date and must be communicated in writing to the challenge organizers at wsdm-cup-2019@spotify.com .
4. A merged team forms one (1) teams for the remainder of the competition, and such team must designate a single account for subsequent submissions in the competition. The other accounts associated with team members in a merged team shall not be used for further submissions.
5. Each team may only make five (5) submissions per day. The last submission day is January 4, 2019.
6. At the end of the challenge, each team is required to open source the source code that was used to generate their final challenge solution under the Apache license 2.0. To be eligible for the leaderboard or prizes, teams are also required to submit papers describing their method to the WSDM Cup Workshop. See the 'Call for Papers' section below for further details.
7. Teams may use external data and pretrained models, provided these are public, and linked with the source code provided for the final submission.
8. Teams may use the Spotify API to augment the dataset, provided this use complies with the terms of service for the Spotify API, and that the augmentation is included with the source code provided for the final submission.
9. Spotify employees and their in-home relatives are prohibited from participating in the contest.
10. Ask us: If you are not sure whether something is allowed or not, do not hesitate to contact us at wsdm-cup-2019@spotify.com ! We can quickly give you the information you need.

## Prizes

The prizes will be administered as part of the 2019 WSDM Cup. The winning team will be awarded **AUD2000**, the second placed team will be awarded **AUD750**, and the third placed team will be awarded **AUD250**. All prizes are in Australian Dollars.

## Call for Papers

Submissions must be in English, in PDF format, and should not exceed four pages in the current ACM two-column conference format (including references and figures). Suitable LaTeX and Word templates are available from the ACM Website. The papers can represent reports of original research, preliminary research results, or proposals for new work. The review process is single-blind.  Please mention the team name in the title or abstract, and provide a link to the repository for the open sourced code in your paper. Papers will be evaluated according to their significance, originality, technical content, style, clarity, and likelihood of generating discussion. The submission deadline is January 11, 2019 (AOE timezone).

Papers should be submitted on EasyChair

## Resources

A starter kit for participants to familiarize themselves with the dataset and challenge mechanics is provided at: Starter Kit

Information about the Spotify API is provided at: Spotify API

For an introduction to some of the factors that affect user skip behaviour, see the following blog entry from Paul Lamere: MusicMachinery - Entry on skips

# Google Sponsored Computational Resources

We are very grateful to Google, who have kindly offered to sponsor 100 USD coupons for Google cloud compute resources for participants of this challenge. Teams that have made a valid submission are invited to send an email to wsdm-cup-2019@spotify.com to request a coupon. This email should have the title 'Coupon' and should provide the team name, and should be sent from the email associated with the account which made the valid submission. Every week a team makes an improved submission on the leaderboard, they will be eligible to request a further 100 USD coupon, for as long as coupons remain. Thus, if a team has already received a coupon, but makes an improved submission in the subsequent week starting Monday, they will be eligible for another request.

# Contact Us

Use one of the public channels:

- **Gitter Channel** : https://gitter.im/crowdAI/spotify-sequential-skip-prediction-challenge
- **Technical issues** : https://github.com/crowdAI/skip-prediction-challenge-starter-kit/issues
- **Discussion Forum** : https://www.crowdai.org/challenges/spotify-sequential-skip-prediction-challenge/topics

We strongly encourage you to use the public channels mentioned above for communications between the participants and the organisers. In extreme cases, if there are any queries or comments that you would like to make using a private communication channel, then you can send us an email at:

- wsdm-cup-2019@spotify.com
- sharada.mohanty@epfl.ch
- brianbrost@spotify.com