

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

Лабораторная работа № 4
по курсу «Методы машинного обучения»

ИСПОЛНИТЕЛЬ:

Мьюу Зо У
ФИО

группа ИУ5-21М

подпись

_____ 2021 г.

ПРЕПОДАВАТЕЛЬ:

ФИО

подпись

"__" _____ 2021 г.

Москва - 2021

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).

Текст программы и экранные формы

In [2]: `pip install surprise`

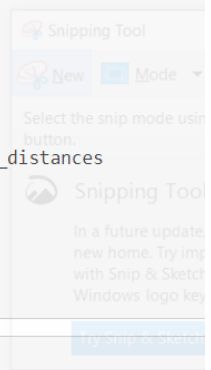
```
Collecting surprise
  Downloading https://files.pythonhosted.org/packages/61/de/e5cba8682201fcf9c3719a6fdda95693468ed061945493dea2dd37c5618b/surprise-0.1-py2.py3-none-any.whl
Collecting scikit-surprise
  Downloading https://files.pythonhosted.org/packages/97/37/5d334adaf5ddd65da99fc65f6507e0e4599d092ba048f4302fe8775619e8/scikit-surprise-1.1.1.tar.gz (11.8MB)
    11.8MB 267kB/s
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages (from scikit-surprise->surprise) (1.0.1)
Requirement already satisfied: numpy>=1.11.2 in /usr/local/lib/python3.7/dist-packages (from scikit-surprise->surprise) (1.19.5)
Requirement already satisfied: scipy>=1.0.0 in /usr/local/lib/python3.7/dist-packages (from scikit-surprise->surprise) (1.4.1)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.7/dist-packages (from scikit-surprise->surprise) (1.15.0)
Building wheels for collected packages: scikit-surprise
  Building wheel for scikit-surprise (setup.py) ... done
  Created wheel for scikit-surprise: filename=scikit_surprise-1.1.1-cp37-cp37m-linux_x86_64.whl size=1617545 sha256=700ba6d99fd71fb0fee6bea64614b120c6f11afd975a4e6a79f53a799cca40a6
  Stored in directory: /root/.cache/pip/wheels/78/9c/3d/41b419c9d2aff5b6e2b4c0fc8d25c538202834058f9ed110d0
Successfully built scikit-surprise
Installing collected packages: scikit-surprise, surprise
Successfully installed scikit-surprise-1.1.1 surprise-0.1
```

In [3]: `import numpy as np
import pandas as pd
from typing import Dict, Tuple
from scipy import stats
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.neighbors import KNeighborsRegressor, KNeighborsClassifier
from sklearn.model_selection import GridSearchCV, RandomizedSearchCV
from sklearn.metrics import confusion_matrix
from sklearn.metrics.pairwise import cosine_similarity, euclidean_distances, manhattan_distances
from surprise import SVD, Dataset, Reader
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib_venn import venn2
%matplotlib inline
sns.set(style="ticks")`

In [4]: `data=pd.read_csv('googleplaystore.csv', sep=",")`

In [5]: `#размер датасета
data.shape`

Out[5]: (10841, 13)



In [6]: data.head()

Out[6]:

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0	4.0.3 and up
1	Coloring book moana	ART_AND_DESIGN	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0	4.0.3 and up
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4	4.0.3 and up
3	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with device	4.2 and up
4	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1	4.4 and up

In [7]: # Колонки с пропусками
hcols_with_na = [c for c in data.columns if data[c].isnull().sum() > 0]
hcols_with_na

Out[7]: ['Rating', 'Type', 'Genres', 'Current Ver', 'Android Ver']

In [12]: df = data[data['Genres'].notnull()]
df = df[~df['Genres'].str.isspace()]

In [13]: App= df['App'].values
App[0:5]

Out[13]: array(['Photo Editor & Candy Camera & Grid & ScrapBook',
 'Coloring book moana',
 'U Launcher Lite – FREE Live Cool Themes, Hide Apps',
 'Sketch - Draw & Paint', 'Pixel Draw - Number Art Coloring Book'],
 dtype=object)

In [14]: Genres= df['Genres'].values
Genres[0:5]

