

Решение по определению цифрового портрета аудитории в мобильной среде

Команда: DST-OFF

2-4 декабря 2021

Тизер



Наше решение позволяет определить распределение пользователей по сегментам на основе неполной и косвенной информации.

Стек решения: gensim, streamlit, sklearn

Уникальность: используется тематическое моделирование для предсказания распределения по сегментам сразу по совокупности данных

Стоимость разработки: 3 млн рублей

Сроки разработки: пилотная версия - 6 месяцев

Описание кейса

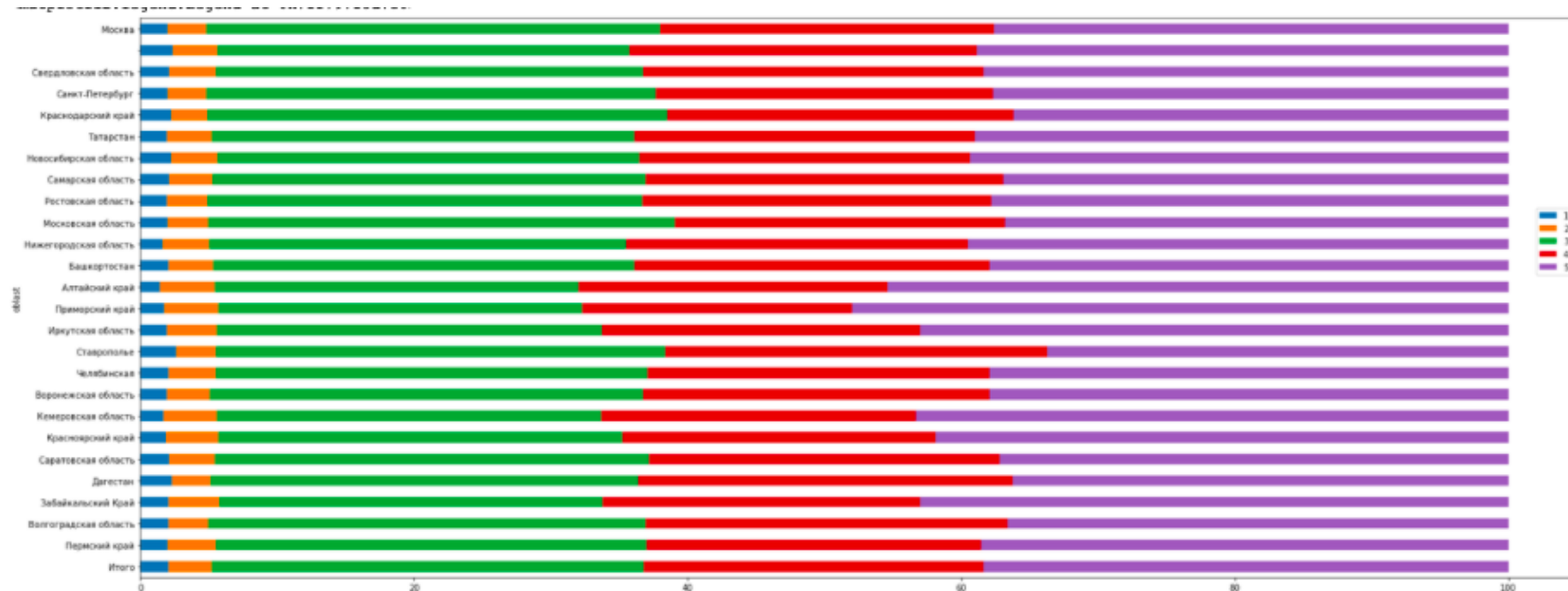
Задачи:

- Обучить классификатор предсказывать сегмент по признакам на тренировочной выборке и проверить работу классификатора на тестовой выборке, в которой отсутствуют метки сегментов;
- Используя таблицу признаков, кластеризовать пользователей и сравнить найденные кластеры с сегментами, представленными в Dataset.

Данные для исследования:

- Сегменты целевых аудиторий — они параметризуются социально-демографическими атрибутами и/или тематическими интересами;
- История активности пользователей в мобильных приложениях — это идентификатор приложения, время и географическая локация;
- Панелисты — каждый панелист принадлежит хотя бы одному сегменту, а действия панелиста размечены в истории активности его сегментами;
- Идентификаторы мобильных устройств, пользователей и панелистов в модельных данных нет.

