

# Детектирование неработающих видео по статистике просмотров

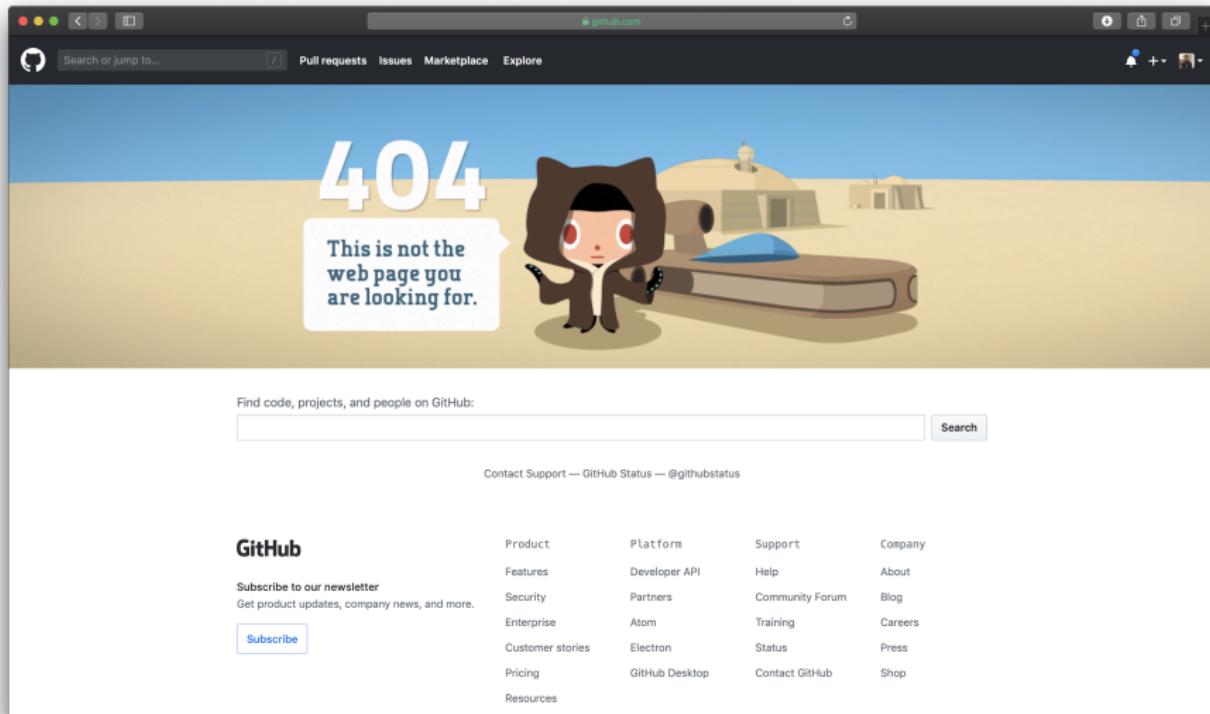
Автор: Шиянов В.А.

Научный руководитель: к.ф.-м.н. Протасов С.И.

МФТИ

25 июня 2019 г.

# Soft 404 страницы



A screenshot of a web browser displaying GitHub's 404 error page. The page has a blue header bar with the GitHub logo and navigation links: Pull requests, Issues, Marketplace, and Explore. Below the header is a large 404 error message with a cartoon illustration of a character in a brown hooded cloak standing next to a large book in a desert setting. A speech bubble from the character says, "This is not the web page you are looking for." At the bottom of the page is a search bar with placeholder text "Find code, projects, and people on GitHub:" and a "Search" button. The footer contains links for GitHub Status, Support, and Company information.

github.com

Search or jump to...

Pull requests Issues Marketplace Explore

# 404

This is not the web page you are looking for.

Find code, projects, and people on GitHub:

Contact Support — GitHub Status — @githubstatus

**GitHub**

Subscribe to our newsletter  
Get product updates, company news, and more.

