



# MUSIC RECSYS HSE PT.1

Герасимов Владислав  
Юркин Андрей  
Поздняков Арсений

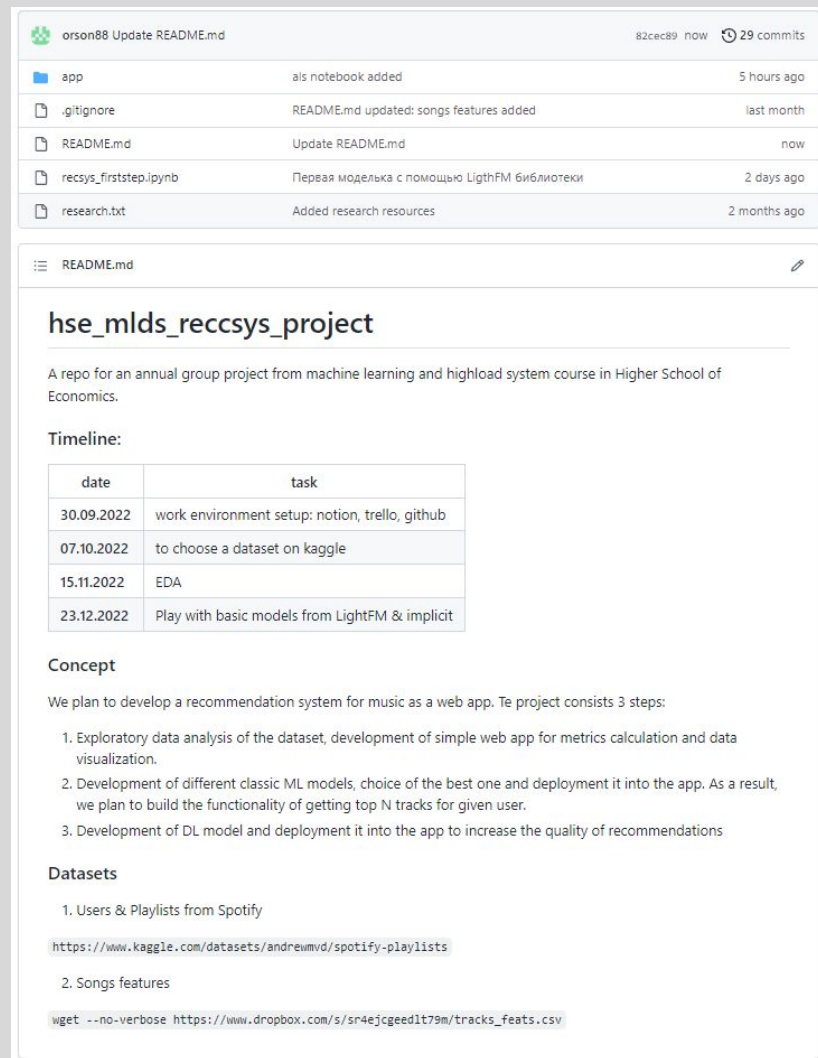
# Чекпоинты:

1. Организационный
2. Сбор данных
3. EDA
4. Базовые модели

# Организационный

- Собрались, сделали репозиторий
- Выбрали тему – музыкальные рекомендации
- Датасет – Плейлисты юзеров из спотифая и описания музыки, заскрапленные со спотифая
- Конечный формат – веб приложение

[https://github.com/orson88/hse\\_mlds\\_recsys\\_project](https://github.com/orson88/hse_mlds_recsys_project)



The screenshot shows the GitHub repository page for 'hse\_mlds\_recsys\_project' by user 'orson88'. The repository has 29 commits. The file list includes 'app', '.gitignore', 'README.md', 'recsys\_firststep.ipynb', and 'research.txt'. The 'README.md' file is open, showing the project title 'hse\_mlds\_recsys\_project', a description, a timeline table, a concept section, and datasets.

**hse\_mlds\_recsys\_project**

A repo for an annual group project from machine learning and highload system course in Higher School of Economics.

