Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Факультет «Информатика и системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

Лабораторная работа №4
«Создание рекомендательной модели»
по дисциплине «Методы машинного обучения»

Выполнил:
студент группы ИУ5-25М
Тураев Г.В.
Подпись и дата:

Проверил:
преподаватель каф. ИУ5
Гапанюк Ю.Е.
Подпись и дата:

Москва 2023

Цель:

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).
4) Сформировать отчет и разместить его в своем репозитории на github.

Выполнение работы

Импортируем библиотеки:

```
[21] import numpy as np
     import pandas as pd
     from typing import Dict, Tuple
     from scipy import stats
     from IPython.display import Image
     from IPython.display import Image
     from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
     #from sklearn.datasets import load_iris, load_boston
     from sklearn.model_selection import cross_val_score
     from sklearn.model_selection import train_test_split
     from sklearn.neighbors import KNeighborsRegressor, KNeighborsClassifier
     from sklearn.model_selection import GridSearchCV, RandomizedSearchCV
     from sklearn.metrics import accuracy_score, balanced_accuracy_score
     from sklearn.metrics import precision_score, recall_score, f1_score, classification_report
     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.ensemble import GradientBoostingClassifier, GradientBoostingRegressor
     from sklearn.ensemble import BaggingClassifier
     from sklearn.ensemble import AdaBoostClassifier
     from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_squared_log_error, median_absolute_error, r2_score
     from sklearn.metrics import roc_curve, roc_auc_score
     from sklearn.metrics.pairwise import cosine_similarity, euclidean_distances, manhattan_distances
     from collections import defaultdict
     import seaborn as sns
     import matplotlib.pyplot as plt
     from matplotlib_venn import venn2
     %matplotlib inline
     sns.set(style="ticks")
```

Чтение и обработка данных:

```
[22] data = pd.read_csv('winemag.csv')
     data.head()
```

	Unnamed: 0	country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety	winery
0	0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	Kerin O'Keefe	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia
1	1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	NaN	Roger Voss	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portuguese Red	Quinta dos Avidagos
2	2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris	Rainstorm
3	3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	NaN	Alexander Peartree	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...	Riesling	St. Julian
4	4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	Pinot Noir	Sweet Cheeks

```
[23] data.shape
```

```
(129971, 14)
```

```
[26] description_data = data[data['description'].notnull()]
     description_data.shape
```

```
(129971, 14)
```

```
[30] title = description_data['title'].values
     title[0:5]
```

```
array(['Nicosia 2013 Vulkà Bianco  (Etna)',
       'Quinta dos Avidagos 2011 Avidagos Red (Douro)',
       'Rainstorm 2013 Pinot Gris (Willamette Valley)',
       'St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan Shore)',
       "Sweet Cheeks 2012 Vintner's Reserve Wild Child Block Pinot Noir (Willamette Valley)"],
      dtype=object)
```

```
[31] descriptions = description_data['description'].values
     descriptions[0:5]
```

```
array(["Aromas include tropical fruit, broom, brimstone and dried herb. The palate isn't overly expressive, offering unripened apple, citrus and dried sage alongside brisk acidity.",
       "This is ripe and fruity, a wine that is smooth while still structured. Firm tannins are filled out with juicy red berry fruits and freshened with acidity. It's  already drinkable, although it will certainly be better from 2016.",
       'Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all stainless-steel fermented.',
       'Pineapple rind, lemon pith and orange blossom start off the aromas. The palate is a bit more opulent, with notes of honey-drizzled guava and mango giving way to a slightly astringent, semidry finish.',
       "Much like the regular bottling from 2012, this comes across as rather rough and tannic, with rustic, earthy, herbal characteristics. Nonetheless, if you think of it as a pleasantly unfussy country wine, it's a good companion to a hearty winter stew."],
      dtype=object)
```

```
[32] description_data.keys()
```

```
Index(['Unnamed: 0', 'country', 'description', 'designation', 'points',
       'price', 'province', 'region_1', 'region_2', 'taster_name',
       'taster_twitter_handle', 'title', 'variety', 'winery'],
      dtype='object')
```

```
[34] wine_ids = description_data['Unnamed: 0'].values
     wine_ids
```

```
array([     0,      1,      2, ..., 129968, 129969, 129970])
```

```
[36] %%time
     tfidf = TfidfVectorizer()
     description_matrix = tfidf.fit_transform(descriptions)
     description_matrix
```

```
CPU times: user 7.51 s, sys: 184 ms, total: 7.7 s
Wall time: 11.9 s
<129971x31275 sparse matrix of type '<class 'numpy.float64'>'
        with 4475479 stored elements in Compressed Sparse Row format>
```

```
[37] description_matrix
```

```
<129971x31275 sparse matrix of type '<class 'numpy.float64'>'
        with 4475479 stored elements in Compressed Sparse Row format>
```

