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

Лабораторная работа №4
по курсу «Методы машинного обучения»

«Создание рекомендательной модели»

ИСПОЛНИТЕЛЬ:

Андронов Д.О.

Группа ИУ5-24М

____

"__"___________2022 г.

Москва 2022

In [2]:

```
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")
```

In [3]:

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Чтение и обработка данных

In [4]:

```
data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/winemag-data-130k-v2.csv')
data.head()
```

Out[4]:

	Unnamed: 0	country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety
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
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
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
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
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

In [5]:

```
data.shape
```

Out[5]:

```
(129971, 14)
```

In [6]:

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

Out[6]:

```
(129971, 14)
```

In [7]:

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

Out[7]:

```
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)
```

In [8]:

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

Out[8]:

```
array(["Aromas include tropical fruit, broom, brimstone and dried herb. The palate isn't overly expressive, offering unripened apple, citrus and dried sage alon
gside 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  alre
ady 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 givi
ng 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 thin
k of it as a pleasantly unfussy country wine, it's a good companion to a hearty winter stew."],
      dtype=object)
```

In [9]:

```
description_data.keys()
```

Out[9]:

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

In [10]:

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

Out[10]:

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

In [11]:

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

```
CPU times: user 3.55 s, sys: 42.3 ms, total: 3.59 s
Wall time: 3.61 s
```

In [12]:

```
description_matrix
```

Out[12]:

```
<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

In [13]:

```
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
```

In [14]:

```
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 cris
p aftertaste.

In [15]:

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

Out[15]:

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

In [16]:

```
skr1 = SimplerKnnRecomender(description_matrix, wine_ids, title, descriptions)
```

In [17]:

```
# 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
```

Out[17]:

	id	title	overview	dist
102760	102760	Leon Beyer 2012 Gewurztraminer (Alsace)	This is a dry wine, very spicy, with a tight, ...	1000000.000000
11	11	Leon Beyer 2012 Gewurztraminer (Alsace)	This is a dry wine, very spicy, with a tight, ...	1000000.000000
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 Gaffeliè...	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

In [18]:

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

Out[18]:

	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 Gaffeliè...	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

In [19]:

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

Out[19]:

	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 Gaffeliè...	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

In [20]:

```
data.head()
```

Out[20]:

	Unnamed: 0	country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety
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
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
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
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
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

In [21]:

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

In [22]:

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

Out[22]:

20

In [23]:

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

Out[23]:

24517

In [24]:

```
# 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
```

In [25]:

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

CPU times: user 762 ms, sys: 11.1 ms, total: 773 ms
Wall time: 777 ms

In [26]:

```
user_item_matrix
```

Out[26]:

	Dão Sul 2006 Berço do Infante Red (Estremadura)	Lemelson 2009 Dry Riesling (Willamette Valley)	Jasper Hill 2014 Georgia's Paddock Shiraz (Heathcote)	Bowers Harbor 2013 Langley Late Harvest Riesling (Old Mission Peninsula)	Stadt Krems 2012 Steinterrassen Riesling (Kremstal)	Youngberg Hill Vineyards 2012 Pinot Blanc (McMinnville)	Jaffurs 2013 Grenache (Santa Barbara County)	Pratsch 2012 Steinberg Grüner Veltliner (Niederösterreich)	Hunnicutt 2006 Zinfandel (Napa Valley)	Château Guilhem 2015 Pot de Vin Syrah Rosé (Pays d'Oc)	...	Dão Su 200 Quinta d Encontr Pret Branc Bag (Bairrada
NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	92.0	0.0	...	0.
Anna Lee C. Iijima	0.0	0.0	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.
Joe Czerwinski	0.0	0.0	91.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.
Alexander Peartree	0.0	0.0	0.0	88.0	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	83.0	...	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.
Michael Schachner	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.
Fiona Adams	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	0.0	0.0	0.0	92.0	0.0	0.0	...	0.
Matt Kettmann	0.0	0.0	0.0	0.0	0.0	0.0	92.0	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.
Jim Gordon	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.
Roger Voss	87.0	0.0	0.0	0.0	89.0	0.0	0.0	0.0	0.0	0.0	...	88.
Mike DeSimone	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.
Virginie Boone	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	91.0	0.0	0.0	0.0	86.0	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.
Jeff Jenssen	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

