

Создание модели машинного обучения для предсказания пола пользователя по активности в интернете

Команда 6



Декабрь 2023

Введение

Разработан подход с применением лингвистической модели с целью предсказания пола пользователя использующего сервис на основе его поведения на сайте и истории заказов. Такая модель поможет компании лучше понимать свою целевую аудиторию и улучшать стратегию продаж.

Команда проекта

- Бобров Александр
- Бобров Максим
- Богомольный Александр
- Каприелов Юрий
- Макаров Александр
- Шкурихин Михаил

Описание решения

Постановка задачи

- Задачу рассматривали как классификацию текста.
- Данные по активности пользователей на сайте рассматривались как текст.
- Данный текст очищался от лишних символов и передавался в лингвистические модели.
- Модели использовали «из коробки».

Рассматриваемые модели

- BERT
- RoBERTa
- BART
- Обучения трансформера «с нуля»

Описание решения

Технологии и инструменты

- Вычислительные мощности
- Предобученные модели
- PyTorch



Результаты на валидационной выборке

Модель	Accuracy
BERT	0.76
RoBBERTa	0.79
DistilBERT	
с дополнительным преобразованием данных	0.808

Анализ результатов

- Данный подход применим к решению такого типа задач
- Возможны улучшения результатов с помощью дополнительной обработки данных и подбора гиперпараметров
- Для конечной цели поставленной заказчиком задачи может быть более интересен «стереотип поведения», чем реальный биологический пол.

Спасибо за внимание!