Фильтрация на основе содержания. Метод k-ближайших соседей.

```
[43] class SimplerKnnRecomender:
        def __init__(self, X_matrix, X_ids, X_title, X_overview):
            """
            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 = pd.DataFrame(
                {'id': pd.Series(X_ids, dtype='int'),
                'title': pd.Series(X_title, dtype='str'),
                'overview': pd.Series(X_overview, 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
```

```
[45] test_id = 11
     print(title[test_id])
     print(descriptions[test_id])

     Leon Beyer 2012 Gewurztraminer (Alsace)
     This is a dry wine, very spicy, with a tight, taut texture and strongly mineral character layered with citrus as well as pepper. It's a food wine with its almost crisp aftertaste.
```

```
[46] test_matrix = description_matrix[test_id]
     test_matrix

     <1x31275 sparse matrix of type '<class 'numpy.float64'>'
            with 25 stored elements in Compressed Sparse Row format>
```

```
[48] skr1 = SimplerKnnRecomender(description_matrix, wine_ids, title, descriptions)
```

```
[50] # 15 вин, наиболее похожих на Leon Beyer 2012 Gewurztraminer (Alsace), в порядке убывания схожести на основе косинусного сходства
     rec1 = skr1.recommend_for_single_object(15, test_matrix)
     rec1
```

	id	title	overview	dist
24045	24045	Domaine Michel Thomas et Fils 2015 Rosé (Sance...	The wine is textured and tight with crisp acid...	633624.990866
90700	90700	Henri de Villamont 2014 Morgeot Premier Cru (...	This wine is still tight and crisp. It has ple...	442624.176096
58330	58330	Schröder & Schÿler 2013 Chartron la Fleur (Bo...	The wine is tight and nervy, very fresh, crisp...	432556.705703
66081	66081	Maison Champy 2014 Viré-Clessé	This taut and structured wine has weight as we...	430242.028148
78572	78572	Domaine Olivier Merlin 2014 Mâcon La Roche Vi...	This wine is tight, structured and taut. Still...	428504.458538
105230	105230	Domaine Nigri 2013 Pierre de Lune (Jurançon Sec)	This rich and ripe wine is full of apricot and...	425886.605501
25907	25907	Louis Max 2014 Mâcon-Villages	Tight and structured, this wine has minerality...	424385.444731
99011	99011	Joseph Drouhin 2013 Les Clos (Macon-Bussières)	This crisp wine offers plenty of acidity as we...	423757.525560
5406	5406	Aveleda 2015 Alvarinho (Vinho Verde)	Ripe Alvarinho gives a wine that is rich as we...	421592.529700
22652	22652	Maison Malet Roquefort 2012 Léo de la Gaffelië...	Very herbaceous in character, this is a wine t...	418388.507228
129715	129715	Boeckel 2012 Vieilles Vignes Sylvaner (Alsace)	Intensely peppery as well as fruity, this is a...	416866.789965
119482	119482	Boeckel 2012 Vieilles Vignes Sylvaner (Alsace)	Intensely peppery as well as fruity, this is a...	416866.789965
21920	21920	Moncigale 2014 Frais et Délicat Rosé (Coteaux ...	This is crisp, fruity with apple and citrus fl...	411434.544994
96505	96505	Domaine Alban Roblin 2014 Rosé (Sancerre)	This is a fresh wine with caramel as well as r...	408987.116976
92292	92292	Domaine Alban Roblin 2014 Rosé (Sancerre)	This is a fresh wine with caramel as well as r...	408987.116976

```
[52] # При поиске с помощью Евклидова расстояния получаем такой же результат
     rec2 = skr1.recommend_for_single_object(15, test_matrix, cos_flag = False)
     rec2
```

