

# Автоматическая сегментация аудитории

команда: ПРОНайк

# Актуальность

Компании сталкиваются с огромными объемами данных и необходимостью быстро адаптироваться к меняющимся предпочтениям клиентов. В условиях растущей конкуренции такой инструмент поможет выделиться на рынке и эффективно реагировать на изменения, что является ключом к успеху в маркетинге и увеличению прибыли.

# Что для нас сегментация аудитории ?



визуализация  
данных с помощью  
графиков

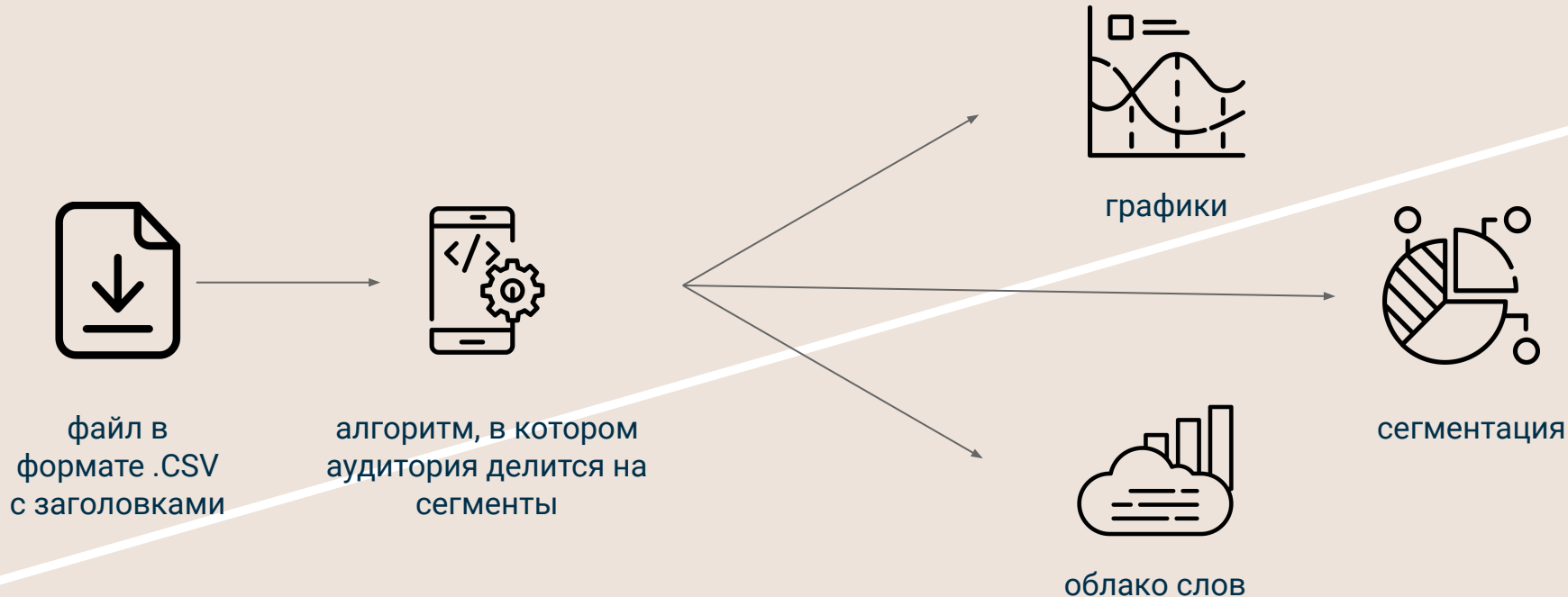


автоматическое  
разделение клиентов  
с описанием

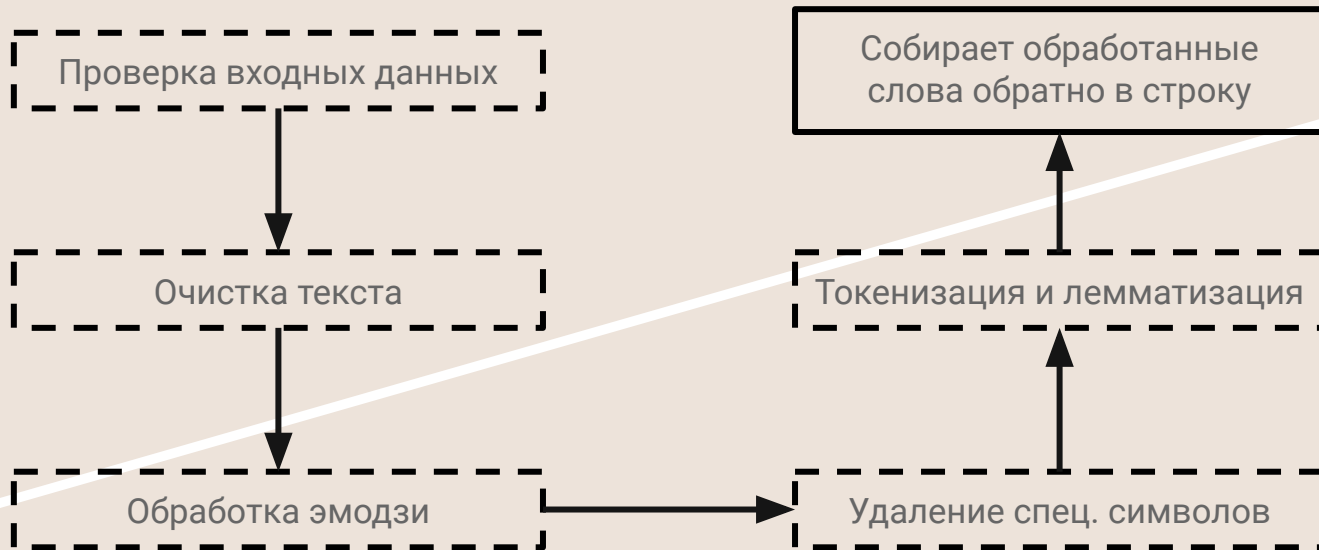


понятная и простая  
визуализация  
сегментации в виде  
облака слов

# Как работает наше решение



# Обработка данных

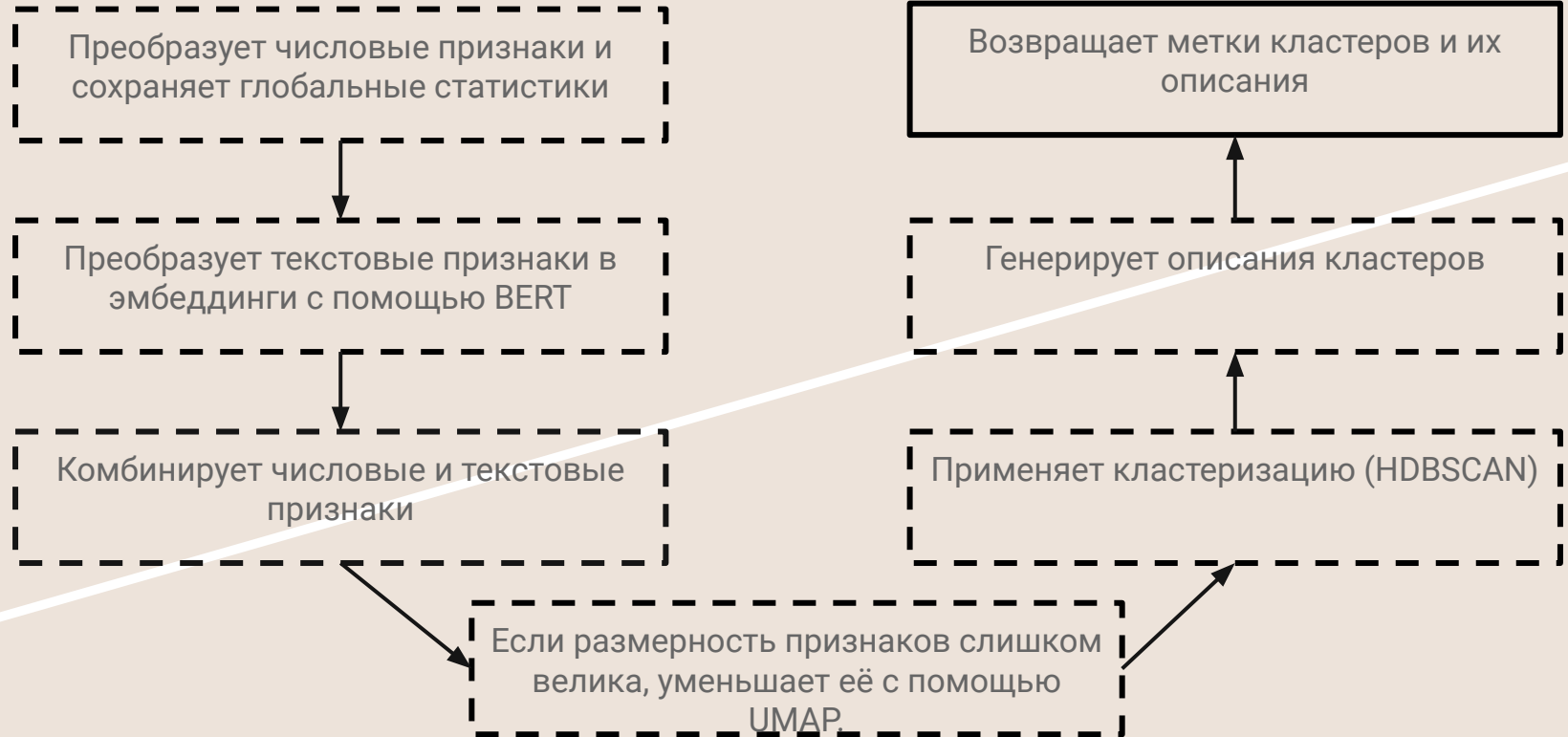