Прототип (демонстрация работы решения)





Фи́чи

- 1. Просмотр видео – слово, корпус текстов подается на модель для тематического моделирования. В роли тем выступают сегменты
- 2:
- 3

Проблематика (проблемы и варианты решений)

Заявленные и выявленные проблемы:

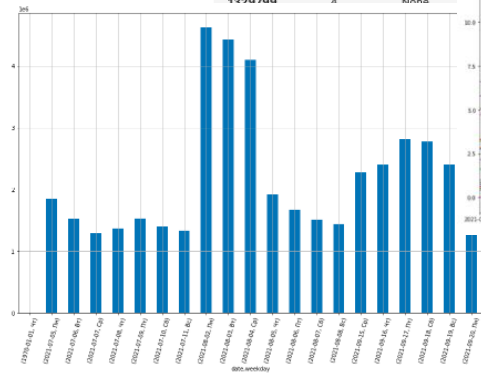
- Изменение разбиения на сегменты от периода к периоду

Функционал решения проблем:

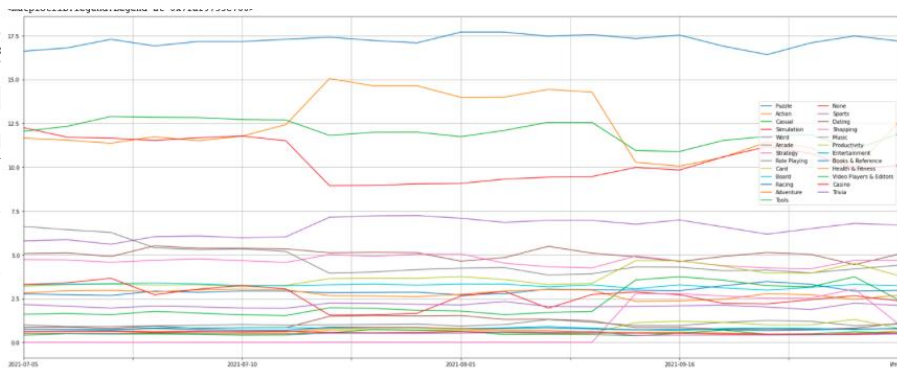
- Формирование отдельной модели для каждого периода / поднабора данных с разными принципами разбиения на сегменты

Аналитика (данные, логика, анализ)

- Используемые данные:
- Логика:
- Проведённый анализ



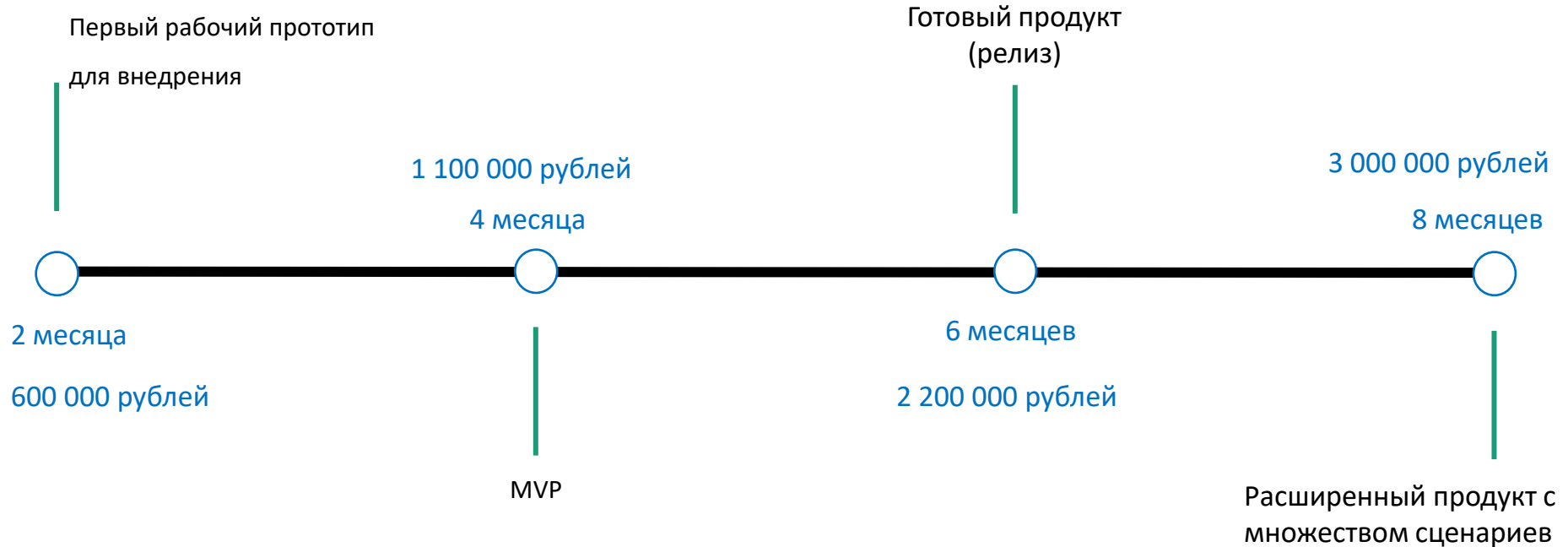
Segment	gamecategory	sut
444051	4	None
826581	4	None
835484	3	None
1550700	4	None