Out[14]: array(['Art & Design', 'Art & Design;Pretend Play', 'Art & Design',
 'Art & Design', 'Art & Design;Creativity'], dtype=object)

In [16]: Category = df['Category'].values
Category[0:5]

Out[16]: array(['ART_AND_DESIGN', 'ART_AND_DESIGN', 'ART_AND_DESIGN',
 'ART_AND_DESIGN', 'ART_AND_DESIGN'], dtype=object)

In [17]: %%time
tfidf = TfidfVectorizer()
matrix = tfidf.fit_transform(Genres)
matrix

CPU times: user 46.4 ms, sys: 1.96 ms, total: 48.3 ms
Wall time: 63.4 ms

```
In [23]: class SimpleKNNRecommender:

    def __init__(self, X_matrix, X_App, X_Genres, X_Category)::
        """
        Входные параметры:
        X_matrix - обучающая выборка (матрица объект-признак)
        """

        #Сохраняем параметры в переменных объекта
        self.X_matrix = X_matrix
        self.df = pd.DataFrame(
            {'App': pd.Series(X_App, dtype='str'),
             'Genres': pd.Series(X_Genres, dtype='str'),
             'Category': pd.Series(X_Category, dtype='str'),
             'dist': pd.Series([], dtype='float')})

    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 манхэттенского 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)
```

```
In [24]: Genres[0]
```

```
Out[24]: 'Art & Design'
```

```
In [25]: mc_matrix = matrix[0]
         mc_matrix
```

```
Out[25]: <1x66 sparse matrix of type '<class 'numpy.float64'>'
         with 2 stored elements in Compressed Sparse Row format>
```

```
In [26]: skr1 = SimpleKNNRecommender(matrix, App, Genres, Category)
```

```
In [27]: rec1 = skr1.recommend_for_single_object(5, mc_matrix)
         rec1
```

```
Out[27]:
```

	App	Genres	Category	dist
23	Mcqueen Coloring pages	Art & Design;Action & Adventure	ART_AND_DESIGN	801421.625333
2111	Mcqueen Coloring pages	Art & Design;Action & Adventure	FAMILY	801421.625333
10438	Dolphin and fish coloring book	Art & Design;Creativity	FAMILY	789039.398405
26	Colorfit - Drawing & Coloring	Art & Design;Creativity	ART_AND_DESIGN	789039.398405
4	Pixel Draw - Number Art Coloring Book	Art & Design;Creativity	ART_AND_DESIGN	789039.398405

```
In [28]: # При поиске с помощью Евклидова расстояния
rec2 = skr1.recommend_for_single_object(5, mc_matrix, cos_flag = False)
rec2
```

```
Out[28]:
```

	App	Genres	Category	dist
2111	Mcqueen Coloring pages	Art & Design;Action & Adventure	FAMILY	630203.736369
23	Mcqueen Coloring pages	Art & Design;Action & Adventure	ART_AND_DESIGN	630203.736369
7027	UNICORN - Color By Number & Pixel Art Coloring	Art & Design;Creativity	FAMILY	649554.619097
26	Colorfit - Drawing & Coloring	Art & Design;Creativity	ART_AND_DESIGN	649554.619097
4	Pixel Draw - Number Art Coloring Book	Art & Design;Creativity	ART_AND_DESIGN	649554.619097

```
In [29]: # Манхэттенское расстояние
rec3 = skr1.recommend_for_single_object(5, mc_matrix,
cos_flag = False, manh_flag = True)
rec3
```

```
Out[29]:
```

	App	Genres	Category	dist
10438	Dolphin and fish coloring book	Art & Design;Creativity	FAMILY	912685.941941
9	Kids Paint Free - Drawing Fun	Art & Design;Creativity	ART_AND_DESIGN	912685.941941
7027	UNICORN - Color By Number & Pixel Art Coloring	Art & Design;Creativity	FAMILY	912685.941941
43	Paint Splash!	Art & Design;Creativity	ART_AND_DESIGN	912685.941941
26	Colorfit - Drawing & Coloring	Art & Design;Creativity	ART_AND_DESIGN	912685.941941