**Timeline:**

date	task
30.09.2022	work environment setup: notion, trello, github
07.10.2022	to choose a dataset on kaggle
15.11.2022	EDA
23.12.2022	Play with basic models from LightFM & implicit

**Concept**

We plan to develop a recommendation system for music as a web app. Te project consists 3 steps:

1. Exploratory data analysis of the dataset, development of simple web app for metrics calculation and data visualization.
2. Development of different classic ML models, choice of the best one and deployment it into the app. As a result, we plan to build the functionality of getting top N tracks for given user.
3. Development of DL model and deployment it into the app to increase the quality of recommendations

**Datasets**

1. Users & Playlists from Spotify  
`https://www.kaggle.com/datasets/andrewmvd/spotify-playlists`
2. Songs features  
`wget --no-verbose https://www.dropbox.com/s/sr4ejcgeedlt79m/tracks_feats.csv`

# Сбор данных

- Нашли данные, заскрапили прочие фишки со Спотифая
- Набрали статей по акустическим фишкам – данных не нашлось
- Датасет – Плейлисты юзеров из спотифая и описания музыки, заскрапленные со спотифая
- Конечный формат – веб приложение

<https://www.kaggle.com/datasets/andrewmvd/spotify-playlists>

<https://rapidapi.com/theaudiodb/api/theaudiodb/>

LARXEL · UPDATED A YEAR AGO

74

New Notebook

Download (192 MB)

## Spotify Playlists

1.2GB of tabular data for music recommendation

Data Card

Code (8)

Discussion (0)

### About Dataset

**About this dataset**

Music is ubiquitous in today's world-almost everyone enjoys listening to music. With the rise of streaming platforms, the amount of music available has substantially increased. While users may seemingly benefit from this plethora of available music, at the same time, it has increasingly made it harder for users to explore new music and find songs they like. Personalized access to music libraries and music recommender systems aim to help users discover and retrieve music they like and enjoy.

This dataset is based on the subset of users in the #nowplaying dataset who publish their #nowplaying tweets via Spotify. In principle, the dataset holds users, their playlists and the tracks contained in these playlists.

**Usability** 10.00

**License** Attribution 4.0 International (CC ...)

**Expected update frequency** Never

Request URL

rapidapi.com

REQUIRED

Header Parameters

X-RapidAPI-Key

SIGN-UP-FOR-KEY

ENUM

REQUIRED

X-RapidAPI-Host

theaudiodb.p.rapidapi.com

STRING

REQUIRED

Required Parameters

s

coldplay

STRING

REQUIRED

artist

t

yellow

STRING

REQUIRED

track name

Response

Body

the range of popularity, the song track the song track media  
ive popularity. "Yellow" has since been covered by various  
recording artists worldwide, and remains one of the band's  
most popular songs."  
strGenre: "Pop-Rock"  
strLocked: "unlocked"  
strMood: "Relaxed"  
strMusicBrainzAlbumID: "1dc4c347-a1db-32aa-b14f-bc9cc507b843"  
strMusicBrainzArtistID: "cc197bad-dc9c-440d-a5b5-d52ba2e1423  
4"  
strMusicBrainzID: "729cf505-94eb-4fbc-bc76-cbae44cff091"  
strMusicVid: "https://www.youtube.com/watch?v=yKNxeF4KMsY"  
strMusicVidCompany: "The Artists Company"  
strMusicVidDirector: " James & Alex"  
strMusicVidScreen1: "https://www.theaudiodb.com/images/media/  
track/mvidscreen/rpuypx1364211921.jpg"  
strMusicVidScreen2: "https://www.theaudiodb.com/images/media/  
track/mvidscreen/ytpqqt1364212027.jpg"  
strMusicVidScreen3: "https://www.theaudiodb.com/images/media/  
track/mvidscreen/vuqptv1364212034.jpg"  
strStyle: "Rock/Pop"  
strTheme: "In Love"  
strTrack: "Yellow"  
strTrack3DCase: null  
strTrackLyrics: ""

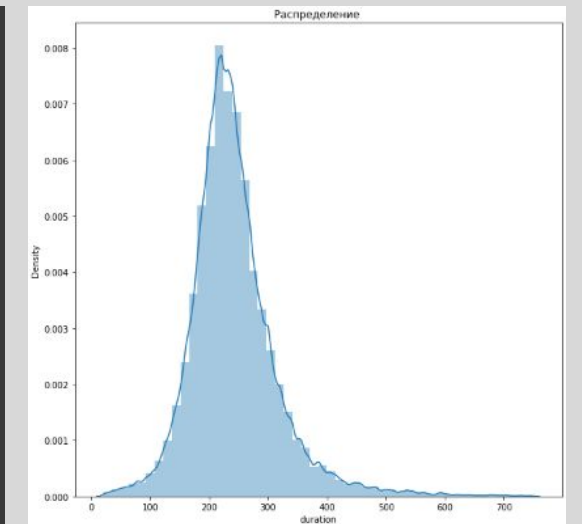
Headers

# EDA

- Принято решение выбирать топ 10000 песен (артист+песня). Получили общее представление о структуре данных.
- 16 т. юзеров
- 2.8 млн. песен