	id	title	overview	dist
24045	24045	Domaine Michel Thomas et Fils 2015 Rosé (Sance...	The wine is textured and tight with crisp acid...	8.560082e+05
90700	90700	Henri de Villamont 2014 Morgeot Premier Cru (...	This wine is still tight and crisp. It has ple...	1.055818e+06
58330	58330	Schröder & Schÿler 2013 Chartron la Fleur (Bo...	The wine is tight and nervy, very fresh, crisp...	1.065311e+06
66081	66081	Maison Champy 2014 Viré-Clessé	This taut and structured wine has weight as we...	1.067481e+06
78572	78572	Domaine Olivier Merlin 2014 Mâcon La Roche Vi...	This wine is tight, structured and taut. Still...	1.069108e+06
105230	105230	Domaine Nigri 2013 Pierre de Lune (Jurançon Sec)	This rich and ripe wine is full of apricot and...	1.071553e+06
25907	25907	Louis Max 2014 Mâcon-Villages	Tight and structured, this wine has minerality...	1.072953e+06
99011	99011	Joseph Drouhin 2013 Les Clos (Macon-Bussières)	This crisp wine offers plenty of acidity as we...	1.073539e+06
5406	5406	Aveleda 2015 Alvarinho (Vinho Verde)	Ripe Alvarinho gives a wine that is rich as we...	1.075553e+06
22652	22652	Maison Malet Roquefort 2012 Léo de la Gaffelië...	Very herbaceous in character, this is a wine t...	1.078528e+06
119482	119482	Boeckel 2012 Vieilles Vignes Sylvaner (Alsace)	Intensely peppery as well as fruity, this is a...	1.079938e+06
129715	129715	Boeckel 2012 Vieilles Vignes Sylvaner (Alsace)	Intensely peppery as well as fruity, this is a...	1.079938e+06
21920	21920	Moncigale 2014 Frais et Délicat Rosé (Coteaux ...	This is crisp, fruity with apple and citrus fl...	1.084957e+06
92292	92292	Domaine Alban Roblin 2014 Rosé (Sancerre)	This is a fresh wine with caramel as well as r...	1.087210e+06
96505	96505	Domaine Alban Roblin 2014 Rosé (Sancerre)	This is a fresh wine with caramel as well as r...	1.087210e+06

```
[53] # Манхэттэнское расстояние дает несколько иные результаты поиска
     rec3 = skr1.recommend_for_single_object(15, test_matrix,
                                             cos_flag = False, manh_flag = True)
     rec3
```

	id	title	overview	dist
24045	24045	Domaine Michel Thomas et Fils 2015 Rosé (Sance...	The wine is textured and tight with crisp acid...	3.865262e+06
22652	22652	Maison Malet Roquefort 2012 Léo de la Gaffelière...	Very herbaceous in character, this is a wine t...	5.251729e+06
35502	35502	Château de Piote 2012 Perles (Crémant de Bord...	Tight and sharp, this is an herbaceous wine wi...	5.312967e+06
58330	58330	Schröder & Schÿler 2013 Chartron la Fleur (Bo...	The wine is tight and nervy, very fresh, crisp...	5.316624e+06
25907	25907	Louis Max 2014 Mâcon-Villages	Tight and structured, this wine has minerality...	5.354298e+06
21920	21920	Moncigale 2014 Frais et Délicat Rosé (Coteaux ...	This is crisp, fruity with apple and citrus fl...	5.452536e+06
97201	97201	Ravoire et Fils 2013 Domaine la Rabiotte Rosé ...	Tight, zingy and crisp, this wine has fresh, c...	5.535851e+06
70762	70762	Château du Seuil 2015 Domaine du Seuil (Borde...	The wine is tight and mineral in character. It...	5.564448e+06
128577	128577	Ravoire et Fils 2014 Domaine Bel Eouve Rosé (C...	This is a tangy, spicy wine, a character that ...	5.628584e+06
78572	78572	Domaine Olivier Merlin 2014 Mâcon La Roche Vi...	This wine is tight, structured and taut. Still...	5.644448e+06
92292	92292	Domaine Alban Roblin 2014 Rosé (Sancerre)	This is a fresh wine with caramel as well as r...	5.653916e+06
96505	96505	Domaine Alban Roblin 2014 Rosé (Sancerre)	This is a fresh wine with caramel as well as r...	5.653916e+06
108912	108912	Quinta do Portal 2012 Colheita Rosé (Douro)	This rosé is almost as rich as a red wine, the...	5.701024e+06
66081	66081	Maison Champy 2014 Viré-Clessé	This taut and structured wine has weight as we...	5.734040e+06
88898	88898	Markus Huber 2009 Hugo Grüner Veltliner (Niede...	Very crisp fruit, with light acidity and a tau...	5.751297e+06

