МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

Лабораторная работа №4
по курсу «Методы машинного обучения»

ИСПОЛНИТЕЛЬНИЦА: Абросимова Н.Г.
ИУ5-24М

_____
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

```
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.tree import DecisionTreeClassifier, DecisionTreeRegressor,
export_graphviz
from sklearn.ensemble import RandomForestClassifier, RandomForestRegressor
from sklearn.ensemble import ExtraTreesClassifier, ExtraTreesRegressor
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")
data=pd.read_csv('Final_Dataframe.csv', sep=",")
#размер датасета
data.shape
(205, 9)
data.head()
```

	brand	laptop_name	display_size	processor_type	graphics_card	disk_space	discount_price	old_price	ratings_5max
0	HP	Notebook 14-df0008nx	14.0	Intel Celeron N4000	Intel HD Graphics 600	64 GB (eMMC)	1259.0	1259.0	0 / 5
1	Lenovo	IdeaPad 330S-14IKB	14.0	Intel Core i5-8250U	Intel UHD Graphics 620	1 TB HDD	1849.0	2099.0	3.3 / 5
2	Huawei	MateBook D Volta	14.0	Intel Core i5-8250U	NVIDIA GeForce MX150 (2 GB)	256 GB SSD	2999.0	3799.0	0 / 5
3	Dell	Inspiron 15 3567	15.6	Intel Core i3-7020U	Intel HD Graphics 620	1 TB HDD	1849.0	1849.0	0 / 5
4	Asus	VivoBook 15 X510UR	15.6	Intel Core i7-8550U	NVIDIA GeForce 930MX (2 GB)	1 TB HDD	2499.0	3149.0	0 / 5

```
# Колонки с пропусками
hcols_with_na = [c for c in data.columns if data[c].isnull().sum() > 0]
hcols_with_na
['laptop_name']
df = data[data['laptop_name'].notnull()]
df = df[~df['laptop_name'].str.isspace()]
brand= df['brand'].values
brand[0:5]
array(['HP', 'Lenovo', 'Huawei', 'Dell', 'Asus'], dtype=object)
laptop_name= df['laptop_name'].values
laptop_name[0:5]
array(['Notebook 14-df0008nx', 'IdeaPad 330S-14IKB', 'MateBook D Volta',
       'Inspiron 15 3567', 'VivoBook 15 X510UR'], dtype=object)
processor_type = df['processor_type'].values
processor_type[0:5]
```

```
array([' Intel Celeron N4000', ' Intel Core i5-8250U',
       ' Intel Core i5-8250U', ' Intel Core i3-7020U',
       ' Intel Core i7-8550U'], dtype=object)
```

```
%%time
tfidfv = TfidfVectorizer()
matrix = tfidfv.fit_transform(laptop_name)
matrix
```

```
Wall time: 14.9 ms
```

Out[10]:

```
<204x163 sparse matrix of type '<class 'numpy.float64'>'
       with 578 stored elements in Compressed Sparse Row format>
```

```
class SimpleKNNRecommender:

    def __init__(self, X_matrix, X_brand, X_laptop_name, X_processor_type):
        """
        Входные параметры:
        X_matrix - обучающая выборка (матрица объект-признак)
        """
        #Сохраняем параметры в переменных объекта
        self._X_matrix = X_matrix
        self.df = pd.DataFrame(
            {'brand': pd.Series(X_brand, dtype='str'),
            'laptop_name': pd.Series(X_laptop_name, dtype='str'),
            'processor_type': pd.Series(X_processor_type, 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 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
laptop_name[0]
'Notebook 14-df0008nx'
mc_matrix = matrix[0]
mc_matrix
<1x163 sparse matrix of type '<class 'numpy.float64'>'
        with 3 stored elements in Compressed Sparse Row format>
skr1 = SimpleKNNRecommender(matrix, brand, laptop_name, processor_type)
rec1 = skr1.recommend_for_single_object(5, mc_matrix)
rec1
```

	brand	laptop_name	processor_type	dist
65	HP	14-cf1001nx	Intel Core i5-8265U	196183.401727
163	HP	14-bp101nx	Intel Core i5-8250U	196183.401727
165	HP	14-ck0008nx	Intel Celeron N4000	196183.401727
9	HP	14-cf0007nx	Intel Core i5-8250U	196183.401727
130	Dell	Inspiron 14 5480	Intel Core i7-8565U	174852.637832

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

	brand	laptop_name	processor_type	dist
165	HP	14-ck0008nx	Intel Celeron N4000	1.267925e+06
163	HP	14-bp101nx	Intel Core i5-8250U	1.267925e+06
9	HP	14-cf0007nx	Intel Core i5-8250U	1.267925e+06
65	HP	14-cf1001nx	Intel Core i5-8265U	1.267925e+06
130	Dell	Inspiron 14 5480	Intel Core i7-8565U	1.284638e+06

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

	brand	laptop_name	processor_type	dist
65	HP	14-cf1001nx	Intel Core i5-8265U	2.288947e+06
9	HP	14-cf0007nx	Intel Core i5-8250U	2.288947e+06
163	HP	14-bp101nx	Intel Core i5-8250U	2.288947e+06
165	HP	14-ck0008nx	Intel Celeron N4000	2.288947e+06
130	Dell	Inspiron 14 5480	Intel Core i7-8565U	2.593494e+06