[Subscribe](#)

	Product	Platform	Support	Company
Features	Developer API	Help	About	
Security	Partners	Community Forum	Blog	
Enterprise	Atom	Training	Careers	
Customer stories	Electron	Status	Press	
Pricing	GitHub Desktop	Contact GitHub	Shop	
Resources				

# Неработающие видео

Яндекс

Бразилия - боливия 3-0 обзор матча кубок америки 15.06.: Найти

Будьте в Плюсе +

Вадим Шиянов

Поиск Картинки Видео Карты Маркет Новости Эфир Коллекции Знатоки Услуги Еще

Источник видео: youtube.com

Отправить на устройство

Video unavailable

This video is no longer available due to a copyright claim by National sports channel LLC.

Бразилия - боливия 3-0 обзор матча кубок америки 15.06.2019

... Боливия - Бразилия 0-3 Обзор Матча Brazil VS Bolivia 3:0 Bolivia VS Brazil 0:3,Copa America 2019. 19,9 тыс. просмотров

Кубок Америки по футболу 2016 Аргентина - Боливия 3-0 (14 июня 2016). Обзор матча

Tags video: Футбол, Евро 2016, Аргентина, Боливия, Футбол, Евро 2016, Аргентина, Боливия, Кубок...

Бразилия - Боливия 3:0 Обзор матча 15/06/2019

Бразилия - Боливия 3-0 - Обзор Матча Кубок Америки 15/06/2019 HD 🔥 Бразилия Боливия Бразилия... 4221 просмотр

Аргентина 3-0 Боливия - Кубок Америки 2016 - 3-й тур - Обзор матча

Аргентина 3-0 Боливия - Кубок Америки 2016 - 3-й тур - Обзор матча. Аргентина 3-0 Боливия... 1194 просмотра

Brazil vs Bolivia 3-0 Бразилия 3-0 Боливия Обзор матча Match review and ALL Goals...

бразилия боливия, бразилия боливия 3-0, бразилия боливия обзор матча, неймар, коутинью, кубок...

Аргентина 3-0 Боливия | Кубок Америки 2016 | Обзор матча HD...

Сборная России вновь обыграла Турцию в матче

Слуга народа - Сезон 1 - Серия 1

Неймара жалится

Бразилия камерун обзор матча HD!

