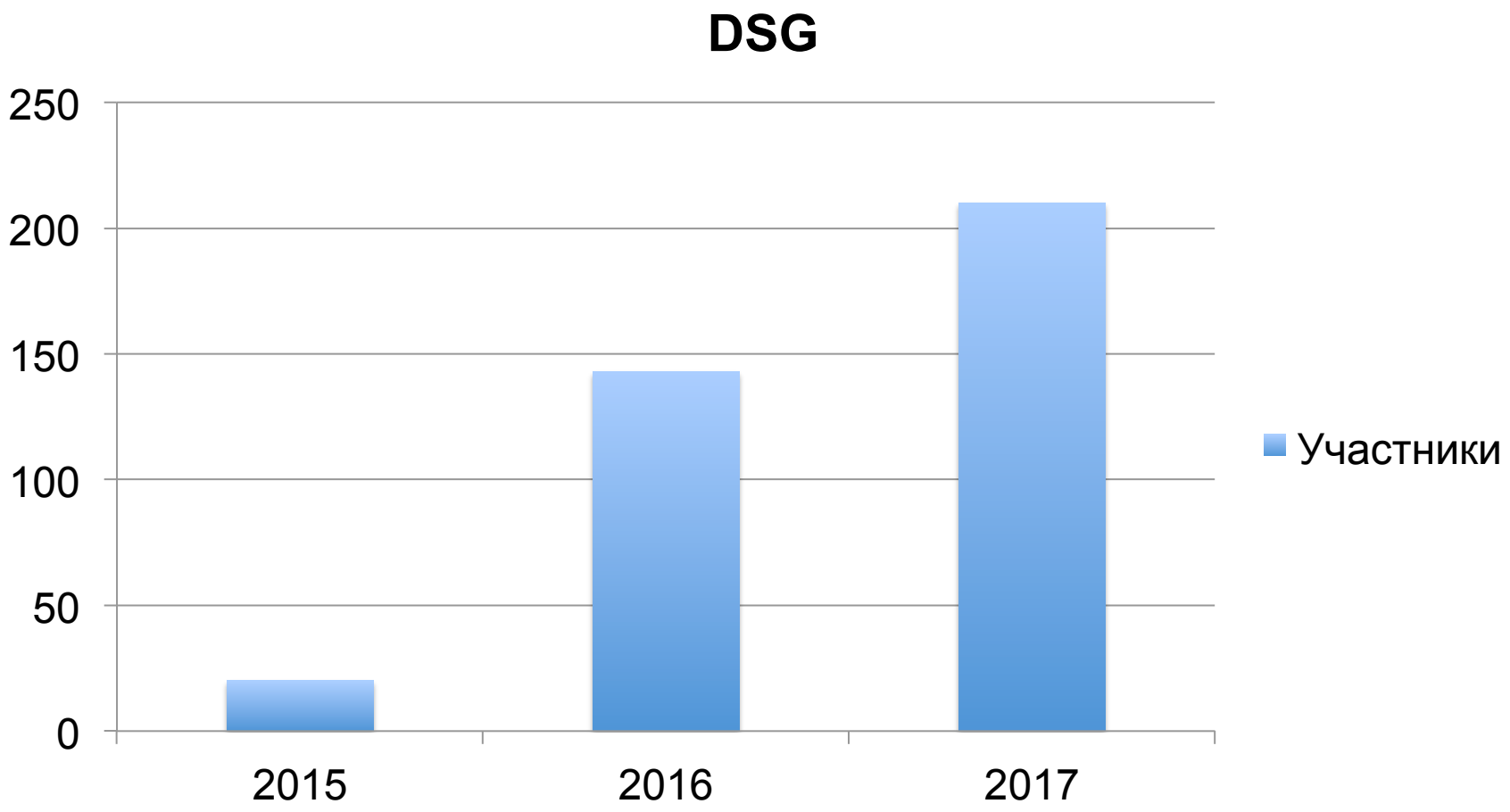
NATIONAL RESEARCH
UNIVERSITY

Data Science Game


DSG
250
200
150
100
50
0
2015
2016
2017
Участники

Формат соревнования

- Онлайн-стадия:
  - Участвуют все команды
  - Соревнование длится полтора месяца
  - Ограниченное количество сабмитов
- Финал:
  - Участвуют 20 лучших команд (но, одна команда от университета, максимум 5 команд от страны)
  - Хакатон ~30 часов.


