

Описание проекта

Основные цели и задачи

Целями проекта было научиться прогнозировать какой товар может понравиться тому или иному пользователю. Данная рекомендательная система может повысить продажи.

Краткая информация о данных

Данные содержат историю покупок с 2002 по 2018 года.

Основные (на мой взгляд) признаки:

- Оценка, которую пользователь поставил товару
- Имя пользователя
- Категория товара
- Дата покупки
- Цена

Этапы работы над проектом

Этапы:

1. Знакомство с данными (импорт и предпросмотр)
2. Процессинг данных (избавится от пустых значений, вылетов(*про вылеты вспомнил только сейчас, когда пишу этот документ*))
3. Фичаинженеринг (покрутить имеющиеся признаки, сделать дополнительные)
4. Обучение и валидация классификаторов
5. Обучение и валиация LightFM
6. Прогноз
7. Отправка прогноза на Kaggle
8. Проба рекомендаций внутри ноутбука

Никнейм

Konstantin_Dudarev

Ссылка на прототип

Прототип не сделан

Роль в команде

Делал все один

Вклад в команду

Работал один

Какой частью работы доволен

Так как сам курс дался тяжело, очень доволен, что все же смог разобраться в том, как работают модели прогноза. Хотя и получить профит от фич и не удалось....

Что не получилось

1. Не удалось использовать фичи
2. Не удалось познакомиться с какими либо еще библиотеками кроме lightFM (если они есть)
3. Не удалось запустить LightFM на Windows (очень долго считала)
4. Не удалось запустить NMSLIB на MacOS (падает ядро)
5. Не удалось сделать прототип

Что получилось

1. Получилось сделать программу минимум, не смотря на то, что сам курс дался с трудом.

Интересное и полезное

1. Как строятся рекомендательные системы
2. Что могут дать рекомендательные системы

Главный результат

Я его сделал.

Мысли о “холодном” старте

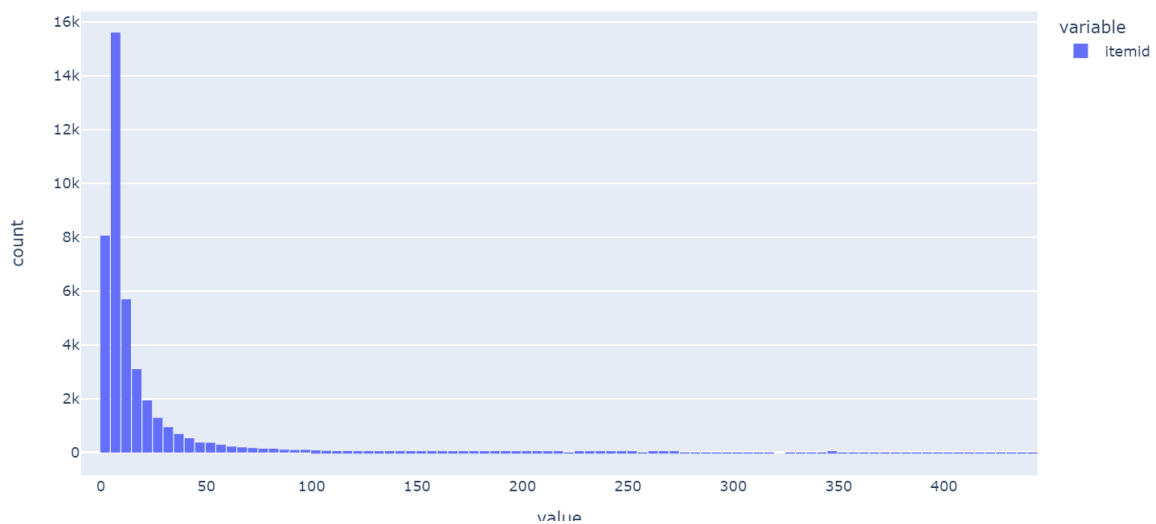
Из количества сделанных заказов можно заметить, что популярностью пользуются какие то определенные книги, при том, их довольно ограниченное количество

```
df.head(20)
```

```
22208  5491
23540  4733
24556  4654
32046  2547
5195   2300
1919   2180
379    2126
29237  1991
4689   1951
1378   1945
15137  1882
32178  1677
5462   1548
33788  1541
33913  1485
5346   1442
10833  1381
6218   1370
2602   1172
5632   1146
Name: itemid, dtype: int64
```

```
df = train['itemid'].value_counts()
```

```
fig = px.histogram(df)
fig.update_layout(bargap=0.1)
fig.show()
```



Эти первые 20 итемов пользуются наибольшей популярностью, их можно показывать всем незнакомцам.

Также, можно собирать инфу о поле, возрасте, и месте положения человека, устройства с которого он заходит чтобы делать более точные рекомендации.