Коллаборативная фильтрация. Метод на основе сингулярного разложения:

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

	Unnamed: 0	country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety	winery
0	0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	Kerin O'Keefe	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia
1	1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	NaN	Roger Voss	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portuguese Red	Quinta dos Avidagos
2	2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris	Rainstorm
3	3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	NaN	Alexander Peartree	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...	Riesling	St. Julian
4	4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	Pinot Noir	Sweet Cheeks

```
[ ] data3 = data[30000:55000]
```

```
[ ] # Количество уникальных дегустаторов
    len(data3['taster_name'].unique())
```

```
20
```

```
[ ] # Количество уникальных вин
    len(data3['title'].unique())
```

```
24517
```

```
[ ] # Сформируем матрицу взаимодействий на основе рейтингов
    # Используется идея из статьи - https://towardsdatascience.com/beginners-guide-to-creating-an-svd-recommender-system-1fd7326d1f65
    def create_utility_matrix(data):
        itemField = 'title'
        userField = 'taster_name'
        valueField = 'points'

        userList = data[userField].tolist()
        itemList = data[itemField].tolist()
        valueList = data[valueField].tolist()

        users = list(set(userList))
        items = list(set(itemList))

        users_index = {users[i]: i for i in range(len(users))}
        pd_dict = {item: [0.0 for i in range(len(users))] for item in items}

        for i in range(0,data.shape[0]):
            item = itemList[i]
            user = userList[i]
            value = valueList[i]
            pd_dict[item][users_index[user]] = value

        X = pd.DataFrame(pd_dict)
        X.index = users

        itemcols = list(X.columns)
        items_index = {itemcols[i]: i for i in range(len(itemcols))}

        return X, users_index, items_index
```

```
[ ] %%time
    user_item_matrix, users_index, items_index = create_utility_matrix(data3)
```

```
CPU times: user 622 ms, sys: 3.76 ms, total: 626 ms
Wall time: 635 ms
```

```
[ ] user_item_matrix
```

	William Knuttel 2006 Clone 777 Pinot Noir (Russian River Valley)	Joseph Fritsch 2013 Pinot Gris (Alsace)	Cayuse 2011 Cailloux Vineyard Viognier (Walla Walla Valley (WA))	Bodegas Murtia 2004 Reserva Monastrell (Jumilla)	Weszeli 2016 Langenlois Grüner Veltliner (Kamptal)	Robert Weil 2015 Tradition Riesling (Rheingau)	Venge 2014 Bacigalupi Vineyard Pinot Noir (Russian River Valley)	Gunsight Rock 2012 Cabernet Sauvignon (Paso Robles)	Barons V 2008 Cabernet Sauvignon (Columbia Valley (WA))	Château Crusquet Sabourin 2015 Blaye Côtes de Bordeaux	...	Gary Farrell 2012 Stiling Vineyard Pinot Noir (Russian River Valley)	Leonesse Cellars 2005 Cabernet Franc-Merlot Red (Temecula)	Château la France 2014 Bordeaux Blanc	Wy'East Vineyards 2014 Estate Grown Pinot Gris (Columbia Gorge (OR))	Fulkerson 2016 Reserve Riesling (Finger Lakes)	FiàNobile 2013 Nero d'Avola (Terre Siciliane)
Fiona Adams	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Lauren Buzzeo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Matt Kettmann	0.0	0.0	0.0	0.0	0.0	0.0	0.0	85.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Roger Voss	0.0	85.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	88.0	...	0.0	0.0	87.0	0.0	0.0	0.0
Anna Lee C. Iijima	0.0	0.0	0.0	0.0	0.0	91.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	86.0	0.0
Mike DeSimone	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
...																	
Alexander Peartree	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Susan Kostrzewa	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Paul Gregutt	0.0	0.0	94.0	0.0	0.0	0.0	0.0	0.0	87.0	0.0	...	0.0	0.0	0.0	91.0	0.0	0.0
Kerin O'Keefe	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	85.0
Joe Czerwinski	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
Anne Krebiehl MW	0.0	0.0	0.0	0.0	89.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0