Вideoобзор матча 1/4 финала ЧМ-2018 по

Бразилия - Катар 2-0 - 05:20 HD

Чили - Аргентина 0-0 ( пен. ) 04:44 HD

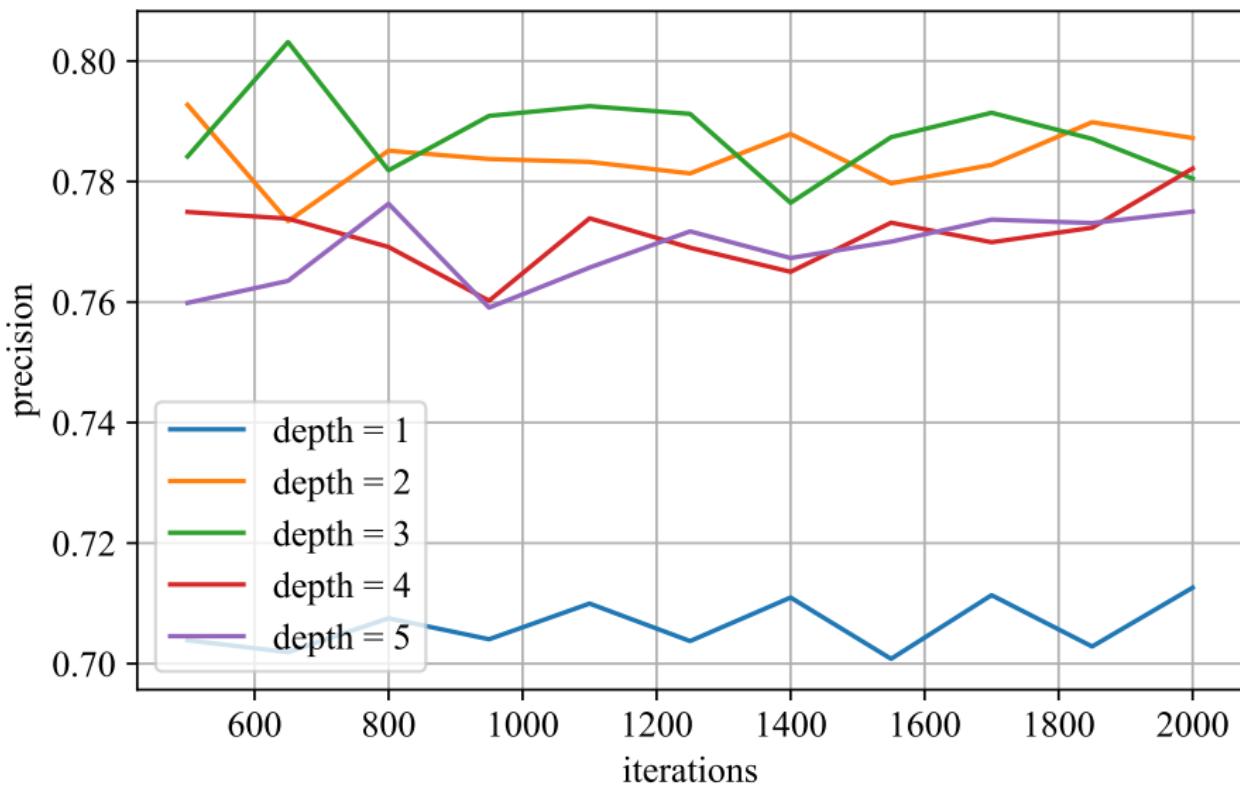
Аргентина - Никарагуа 1-1 11:35

Аргентина Боливия 2:0 05:56 HD

# Статистика просмотров

Плеер	Видео	Длительность, с	Просмотр, с	Старт	Ошибка
Плеер1	Видео1	100	10	True	
Плеер2	Видео2	200	180	True	
Плеер3	Видео3	0	5		True
...	...	...	...	...	...

# Обучение модели



# Обучение модели

Модель	Точность, %	Полнота, %
RandomForestClassifier, 500 деревьев	72.0	50.8
ExtraTreesClassifier, 1400 деревьев	74.3	47.0
ExtraTreesClassifier, 1100 деревьев	73.9	47.2
CatBoostClassifier, 650 деревьев глубины 3	<b>80.3</b>	51.4
CatBoostClassifier, 500 деревьев глубины 4	77.5	<b>53.2</b>

## Анализ модели

События плеера имеют большой вес в результирующей модели.  
Попробуем добавить идентификатор плеера в признаки.

Модель	Точность, %	Полнота, %
CatBoostClassifier, 1250 деревьев глубины 2	<b>72.6</b>	65.0
CatBoostClassifier, 800 деревьев глубины 5	71.3	<b>68.1</b>

## Одноплеерная модель

Попробуем обучить модель на данных единственного плеера.  
Для разметки воспользуемся историей обхода.

Модель	Точность, %	Полнота, %
CatBoostClassifier, 1550 деревьев глубины 3	<b>89.3</b>	51.4
CatBoostClassifier, 1850 деревьев глубины 5	88.2	<b>54.7</b>

## Дальнейшие планы

- внедрение одноплеерной модели, как вспомогательного для обхода механизма, анализ и последовательное улучшение точности и полноты ее предсказания;
- изучение других возможностей предсказывать неработающие видеозаписи по статистике просмотров.

# Заключение

- была изучена предметная область;
- было предложено использовать статистику просмотров для детектирования неработающих видео;
- был предложен ряд моделей машинного обучения, удалось добиться точности в 80.3 % и полноты в 51.4 %;
- был предложен одноплеерный подход, с помощью которого удалось достичь точности в 89.3 % и полноты в 51.4 %.