1755


HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY


Skoltech
Сколковский институт науки и технологий
Toulouse School of Economics


INDIAN INSTITUTE OF MANAGEMENT
CALCUTTA


USP


1870


STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY


THE UNIVERSITY OF EDINBURGH


THE UNIVERSITY of EDINBURGH


Unifal MG
Universidade Federal de Alfenas


Grenoble INP
ensimag


UNIVERSIDAD NACIONAL DE INGENIERIA
SCIENTIA ET LABOR
LIMA PERU
1876


УКРАЇНСЬКИЙ КАТОЛИЦЬКИЙ УНІВЕРСИТЕТ


UNIVERSITÉ TOULOUSE III PAUL SABATIER
Université de Toulouse


HUMBOLDT-UNIVERSITÄT ZU BERLIN


Barcelona GSE


UPMC
SORBONNE UNIVERSITÉS

Онлайн-стадия

- Данные предоставлены Deezer
- Дано: история прослушиваний различной музыки (id альбома, исполнителя, жанра а также некоторые другие характеристики), скипнута ли была песня или нет, ~20000 пользователей
- Предсказать: Скипнет ли пользователь следующую песню.

Сортировка Фильтр Сводная таблица Что если Обновить Текст База данных HTML FileMaker Проверка дубликаты Группировать Разгруппировать

A3 | 25571

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	genre_id	ts_listen	media_id	album_id	context_type	release_date	platform_na	platform_far	media_durat	listen_type	user_gender	user_id	artist_id	user_age	is_listened
2	25471	1480597215	222606	41774	12	20040704	1	0	223	0	0	9241	55164	29	0
3	25571	1480544735	250467	43941	0	20060301	2	1	171	0	0	16547	55830	30	1
4	16	1479563953	305197	48078	1	20140714	2	1	149	1	1	7665	2704	29	1
5	7	1480152098	900502	71521	0	20001030	0	0	240	0	1	1580	938	30	0
6	7	1478368974	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
7	7	1478382544	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
8	7	1478338409	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
9	7	1478353709	542335	71718	1	20080215	1	0	150	1	1	10325	2939	29	1
10	7	1479130924	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
11	7	1479214304	542335	71718	1	20080215	1	2	150	1	1	51	2939	28	1
12	7	1478673075	542335	71718	0	20080215	0	0	150	0	0	1089	2939	27	1
13	7	1479279494	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
14	7	1478356603	542341	71718	1	20080215	0	0	188	1	0	822	2939	21	1
15	7	1479968326	542335	71718	1	20080215	0	0	150	1	0	2946	2939	20	0
16	7	1479991070	542335	71718	1	20080215	1	0	150	1	0	5248	2939	23	0
17	7	1478445244	542335	71718	12	20080215	2	1	150	0	0	1556	2939	25	1
18	7	1478644646	542335	71718	11	20080215	2	1	150	0	0	1709	2939	21	1
19	7	1478892170	542335	71718	14	20080215	2	1	150	0	0	336	2939	20	1
20	7	1480084581	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
21	7	1480454926	542335	71718	1	20080215	0	0	150	1	1	3570	2939	25	1
22	7	1480236062	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1
23	7	1478806000	542335	71718	0	20080215	2	1	150	0	0	1284	2939	29	1
24	7	1479409743	542335	71718	6	20080215	0	0	150	1	1	754	2939	29	1
25	7	1480113489	542335	71718	1	20080215	0	2	150	1	1	7599	2939	28	1
26	7	1480340322	542335	71718	1	20080215	0	0	150	1	1	3570	2939	25	0
27	7	1478079016	542346	71718	3	20080215	2	1	197	0	0	65	2939	26	1
28	7	1478085122	542340	71718	3	20080215	2	1	210	0	0	65	2939	26	1
29	7	1478084056	542347	71718	3	20080215	2	1	188	0	0	65	2939	26	1
30	7	1479757473	542335	71718	0	20080215	1	0	150	0	0	1709	2939	21	1
31	7	1480239881	542335	71718	0	20080215	0	0	150	0	0	1089	2939	27	0
32	7	1480081887	542335	71718	6	20080215	2	1	150	1	0	551	2939	26	1
33	7	1479830337	542335	71718	1	20080215	2	1	150	1	0	242	2939	29	1
34	7	1479953799	542335	71718	1	20080215	0	0	150	1	1	3570	2939	25	1
35	7	1478037666	542335	71718	5	20080215	2	1	150	1	0	1325	2939	21	1
36	7	1478948772	542335	71718	10	20080215	2	1	150	0	0	1521	2939	28	1
37	7	1478956898	542335	71718	6	20080215	0	0	150	1	1	754	2939	29	1
38	7	1480068039	542335	71718	5	20080215	2	1	150	1	0	242	2939	29	1
39	7	1480514658	542335	71718	0	20080215	0	0	150	0	1	1812	2939	24	1