20 rows × 24517 columns

```
[ ] # Выделение тестовой строки
    user_item_matrix__test = user_item_matrix.loc[['Kerin O'Keefe']]
    user_item_matrix__test
```

	William Knuttel 2006 Clone 777 Pinot Noir (Russian River Valley)	Joseph Fritsch 2013 Pinot Gris (Alsace)	Cayuse 2011 Cailloux Vineyard Viognier (Walla Walla Valley (WA))	Bodegas Murtia 2004 Reserva Monastrell (Jumilla)	Weszeli 2016 Langenlois Grüner Veltliner (Kamptal)	Robert Weil 2015 Tradition Riesling (Rheingau)	Venge 2014 Bacigalupi Vineyard Pinot Noir (Russian River Valley)	Gunsight Rock 2012 Cabernet Sauvignon (Paso Robles)	Barons V 2008 Cabernet Sauvignon (Columbia Valley (WA))	Château Crusquet Sabourin 2015 Blaye Côtes de Bordeaux	...	Gary Farrell 2012 Stiling Vineyard Pinot Noir (Russian River Valley)	Leonesse Cellars 2005 Cabernet Franc-Merlot Red (Temecula)	Château la France 2014 Bordeaux Blanc	Wy'East Vineyards 2014 Estate Grown Pinot Gris (Columbia Gorge (OR))	Fulkerson 2016 Reserve Riesling (Finger Lakes)	FiàNobile 2013 Nero d'Avola (Terre Siciliane)	Hus Estate Gewurzt (A
Kerin O'Keefe	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	85.0	

1 rows × 24517 columns

```
[ ] #taster_names = description_data['taster_name'].unique()
    taster_names = np.delete(data3['taster_name'].unique(), 0)
    taster_names = np.delete(taster_names, 7)
    taster_names
```

```
array(['Jim Gordon', 'Michael Schachner', 'Matt Kettmann',
       'Sean P. Sullivan', 'Roger Voss', 'Virginie Boone',
       'Joe Czerwinski', 'Paul Gregutt', 'Mike DeSimone', 'Jeff Jenssen',
       nan, 'Anna Lee C. Iijima', 'Susan Kostrzewa', 'Lauren Buzzeo',
       'Alexander Peartree', 'Fiona Adams', 'Carrie Dykes',
       'Christina Pickard'], dtype=object)
```

```
[ ] # Оставшаяся часть матрицы для обучения
    user_item_matrix__train = user_item_matrix.loc[taster_names]
    user_item_matrix__train
```

	William Knuttel 2006 Clone 777 Pinot Noir (Russian River Valley)	Joseph Fritsch 2013 Pinot Gris (Alsace)	Cayuse 2011 Cailloux Vineyard Viognier (Walla Walla Valley (WA))	Bodegas Murtia 2004 Reserva Monastrell (Jumilla)	Weszeli 2016 Langenlois Grüner Veltliner (Kamptal)	Robert Weil 2015 Tradition Riesling (Rheingau)	Venge 2014 Bacigalupi Vineyard Pinot Noir (Russian River Valley)	Gunsight Rock 2012 Cabernet Sauvignon (Paso Robles)	Barons V 2008 Cabernet Sauvignon (Columbia Valley (WA))	Château Crusquet Sabourin 2015 Blaye Côtes de Bordeaux	...	Gary Farrell 2012 Stiling Vineyard Pinot Noir (Russian River Valley)	Leonesse Cellars 2005 Cabernet Franc-Merlot Red (Temecula)	Château la France 2014 Bordeaux Blanc	Wy'East Vineyards 2014 Estate Grown Pinot Gris (Columbia Gorge (OR))	Fulkerson 2016 Reserve Riesling (Finger Lakes)	FiàNobile 2013 Nero d'Avola (Terre Siciliane)	E G
Jim Gordon	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Michael Schachner	0.0	0.0	0.0	90.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Matt Kettmann	0.0	0.0	0.0	0.0	0.0	0.0	0.0	85.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Sean P. Sullivan	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Roger Voss	0.0	85.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	88.0	...	0.0	0.0	87.0	0.0	0.0	0.0	
Virginie Boone	0.0	0.0	0.0	0.0	0.0	0.0	91.0	0.0	0.0	0.0	...	91.0	0.0	0.0	0.0	0.0	0.0	
Joe																		

...

