

Московский государственный технический университет им. Н.Э. Баумана
Кафедра «Системы обработки информации и управления»



Лабораторная работа №4
по дисциплине
«Методы машинного обучения»

Выполнил:
студент группы ИУ5-22М
Вей Пхью Ту

Москва — 2021 г.

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

```
[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")
```

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

```
[33] #размер датасета
data.shape
```

```
(10841, 13)
```

```
[34] data.head()
```

	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

```
[35] list(zip(data.columns, [i for i in data.dtypes]))
```

```
[('App', dtype('O')),  
 ('Category', dtype('O')),  
 ('Rating', dtype('float64')),  
 ('Reviews', dtype('int64')),  
 ('Size', dtype('O')),  
 ('Installs', dtype('O')),  
 ('Type', dtype('O')),  
 ('Price', dtype('O')),  
 ('Content Rating', dtype('O')),  
 ('Genres', dtype('O')),  
 ('Last Updated', dtype('O')),  
 ('Current Ver', dtype('O')),  
 ('Android Ver', dtype('O'))]
```

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

```
['Rating', 'Type', 'Genres', 'Current Ver', 'Android Ver']
```

```
[37] df = data[data['Genres'].notnull()]
df = df[~df['Genres'].str.isspace()]
```

```
[38] App= df['App'].values
App[0:5]
```

```
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)
```

```
[39] Genres= df['Genres'].values
Genres[0:5]
```

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

```
[41] Installs= df['Installs'].values
Installs[0:5]
```

```
array(['10,000+', '500,000+', '5,000,000+', '50,000,000+', '100,000+'],
      dtype=object)
```

```
[42] %%time
tfidf = TfidfVectorizer()
matrix = tfidf.fit_transform(Genres)
matrix
```

```
CPU times: user 47.6 ms, sys: 0 ns, total: 47.6 ms
Wall time: 63 ms
```

```
[43] class SimpleKNNRecommender:
```

```
    def __init__(self, X_matrix, X_Genres, X_App, X_Installs):
```

```
        """
```

```
        Входные параметры:
```

```
        X_matrix - обучающая выборка (матрица объект-признак)
```

```
        """
```

```
        #Сохраняем параметры в переменных объекта
```

```
        self._X_matrix = X_matrix
```

```
        self.df = pd.DataFrame(
```

```
            {'Genres': pd.Series(X_Genres, dtype='str'),
```

```
            'App': pd.Series(X_App, dtype='str'),
```

```
            'Installs': pd.Series(X_Installs, 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
```

```

[43]     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

```

```

[44] Genres[0]

```

```

'Art & Design'

```

```

[45] mc_matrix = matrix[0]
      mc_matrix

```

```

<1x66 sparse matrix of type '<class 'numpy.float64'>'
  with 2 stored elements in Compressed Sparse Row format>

```

```

[46] skr1 = SimpleKNNRecommender(matrix, Genres, App, Installs)

```

```
[47] rec1 = skr1.recommend_for_single_object(5, mc_matrix)
      rec1
```

	Genres	App	Installs	dist
23	Art & Design;Action & Adventure	Mcqueen Coloring pages	100,000+	801421.625333
2111	Art & Design;Action & Adventure	Mcqueen Coloring pages	100,000+	801421.625333
10438	Art & Design;Creativity	Dolphin and fish coloring book	500,000+	789039.398405
26	Art & Design;Creativity	Colorfit - Drawing & Coloring	500,000+	789039.398405
4	Art & Design;Creativity	Pixel Draw - Number Art Coloring Book	100,000+	789039.398405

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

	Genres	App	Installs	dist
2111	Art & Design;Action & Adventure	Mcqueen Coloring pages	100,000+	630203.736369
23	Art & Design;Action & Adventure	Mcqueen Coloring pages	100,000+	630203.736369
7027	Art & Design;Creativity	UNICORN - Color By Number & Pixel Art Coloring	500,000+	649554.619097
26	Art & Design;Creativity	Colorfit - Drawing & Coloring	500,000+	649554.619097
4	Art & Design;Creativity	Pixel Draw - Number Art Coloring Book	100,000+	649554.619097

```
# Манхэттэнское расстояние
rec3 = skr1.recommend_for_single_object(5, mc_matrix,
                                         cos_flag = False, manh_flag = True)
rec3
```

	Genres	App	Installs	dist
10438	Art & Design;Creativity	Dolphin and fish coloring book	500,000+	912685.941941
9	Art & Design;Creativity	Kids Paint Free - Drawing Fun	10,000+	912685.941941
7027	Art & Design;Creativity	UNICORN - Color By Number & Pixel Art Coloring	500,000+	912685.941941
43	Art & Design;Creativity	Paint Splash!	100,000+	912685.941941
26	Art & Design;Creativity	Colorfit - Drawing & Coloring	500,000+	912685.941941