# Оценка тональности

Для анализа тональности использовалась модель `rubert-base-cased-sentiment` — предобученная языковая модель на основе BERT.

Точность этой модели может достигать 0.8-0.9 accuracy

# Сегментация аудитории



# Сегментация аудитории

## пример описания кластера

Кластер 0:

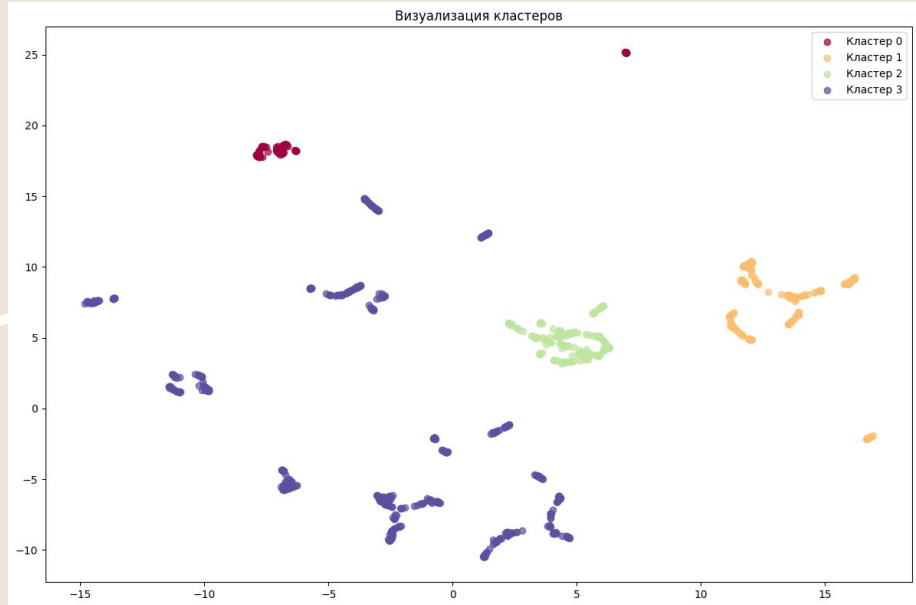
Общее описание: барнаул + ios

Размер кластера: 72 элемента

Характеристики:

- Возраст:  $42.72 \pm 16.90$
- Доход:  $92\,134.52 \pm 38\,179.24$
- Сумма покупки:  $73\,593.66 \pm 42\,051.09$
- Частота покупок:  $10.18 \pm 6.15$
- Кредитный рейтинг:  $615.35 \pm 148.88$
- Отмененные заказы:  $4.68 \pm 3.53$
- Количество жалоб:  $2.49 \pm 1.72$
- Ключевые слова: барнаул, ios, linux, сургут, самара

## пример визуализации кластера

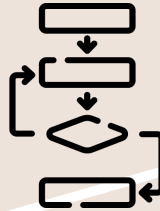




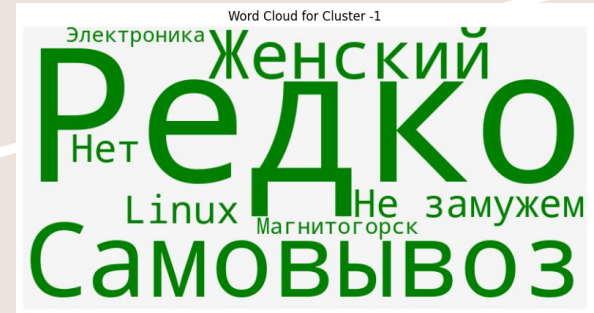
# Облако слов



сегментированная  
аудитория



алгоритм



пример облака слов,  
которое характеризует  
кластер(сегмент)

# MVP

## Инструмент для сегментации людей методами машинного обучения

Загрузите файл для начала статического анализа. Поддерживаются форматы CSV и Excel.

Обзор... Файл не выбран.

☐ Использовать тональную аналитику

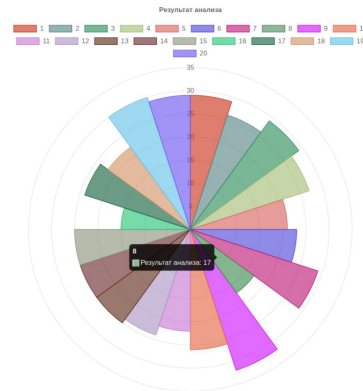
Приступить к анализу ->

© 2024 Анализатор данных

## Статистическая информация по файлу

Выберите колонку для анализа:

- ☐ Пол
  - ☐ Возраст
  - ☐ Город
  - ☐ Доход
  - ☐ Любимый продукт
  - ☐ Сумма покупки
  - ☒ Частота покупок
  - ☐ Семейное положение
  - ☐ Есть дети
  - ☐ Кредитный рейтинг
  - ☐ Отмененные заказы
  - ☐ Частота посещений магазина
  - ☐ Количество жалоб
  - ☐ Предпочитаемый способ доставки
  - ☐ Операционная система
  - ☐ Отзывы
- Выберите тип графика: PolarArea

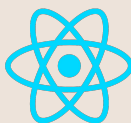


Перейти к анализу методами машинного обучения

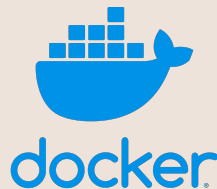
Вернуться назад

© 2024 Анализатор данных

Стек технологий:



React



Flask

# Начинаем кластеризовать данные

Выберите колонки для анализа:

Выбрать все

☒ Пол

☐ Возраст

☐ Город

☐ Доход

☐ Любимый продукт

☒ Сумма покупок

☒ Частота покупок

☒ Семейное положение

☒ Есть дети

☐ Кредитный рейтинг

☐ Отмененные заказы

☐ Частота посещений магазина

☐ Количество жалоб

☒ Предпочитаемый способ доставки

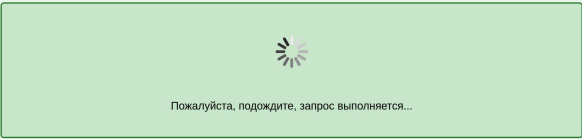
☐ Операционная система

☐ Отзывы

Запустить анализ

Вернуться назад

# Инструмент для сегментации людей методами машинного обучения



Показать кластеры

Вернуться назад

# Инструмент для сегментации людей методами машинного обучения

Word Cloud for Cluster 0

Курьерская доставка  
Товары для животных  
Мужской Ежемесячно  
Linux Да Женский  
Не замужем Самовывоз  
Android Еженедельно

Word Cloud for Cluster 1

Windows

Word Cloud for Cluster 2

Редко

Word Cloud for Cluster 3

Да

Вернуться назад

# Стоимость внедрения

Срок внедрения: 4-5 месяцев

**Data Scientists** (2 человека):

$300,000 * 5 \text{ месяцев} = 1,500,000$  —  
**2,000,000 рублей.**

**Backend Developers** (1-2 человека):

$120,000 * 5 \text{ месяцев} = 600,000$  —  
**1,500,000 рублей.**

**Frontend Developers** (1-2 человека):

$100,000 * 5 \text{ месяцев} = 500,000$  —  
**1,300,000 рублей.**

**Product Manager** (1 человек):

$150,000 * 5 \text{ месяцев} = 750,000$  —  
**1,000,000 рублей.**

**QA Engineer** (1 человек):

$80,000 * 5 \text{ месяцев} = 400,000$  —  
**600,000 рублей.**

**UI/UX Designer** (1 человек):

$100,000 * 5 \text{ месяцев} = 500,000$  —  
**650,000 рублей.**

**Технологические расходы:**

- Минимум: **600,000 рублей.**
- Максимум: **1,150,000 рублей.**

**Операционные расходы:**

- Минимум: **300,000 рублей.**
- Максимум: **700,000 рублей.**

**Общие затраты на внедрение  
продукта:**

- Минимум: **5,300,000 рублей.**
- Максимум: **9,700,000 рублей.**

# Бизнес-модель

## Гибридная модель (B2B + B2C)

### B2B

#### Преимущества:

- **Целевая аудитория:** Компании, которые работают с большими объемами данных и нуждаются в анализе текстов, кластеризации информации, сегментации клиентов, улучшении рекомендательных систем и других задачах.
- **Подходящие отрасли:** Маркетинг, e-commerce, финансовый сектор, аналитика, здравоохранение, медиасектор.

### B2C

#### Преимущества:

- **Целевая аудитория:** Пользователи, заинтересованные в улучшении личной аналитики, исследовании текстовых данных, создании персонализированных рекомендаций и т. д.
- **Подходит для стартапов, фрилансеров, исследователей данных, писателей и маркетологов.**

## Основные каналы монетизации:

- **Платная подписка**

Пример расчета для B2B-модели с подпиской:

1. **Целевая аудитория:** 50 компаний.
2. **Средняя стоимость подписки:** 250,000 руб. в месяц.
3. **Прогнозируемый доход за год:** 50 компаний × 250,000 руб. × 12 месяцев = 150,000,000 руб.

Для B2C-модели:

1. **Целевая аудитория:** 5,000 пользователей.
2. **Средняя стоимость подписки:** 1,500 руб. в месяц.
3. **Прогнозируемый доход за год:** 5,000 пользователей × 1,500 руб. × 12 месяцев = 90,000,000 руб.

# Ссылки



QR-code на  
ShowCase



QR-code на GitHub  
репозиторий