Susan Kostrzewa	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Lauren Buzzeo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Alexander Peartree	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Fiona Adams	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Carrie Dykes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	
Christina Pickard	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	

18 rows × 24517 columns

```
[ ] %%time
    U, S, VT = np.linalg.svd(user_item_matrix__train.T)
    V = VT.T
```

```
[ ] # Матрица соотношения между дегустаторами и латентными факторами
    U.shape
```

```
[ ] # Матрица соотношения между объектами и латентными факторами
    V.shape
```

```
(18, 18)
```

```
[ ] S.shape
```

```
(18,)
```

```
[ ] Sigma = np.diag(S)
    Sigma.shape
```

```
(18, 18)
```

```
# Диагональная матрица сингулярных значений
Sigma
```

```
array([[6328.37615756,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        , 6214.00788753,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        , 4603.41568838,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        , 3880.90866797,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
        3683.95055254,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        , 3004.09187609,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
                                   ...
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,  390.0179483 ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
         177.03107072,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,  124.45882853]])
```

```
[ ]  # Используем 3 первых сингулярных значения
     r=3
     Ur = U[:, :r]
     Sr = Sigma[:r, :r]
     Vr = V[:, :r]
     # Матрица соотношения между новым дегустатором и латентными факторами
     test_user = np.mat(user_item_matrix__test.values)
     test_user.shape, test_user
```

```
((1, 24517), matrix([[0., 0., 0., ..., 0., 0., 0.]]))
```

```
[ ]  tmp = test_user * Ur * np.linalg.inv(Sr)
     tmp
```

```
matrix([[ 3.78394162e-04, -4.35827216e-06,  2.92221682e-18]])
```

```
[ ]  test_user_result = np.array([tmp[0,0], tmp[0,1], tmp[0,2]])
     test_user_result
```

```
array([ 3.78394162e-04, -4.35827216e-06,  2.92221682e-18])
```

```
[ ]  # Вычисляем косинусную близость между текущим дегустатором
     # и остальными дегустаторами
     cos_sim = cosine_similarity(Vr, test_user_result.reshape(1, -1))
     cos_sim[:10]
```

```
array([[ 9.99999728e-01],
       [-1.53344496e-18],
       [ 5.56050464e-34],
       [ 3.47895878e-33],
       [-4.12491330e-04],
       [ 9.99999975e-01],
       [ 3.44380739e-37],
       [-1.04994959e-03],
       [ 0.00000000e+00],
       [ 0.00000000e+00]])
```

```
[ ]  # Преобразуем размерность массива
     cos_sim_list = cos_sim.reshape(-1, cos_sim.shape[0])[0]
     cos_sim_list[:10]
```

```
array([ 9.99999728e-01, -1.53344496e-18,  5.56050464e-34,  3.47895878e-33,
       -4.12491330e-04,  9.99999975e-01,  3.44380739e-37, -1.04994959e-03,
        0.00000000e+00,  0.00000000e+00])
```

```
[ ]  # Находим наиболее близкого дегустатора
     recommended_user_id = np.argsort(-cos_sim_list)[0]
     recommended_user_id
```

```
5
```

```
[ ]  test_user
```

```
matrix([[0., 0., 0., ..., 0., 0., 0.]])
```

```
[ ]  # Получение названия вина
     wine_list = list(user_item_matrix.columns)
     def film_name_by_movieid(ind):
         try:
             wine = wine_list[ind]
             #print(wineId)
             #flt_links = data3[data['movieId'] == wineId]
             #tmdbId = int(flt_links['tmdbId'].values[0])
             #md_links = df_md[df_md['id'] == tmdbId]
             #res = md_links['title'].values[0]