Столбец	Объяснение
genre_id	Идентификатор жанра
media_id	Идентификатор песни
album_id	Идентификатор альбома
artist_id	Идентификатор исполнителя
user_id	Идентификатор пользователя
ts_listen	Время начала прослушивания
context_type	Контекст прослушивания
release_date	Дата выпуска песни
platform_name	Операционка
platform_family	Компьютер/планшет/…
media_duration	Длительность песни
listen_type	Тип прослушивания
user_gender	Пол пользователя
user_age	Возраст пользователя
is_listened	Классификация

Проблемы данных

- Холодный старт
  - ~1000 пользователей имели историю из < 10 песен, 33% пользователей <100 песен.
- Неверные жанры
  - По информации с сайта жанры были совсем другими, не совпадали по цифрам и даже по одинаковым категориям.

Как решали?

- Feature Engineering
  - Создание дополнительных признаков, таких как вероятность скипа, в зависимости от жанров, исполнителей, пользователя, длительности и т.д.
- XGBoost, регрессия, индивидуальные и общие классификаторы

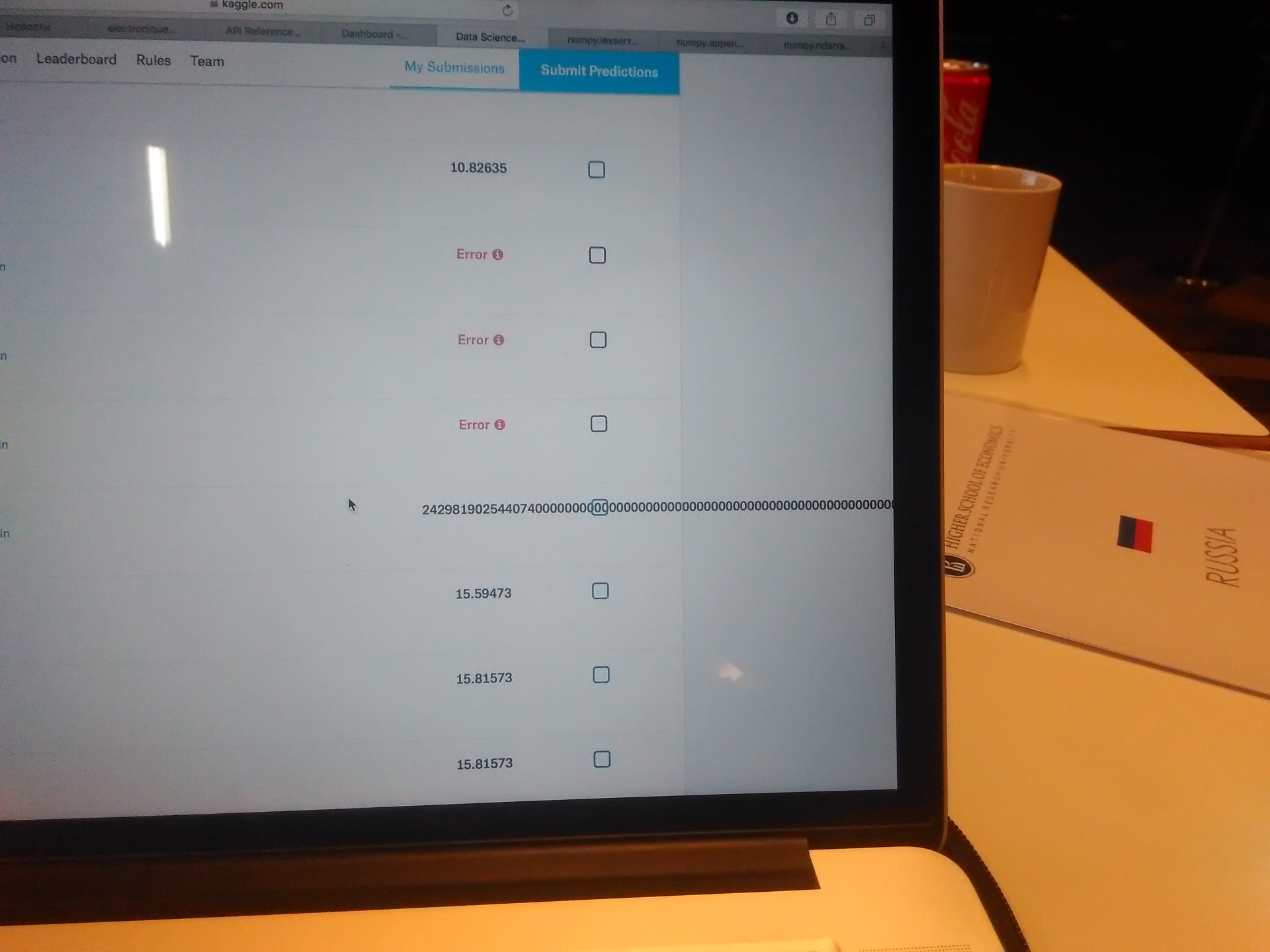
1	▼ 2	ADASE - SkolTech - Russia	1		0.68381	62	4mo
2	▲ 1	lebed i 3 raka - MSU - Russia	2		0.68341	35	4mo
3	▼ 1	DataMafia-IIMC-India			0.68237	96	4mo
4	▲ 2	Do u wanna tell about our Go...	3		0.68144	87	4mo
5	▼ 9	E3 Analytics-UNI-Peru			0.68036	121	4mo
6	▲ 1	TSEureka - TSE - France			0.68027	64	4mo
7	▼ 9	Outliers - SPbSU - Russia	4		0.67919	47	4mo
8	▼ 4	UCUpnic - UCU - Ukraine			0.67902	63	4mo
9	▼ 2	Peace Data - Skoltech - Russia	4		0.67890	98	4mo
10	▼ 13	Mean Predictors - Humboldt ...			0.67877	136	4mo
11	▲ 4	TEUBREUX - IMT Atlantique -...			0.67821	112	4mo
12	▼ 8	Rebyatishki - MIPT - Russia	5		0.67782	72	4mo

