



Финальный проект

ЗАДАЧА

- Цель клиента - рост продаж на 20%
- Решение - создание рекомендательной системы, которая в идеальном сценарии должна дать +20% к обороту
- Метрика -- Precision@3, так как есть только 3 рекомендательные позиции на сайте
- Целевое значение метрики - это 20% к конверсии сайта (порядка 0,84%), то есть порядка Precision@3 более 1%

```
conversion = events.query('event == "transaction"').event.count() / events.query('event == "view"').event.count()
```

✓ 0.7s

Python

```
print(f'Conversion rate is {conversion:.2%}')
```

✓ 0.8s


Python

Conversion rate is 0.84%

ОПИСАНИЕ ДАННЫХ

- events csv - данные о посетителях сайта и их взаимодействиях
 - 'timestamp' - временная метка события (unix-time)
 - 'visitorid' - id посетителя
 - 'event' - тип события: view, addtocart и transaction
 - 'itemid' - id товара
 - 'transactionid' - id транзакции
- item_properties_part(1, 2).csv - данные о свойствах товаров
 - 'timestamp' - временная метка присвоения свойства
 - 'itemid' - id товара
 - 'property' - закодированное название свойства
 - 'value' - закодированное значение свойства
- category-tree.csv - дополнительные данные о соответствии категории товара (categoryid) и надкатегории (parentid)

Нужно отметить, что не все товары присутствуют в справочнике по данным о свойствах товаров (item_properties...csv)




Для построения рекомендательных моделей предлагается взять данные, где были транзакции. В качестве **user у нас будет выступать visitorid**, в качестве **item - itemid**

ВЫБРАННАЯ МОДЕЛЬ

Были проведены следующие эксперименты:

- построение рекомендация на основе данных о товарах (контекстные рекомендации) на базе FastText. Целевая метрика Precision@k получилась равной 0,12%
- матричные методы (ALS) - 0.09%
- XGBoost в режиме классификации (транзакция / не транзакция) ожидаемо неотработал из-за огромного дисбаланса классов (доля класса 1 равна конверсии сервиса)
- **коллаборативная фильтрация (LightFM) - 0.93%**



Модель LightFM показала себя значительно лучше на фоне других - **Precision@3** на уровне **0.93%** на тестовой выборке, что на 11% выше, чем конверсия сайта