             return wine
         except:
             return ''
```

```
[ ]  # Вина, которые оценивал текущий дегустатор:
     i=1
     for idx, item in enumerate(np.ndarray.flatten(np.array(test_user))):
         if item > 0:
             film_title = film_name_by_movieid(idx)
             print('{} - {} - {}'.format(idx, film_title, item))
             if i==20:
                 break
             else:
                 i+=1
```

```
20 - Colterenzio 2014 Pinot Grigio (Alto Adige) - 86.0
31 - Castello di Monsanto 2011 Il Poggio Riserva  (Chianti Classico) - 91.0
37 - Paltrinieri 2012 Grosso Metodo Classico  (Lambrusco di Modena) - 92.0
43 - Marchesi Antinori 2013 Pèppoli  (Chianti Classico) - 89.0
62 - San Lorenzo Vini 2014  Trebbiano d'Abruzzo - 84.0
71 - Feudi di San Gregorio 2014 Studi Fraedane  (Fiano di Avellino) - 89.0
75 - Ca' del Baio 2012 Marcarini  (Barbaresco) - 92.0
80 - Monteraponi 2014  Chianti Classico - 88.0
100 - Marchesi de' Frescobaldi 2011 Mormoreto Red (Toscana) - 92.0
115 - Fazio 2015 Calebianche Catarratto (Erice) - 86.0
124 - Caccia al Piano 1868 2011 Levia Gravia  (Bolgheri Superiore) - 90.0
145 - Feudi di San Gregorio 2012 Serpico Aglianico (Irpinia) - 93.0
146 - Principe di Corleone 2015 Bianca di Corte White (Sicilia) - 86.0
155 - Adriano Marco & Vittorio 2013 Sanadaive  (Barbaresco) - 94.0
171 - Pratello 2014 Catulliano  (Lugana) - 89.0
172 - Jermann 2015 Pinot Grigio (Friuli Venezia Giulia) - 90.0
187 - Caldaro 2013 Solos Schiava (Alto Adige) - 88.0
192 - Mario Gagliasso 2011 Rocche dell'Annunziata  (Barolo) - 88.0
205 - Lechthaler 2014 Torre di Luna Pinot Grigio (Trentino) - 83.0
214 - Feudi di San Gregorio 2013 Studi Campo Aperto  (Fiano di Avellino) - 93.0
```

```
[ ]  # Вина, которые оценивал наиболее схожий дегустатор:
     i=1
     recommended_user_item_matrix = user_item_matrix.loc[['Roger Voss']]
     for idx, item in enumerate(np.ndarray.flatten(np.array(recommended_user_item_matrix))):
         if item > 0:
             film_title = film_name_by_movieid(idx)
             print('{} - {} - {}'.format(idx, film_title, item))
             if i==20:
                 break
             else:
                 i+=1
```

```
1 - Joseph Fritsch 2013 Pinot Gris (Alsace) - 85.0
9 - Château Crusquet Sabourin 2015  Blaye Côtes de Bordeaux - 88.0
23 - Domaine de Cause 2013 Tradition Malbec (Cahors) - 86.0
26 - Domaine Rolet Père et Fils 2005 Vin Jaune Savagnin (Arbois) - 94.0
33 - Finisterra 2009 Red (Alentejano) - 84.0
38 - Château Lanbersac 2010 Cuvée Vieilles Vignes  (Puisseguin Saint-Émilion) - 90.0
40 - Domaine de Ménard 2016 Colombard-Ugni Blanc (Côtes de Gascogne) - 85.0
54 - Companhia das Quintas 2009 Quinta da Fronteira Grande Escolha Red (Douro) - 93.0
60 - Clos Floridène 2015 Le Rosé de Floridène Rosé (Bordeaux Rosé) - 85.0
68 - Pascal Bouchard 2011 Blanchot Grand Cru  (Chablis) - 94.0
81 - Campelo 2014 Tres Castas White (Vinho Verde) - 85.0
89 - Château Laulerie 2011 White (Bergerac Sec) - 85.0
90 - José Maria da Fonseca 2013 Twin Vines White (Vinho Verde) - 84.0
94 - Quinta de la Rosa 2010 Dourosa Red (Douro) - 88.0
96 - Domaines Bunan 2015 Château la Rouvière Rosé (Bandol) - 92.0
102 - Château Fourcas Dupré 2007  Listrac-Médoc - 88.0
103 - Domäne Wachau 2007 Terrassen Federspiel Riesling (Wachau) - 87.0
111 - Joseph Fritsch 2013 Pinot Blanc (Alsace) - 85.0
112 - DFJ Vinhos 2011 Scancio Private Selection Vinhas Velhas Red (Lisboa) - 91.0
116 - Bernard Magrez 2015 L'Excellence des Muraires Rosé (Côtes de Provence) - 92.0
```

Вывод: как видно, фильтрация на основе содержания и коллаборативная фильтрация показывают различные результаты работы в рамках рекомендательных систем.