	uid	artist_x	song_name_x	playlist	num_id	artist_y	song_name_y	album	genre	mood	style	theme	duration
0	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello	(The Angels Wanna Wear My) Red Shoes	HARD ROCK 2010	751205	Elvis Costello	(The Angels Wanna Wear My) Red Shoes	My Aim Is True	Pop-Punk	NaN	NaN	NaN	170.0
1	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello & The Attractions	(What's So Funny 'Bout) Peace, Love And Unders...	HARD ROCK 2010	750786	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	9cc0cfd4d7d7885102480dd99e7a90d6	Tiffany Page	7 Years Too Late	HARD ROCK 2010	2548456	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello & The Attractions	Accidents Will Happen	HARD ROCK 2010	750791	Elvis Costello & The Attractions	Accidents Will Happen	Armed Forces	New Age	NaN	NaN	NaN	178.0
4	9cc0cfd4d7d7885102480dd99e7a90d6	Elvis Costello	Alison	HARD ROCK 2010	751243	Elvis Costello	Alison	My Aim Is True	Pop-Punk	NaN	NaN	NaN	205.0

Daft Punk	37640	Starred	1358091
Coldplay	33399	Liked from Radio	185894
The Rolling Stones	30929	Favoritas de la radio	31707
Queen	30837	Rock	30807
Radiohead	30738	Christmas	22751
Kanye West	28978	2014	22419
Michael Jackson	28964	2013	20425
Eminem	28842	Work	18778
JAY Z	28532	Jazz	18371
Muse	28270	Indie	18145
David Bowie	27914	everything	16443
Lady Gaga	27712	Classical	16394
Bruce Springsteen	25763	Country	15660
Bob Dylan	24677	New Playlist	15451
U2	24403	Music	15161
Beyoncé	24248	Generos	15021
Arctic Monkeys	23945	Rap	14506
Pearl Jam	23844	iPhone	13483
Rihanna	23508	Chill	13481
Nirvana	23048	2012	12968
Foo Fighters	22878	Metal	11277
Johnny Cash	22571	All Live Files	11174
David Guetta	21798	My Shazam Tracks	11097
John Mayer	20690	2015	10996
Red Hot Chili Peppers	20527	Random	10876
Name: artist_x, dtype: int64		Name: playlist, dtype: int64	

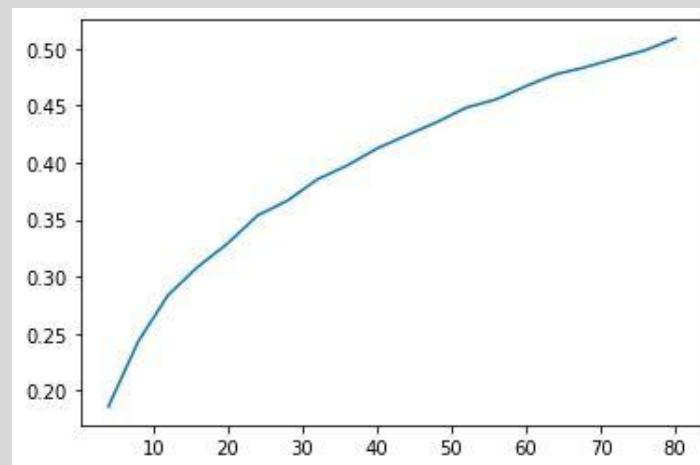


<https://colab.research.google.com/drive/1PTA-3IWY0VAryFYCLk2HJx86NWc5ofu1?usp=sharing>

# Модели

- Практиковались в LightFM и Implicit, библиотеках для recsys моделей и нужных матричных преобразований.
- Разбили трейн и тест случайно, проверив наличие всех тест-юзеров в трейне
- Выбрали Precision@K

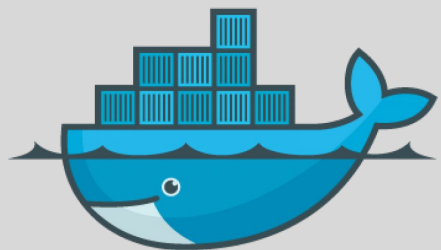
Precision@K = 0.55



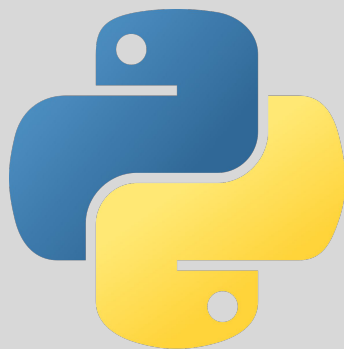
<https://colab.research.google.com/drive/1PTA-3IWY0VAryFYCLk2HJx86NWc5ofu1?usp=sharing>

# Планы

- Сервис на FastAPI
- Использовать доп. Фичи
- Другие модели



docker



Спасибо за внимание!