

цифровой  
прорыв

сезон: III

# КЕЙС

RUTUBE



Предсказание социально-  
демографических  
характеристик  
пользователя



Министерство  
экономического развития  
Российской Федерации

РОССИЯ –  
СТРАНА  
ВОЗМОЖНОСТЕЙ

# Кейсодержатель

RUTUBE

## 01 Сфера деятельности

Видеохостинг

## 02 Краткое описание кейса

С помощью истории просмотров предсказать возраст, пол пользователя

### → Сайт организации

<https://rutube.ru/>

# Постановка задачи

Пользователи RUTUBE не всегда оставляют свои данные, тип возрасту и пола на платформе, что усложняет выдачу им рекомендаций. Чтобы улучшить пользовательский опыт на видеоплатформе, необходимо предсказать социально-демографические характеристики на основе данных пользователей, которые заполнили анкету.

## Проблематика

Пользователи видеоплатформы либо не регистрируются, либо не оставляют своих данных, которые могут помочь в формировании рекомендаций.

# Решение

Прототип системы, принимающей на вход id пользователя и выдающей на выходе его социально-демографические характеристики.



# Стек технологий, рекомендуемых к использованию

## 01

*Язык программирования - Python*

*Библиотеки для использования - scikit-learn,  
библиотеки градиентного бустинга, torch, tensorflow,  
pandas, seaborn.*

*Обязательные условия - решение должно работать  
без доступа к интернету*

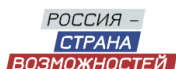
# Необходимые данные, дополнения, пояснения, уточнения

## 02

*У пользователей схожих социо-демографических групп могут быть схожие просмотры видео, их могут интересовать одни и те же авторы/категории. Интересно происследовать эту тематику и на основе данных пользователей, по которым известны пол и возраст сформировать модель, которая для любого пользователя выдаст его вероятные параметры.*



Министерство  
экономического развития  
Российской Федерации



цифровой  
прорыв 

сезон: **ИИ**





# Оценка

→ Для оценки решений применяется метод экспертных оценок и автоматизированные средства оценивания.

→ Жюри состоит из отраслевых и технических членов жюри.

→ На основании описанных далее характеристик, жюри выставляет оценки.

Возможность скачивания тестового датасета и отправки сабмитов открывается за 2 часа до стоп-кода.  
Интервал успешных отправок: 10 минут.

→ Итоговая оценка определяется как сумма баллов всех членов жюри, суммируемая с оценкой автоматизированной системы, нормализованной в 50% от итоговой оценки.

# Технический член жюри оценивает решение по следующим критериям:

## 01

Документация и комментарии к коду

Шкала: 0-1-2-3

## 02

Обоснованность выбранного метода  
(описание подходов к решению, их  
обоснование и релевантность задаче)

Шкала: 0-2-4-6

## 03

Проведение разведочного анализа данных  
(EDA)

Шкала 0-2-4-6

## 04

Выступление команды (умение  
презентовать результаты своей работы,  
строить логичный, понятный и интересный  
рассказ для презентации результатов своей  
работы)

Шкала 0-1-2

Автоматизированные средства оценивания точности  
работы предложенных участниками алгоритмов  
(решений) выставляют оценку в диапазоне 0-1, где 1  
равно 100% точности работы решения.

Итоговая оценка определяется как сумма баллов всех  
членов жюри, суммируемая с оценкой  
автоматизированной системы, нормализованной в 50%  
от итоговой оценки.

Метрика: взвешенная F1 по полу и возрасту



# Отраслевой член жюри оценивает решение по следующим критериям:

## 01

Релевантность поставленной задаче (команда погрузилась в отрасль, проблематику, предложенное решение соответствует поставленной задаче, проблема и решение структурированы)

Шкала 0-1-2-3

## 02

Проведение исследования оценки качества модели в зависимости от количества просмотров пользователя


Шкала 0-2-4-6

## 03

Выступление команды (умение презентовать результаты своей работы, строить логичный, понятный и интересный рассказ для презентации результатов своей работы)

Шкала 0-1-2



цифровой  
прорыв 

сезон: III



Министерство  
экономического развития  
Российской Федерации

РОССИЯ –  
СТРАНА  
ВОЗМОЖНОСТЕЙ

