Лабораторная работа №4 по курсу "Методы машинного обучения"

Выполнил: Саврасов П.А. группа ИУ5-24М

Задание

1. Выбрать произвольный набор данных (датасет), предназначенный для построения рекомендательных моделей.
2. Опираясь на материалы лекции, сформировать рекомендации для одного пользователя (объекта) двумя произвольными способами.
3. Сравнить полученные рекомендации (если это возможно, то с применением метрик).

In [64]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.metrics.pairwise import cosine_similarity
from sklearn.metrics.pairwise import euclidean_distances
from sklearn.metrics.pairwise import manhattan_distances

import warnings
warnings.filterwarnings('ignore')
```

In [65]:
```
data = pd.read_csv('amazon.csv', sep=",")
data.head()
```

Out[65]:

	id	name	asins	brand	categories	keys	manufacturer	reviews.date	reviews.dateAdded	
0	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	07T09 04-
1	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	07T09 04-
2	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	07T09 04-
3	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-13T00:00:00.000Z	2017-07-03T23:33:15Z	07T09 04-
4	AVqkIhwDv8e3D1O-lebb	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	B01AHB9CN2	Amazon	Electronics,iPad & Tablets,All Tablets,Fire Ta...	841667104676,amazon/53004484,amazon/b01ahb9cn2...	Amazon	2017-01-12T00:00:00.000Z	2017-07-03T23:33:15Z	07T09 04-

5 rows × 21 columns

Выделим необходимые для рекомендательной системы признаки:

- name - имя товара
- reviews.title - заголовок отзыва
- reviews.text - текст отзыва
- reviews.rating - рейтинг отзыва
- reviews.numHelpful - полезность отзыва
- reviews.doRecommend - рекрмендация товара в отзыве

In [66]:
```
data = data[['name', 'reviews.title', 'reviews.text', 'reviews.rating', 'reviews.numHelpful', 'reviews.doRecommend']]
data.head()
```

Out[66]:

	name	reviews.title	reviews.text	reviews.rating	reviews.numHelpful	reviews.doRecommend
0	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	Kindle	This product so far has not disappointed. My c...	5.0	0.0	True
1	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	very fast	great for beginner or experienced person. Boug...	5.0	0.0	True
2	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	Beginner tablet for our 9 year old son.	Inexpensive tablet for him to use and learn on...	5.0	0.0	True
3	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	Good!!!	I've had my Fire HD 8 two weeks now and I love...	4.0	0.0	True
4	All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi,...	Fantastic Tablet for kids	I bought this for my grand daughter when she c...	5.0	0.0	True

Всего в датасете 48 товаров

In [67]:
```
title = pd.unique(data['name'])
len(pd.unique(data['name']))
```

Out[67]: 49

In [68]:
```
data.isnull().sum()
```

Out[68]:
```
name                6760
reviews.title          5
reviews.text           1
reviews.rating        33
reviews.numHelpful   529
reviews.doRecommend  594
dtype: int64
```

Уберём записи с пустым товаром и текстом

In [69]:
```
data = data.dropna(subset=['name'])
data = data.dropna(subset=['reviews.text'])
data = data.dropna(subset=['reviews.title'])
data.isnull().sum()
```

Out[69]:
```
name                  0
reviews.title         0
reviews.text          0
reviews.rating       32
reviews.numHelpful  456
reviews.doRecommend  489
dtype: int64
```

In [70]:
```
tfidfv = TfidfVectorizer()
overview_matrix = tfidfv.fit_transform(data['reviews.text'])
overview_matrix
```

Out[70]:
```
<27896x12433 sparse matrix of type '<class 'numpy.float64'>'
	with 654698 stored elements in Compressed Sparse Row format>
```

In [71]:
```
class SimpleKNNRecommender:

    def __init__(self, X_matrix, df):
        """
        Входные параметры:
        X_matrix - обучающая выборка (матрица объект-признак)
        X_ids - массив идентификаторов объектов
        X_title - массив названий объектов
        X_overview - массив описаний объектов
        """
        #Сохраняем параметры в переменных объекта
        self._X_matrix = X_matrix
        self.df = df


    def recommend_for_single_object(self, K: int, \
                X_matrix_object, cos_flag = True, manh_flag = False):
        """
        Метод формирования рекомендаций для одного объекта.
        Входные параметры:
        K - количество рекомендуемых соседей
        X_matrix_object - строка матрицы объект-признак, соответствующая объекту
        cos_flag - флаг вычисления косинусного расстояния
        manh_flag - флаг вычисления манхэттэнского расстояния
        Возвращаемое значение: K найденных соседей
        """

        scale = 1000000
        # Вычисляем косинусную близость
        if cos_flag:
            dist = cosine_similarity(self._X_matrix, X_matrix_object)
            self.df['dist'] = dist * scale
            res = self.df.sort_values(by='dist', ascending=False)
            # Не учитываем рекомендации с единичным расстоянием,
            # так как это искомый объект
            res = res[res['dist'] < scale]

        else:
            if manh_flag:
                dist = manhattan_distances(self._X_matrix, X_matrix_object)
            else:
                dist = euclidean_distances(self._X_matrix, X_matrix_object)
            self.df['dist'] = dist * scale
            res = self.df.sort_values(by='dist', ascending=True)
            # Не учитываем рекомендации с единичным расстоянием,
            # так как это искомый объект
            res = res[res['dist'] > 0.0]

        # Оставляем K первых рекомендаций
        res = res.head(K)
        return res
```

In [72]:
```
productId = 18000
data['name'][productId]
```

Out[72]: 'Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black,,,'

In [73]:
```
prod_matrix = overview_matrix[productId]
prod_matrix
```

Out[73]:
```
<1x12433 sparse matrix of type '<class 'numpy.float64'>'
	with 53 stored elements in Compressed Sparse Row format>
```

In [74]:
```
skr1 = SimpleKNNRecommender(overview_matrix, data)
```

In [79]:
```
rec1 = skr1.recommend_for_single_object(15, prod_matrix)
for prod in pd.unique(rec1['name']):
    print(prod)
```

```
Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black,,,
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta
Kindle Voyage E-reader, 6 High-Resolution Display (300 ppi) with Adaptive Built-in Light, PagePress Sensors, Wi-Fi - Includes Special Offers,
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Black
Brand New Amazon Kindle Fire 16gb 7 Ips Display Tablet Wifi 16 Gb Blue,,,
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Magenta
```

In [80]:
```
rec2 = skr1.recommend_for_single_object(15, prod_matrix, cos_flag = False)
for prod in pd.unique(rec2['name']):
    print(prod)
```

```
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Magenta
All-New Fire HD 8 Tablet, 8 HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta
Kindle Voyage E-reader, 6 High-Resolution Display (300 ppi) with Adaptive Built-in Light, PagePress Sensors, Wi-Fi - Includes Special Offers,
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Black
Amazon Kindle Paperwhite - eBook reader - 4 GB - 6 monochrome Paperwhite - touchscreen - Wi-Fi - black,,,
Brand New Amazon Kindle Fire 16gb 7 Ips Display Tablet Wifi 16 Gb Blue,,,
```

In [81]:
```
rec3 = skr1.recommend_for_single_object(15, prod_matrix,
                                        cos_flag = False, manh_flag = True)
for prod in pd.unique(rec3['name']):
    print(prod)
```

```
Fire Tablet, 7 Display, Wi-Fi, 8 GB - Includes Special Offers, Magenta
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
Amazon 5W USB Official OEM Charger and Power Adapter for Fire Tablets and Kindle eReaders,,,
All-New Kindle E-reader - Black, 6 Glare-Free Touchscreen Display, Wi-Fi -  Includes Special Offers,,
Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Green Kid-Proof Case
```

In [ ]:
```

```