13	▲ 5	Medicinovo Inc. - Stevens Inst...		0.67730	87	4mo
14	▲ 5	The Ginger Nuts - University ...		0.67714	170	4mo
15	—	Imagouille - ENSIMAG - France		0.67706	132	4mo
16	▼ 3	Confounders		0.67543	112	4mo
17	▼ 4	UPMC LSTA - Univ. Paris 6 - F...		0.67458	99	4mo
18	▲ 1	SID - UPS - France		0.67295	100	4mo
19	▲ 6	Make Latin America Great Ag...		0.67251	81	4mo
20	▲ 14	We are Sherlocked - Universit...		0.67251	80	4mo
21	▼ 4	Ca(u)s(u)al Economists - TSE ...		0.67192	133	4mo
22	▼ 7	DeepThinkers - IIM Calcutta - ...		0.67189	81	4mo
23	▼ 4	Team Maia - USP - Brazil		0.67183	62	4mo
24	▲ 6	Lab Rats - HSE NN - Russia		0.67174	79	4mo

Your Best Entry ↑

1	▲ 1	lebed i 3 raka - MSU - Russia	1		0.68555	35	4mo
2	▲ 2	Do u wanna tell about our Go...	2		0.68422	87	4mo
3	▼ 2	ADASE - SkolTech - Russia	3		0.67886	62	4mo
4	▼ 1	DataMafia-IIMC-India			0.67767	96	4mo
5	▲ 1	TSEureka - TSE - France			0.67714	64	4mo
6	▲ 14	We are Sherlocked - Universit...			0.67501	80	4mo
7	▲ 4	TEUBREUX - IMT Atlantique -...			0.67499	112	4mo
8	▲ 5	Medicinovo Inc. - Stevens Inst...			0.67464	87	4mo
9	▲ 5	The Ginger Nuts - University ...			0.67455	170	4mo
10	▲ 16	TheFirstBrazilianSniper - Fed...			0.67372	74	4mo
11	▼ 2	Peace Data - Skoltech - Russia	3		0.67361	98	4mo
12	▼ 4	UCUpnic - UCU - Ukraine			0.67354	63	4mo
13	▲ 6	Make Latin America Great Ag...			0.67341	81	4mo
14	▼ 9	E3 Analytics-UNI-Peru			0.67310	121	4mo
15	—	Imagouille - ENSIMAG - France			0.67296	132	4mo
16	▼ 9	Outliers - SPbSU - Russia	4		0.67236	47	4mo
17	▲ 1	SID - UPS - France			0.67207	100	4mo
18	▲ 6	Lab Rats - HSE NN - Russia	5		0.67193	79	4mo


ASUS
DELL


kaggle.com
Leaderboard
Rules
Team
My Submissions
Submit Predictions
10.82635
Error
Error
Error
15.59473
15.81573
15.81573
HIGHER SCHOOL OF ECONOMICS
RUSSIA