Region	2021-07-05	2021-07-10	2021-07-15	2021-07-20	2021-07-25	2021-07-30	2021-08-05	2021-08-10	2021-08-15	2021-08-20	2021-08-25	2021-08-30	2021-09-05	2021-09-10
Москва	2.16	2.17	2.18	2.19	2.20	2.21	2.22	2.23	2.24	2.25	2.26	2.27	2.28	2.29
Свердловская область	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Санкт-Петербург	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Краснодарский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Татарстан	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Новосибирская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Самарская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Московская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Ростовская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Нижегородская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Приморский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Алтайский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Башкортостан	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Иркутская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Ставрополье	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Челябинская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Красноярский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Кемеровская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Воронежская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Саратовская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Забайкальский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Хабаровский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Пермский край	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Волгоградская область	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Дагестан	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Итого	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

	month			7			8			9				
Segment	1	3	4	3	4	5	1	2	3	4	5			
oblast														
Москва	6.5	57.0	36.5	24.4	27.5	48.2	1.3	6.6	27.3	11.6	51.2			
Свердловская область	7.2	52.1	40.8	22.1	27.3	50.6	1.4	10.1	28.5	10.8	49.2			
Санкт-Петербург	6.6	55.8	37.6	24.4	27.6	48.0	1.4	6.7	27.7	11.3	50.8			
Краснодарский край	7.0	54.5	38.5	25.0	26.9	48.2	1.4	9.1	28.7	11.4	49.4			
Татарстан	7.0	52.0	40.9	23.2	27.9	48.9	1.4	9.5	28.8	11.7	48.6			
Новосибирская область	7.2	52.9	39.9	20.0	26.5	53.5	1.2	10.4	27.6	9.1	51.7			
Самарская область	6.9	52.7	40.4	22.6	28.3	49.1	1.5	9.9	28.6	12.4	47.7			
Московская область	6.5	59.9	33.6	25.7	28.5	45.8	1.4	8.3	27.5	13.1	49.7			
Ростовская область	6.9	55.1	38.0	23.7	28.3	48.0	1.5	9.5	28.2	12.9	47.9			
Нижегородская область	6.8	54.6	38.6	23.5	29.0	47.6	1.5	9.7	28.7	13.2	47.0			
Приморский край	7.6	53.0	39.3	17.4	22.4	60.2	1.0	10.4	24.8	7.2	56.6			
Алтайский край	7.3	52.9	39.7	19.1	27.1	53.8	1.3	10.5	27.1	10.7	50.5			
Башкортостан	7.1	52.8	40.1	21.6	28.9	49.5	1.5	9.7	29.3	13.2	46.3			
Иркутская область	7.4	52.8	39.7	17.9	26.2	55.9	1.1	10.3	26.4	10.0	52.2			
Ставрополье	7.7	52.1	40.3	23.3	29.3	47.5	1.5	9.9	28.3	13.9	46.4			
Челябинская	6.8	54.5	38.6	22.1	28.6	49.2	1.5	9.8	28.9	12.5	47.2			
Красноярский край	7.0	55.0	38.0	19.7	27.3	53.0	1.3	10.0	27.8	10.5	50.4			
Кемеровская область	7.2	54.1	38.7	19.3	27.2	53.5	1.4	10.4	27.5	10.8	50.1			
Воронежская область	7.0	54.2	38.8	23.3	28.2	48.5	1.5	9.5	29.1	12.6	47.3			
Саратовская область	7.0	54.0	39.0	22.7	28.4	48.9	1.6	10.0	28.5	12.6	47.3			
Забайкальский Край	7.4	51.4	41.2	18.1	25.9	56.1	1.2	10.6	25.8	8.7	53.7			
Хабаровский край	7.6	53.2	39.2	16.6	22.7	60.6	1.0	11.1	24.6	7.0	56.2			
Пермский край	7.0	54.8	38.1	22.6	28.1	49.4	1.5	9.8	29.2	12.6	47.0			
Волгоградская область	7.0	54.0	39.1	23.4	29.0	47.6	1.5	9.3	28.4	13.5	47.3			
Дагестан	8.0	52.5	39.5	23.3	29.0	47.7	1.3	9.6	26.6	15.2	47.4			
Итого	6.9	54.6	38.5	22.9	27.5	49.6	1.4	9.4	27.9	11.5	49.9			

Реализация решения (сроки, стоимость и порядок внедрения)



Состав команды DST-OFF



НИКОЛАЙ ГАНИБАЕВ
Капитан команды

Telegram: nganibaev

Email: ganibaev@gmail.com

Тел. +7 903 851 5919

Студент SkillFactory

- Python;
- Финансовое планирование;
- Методы моделирования процессов и программные средства для построения моделей;
- Основы программирования: типы и структуры данных;
- SQL;
- Архитектура приложений и базы данных.



АЛЕКСАНДР ГАНИБАЕВ
Аналитик, разработчик

Telegram: aganibaev

Email: acleriot@gmail.com

Тел. +7 960 111 4192

Студент ВШЭ

- C/C++;
- Python;
- Основы программирования: типы и структуры данных;
- Теория алгоритмов;
- R.



ВАСИФ ФАРАДЖОВ
Data Scientist

Telegram: valthazari

Email: vasif.faradzhov@yandex.ru

Тел. +7 906 062 9089

Студент SkillFactory

- Python;
- Маркетинг и дизайн
- Project Management;
- Теория машинного обучения.