In [27]:

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

Out[27]:

	Dão Sul 2006 Berço do Infante Red (Estremadura)	Lemelson 2009 Dry Riesling (Willamette Valley)	Jasper Hill 2014 Georgia's Paddock Shiraz (Heathcote)	Bowers Harbor 2013 Langley Late Harvest Riesling (Old Mission Peninsula)	Stadt Krems 2012 Steinterrassen Riesling (Kremstal)	Youngberg Hill Vineyards 2012 Pinot Blanc (McMinnville)	Jaffurs 2013 Grenache (Santa Barbara County)	Pratsch 2012 Steinberg Grüner Veltliner (Niederösterreich)	Hunnicutt 2006 Zinfandel (Napa Valley)	Château Guilhem 2015 Pot de Vin Syrah Rosé (Pays d'Oc)	...	Dão Sul 2004 Quinta do Encontro Preto Branco Baga (Bairrada)
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

1 rows × 24517 columns

In [28]:

```
#taster_names = description_data['taster_name'].unique()
```

```
#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
```

Out[28]:

```
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)
```

In [29]:

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

Out[29]:

	Dão Sul 2006 Berço do Infante Red (Estremadura)	Lemelson 2009 Dry Riesling (Willamette Valley)	Jasper Hill 2014 Georgia's Paddock Shiraz (Heathcote)	Bowers Harbor 2013 Langley Late Harvest Riesling (Old Mission Peninsula)	Stadt Krems 2012 Steinterrassen Riesling (Kremstal)	Youngberg Hill Vineyards 2012 Pinot Blanc (McMinnville)	Jaffurs 2013 Grenache (Santa Barbara County)	Pratsch 2012 Steinberg Grüner Veltliner (Niederösterreich)	Hunnicutt 2006 Zinfandel (Napa Valley)	Château Guilhem 2015 Pot de Vin Syrah Rosé (Pays d'Oc)	...	Dão Sul 2004 Quinta do Encontro Preto Branco Baga (Bairrada)
Jim Gordon	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	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	92.0	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
Roger Voss	87.0	0.0	0.0	0.0	89.0	0.0	0.0	0.0	0.0	0.0	...	88.0
Virginie Boone	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
Joe Czerwinski	0.0	0.0	91.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
Paul Gregutt	0.0	91.0	0.0	0.0	0.0	86.0	0.0	0.0	0.0	0.0	...	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
Jeff Jenssen	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
NaN	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	92.0	0.0	...	0.0
Anna Lee C. Iijima	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
Lauren Buzzeo	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	83.0	...	0.0
Alexander Peartree	0.0	0.0	0.0	88.0	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
Carrie Dykes	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

18 rows × 24517 columns

In [30]:

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

```
CPU times: user 33.7 s, sys: 3.38 s, total: 37.1 s
Wall time: 23.2 s
```

In [31]:

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

Out[31]:

```
(24517, 24517)
```

In [32]:

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

Out[32]:

```
(18, 18)
```

In [33]:

```
S.shape
```

Out[33]:

```
(18,)
```

In [34]:

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

Out[34]:

```
(18, 18)
```

In [35]:

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

Out[35]:

```
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.        , 2830.86117639,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        , 2714.05324192,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
        2451.15054617,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        , 2416.46258391,    0.        ,    0.        ,
           0.        ,    0.        ,    0.        ,    0.        ,
           0.        ,    0.        ],
       [   0.        ,    0.        ,    0.        ,    0.        ,
```

```
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        , 1559.43034471,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ],
     [   0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        , 1283.41926119,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ],
     [   0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
     922.26406197,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ],
     [   0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,  825.2545062 ,    0.        ,    0.        ,
       0.        ,    0.        ],
     [   0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,    0.        ,    0.        ,
       0.        ,    0.        ,  745.06040024,    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]])
```

In [36]:

```
# 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
```

Out[36]:

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

In [37]:

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

Out[37]:

```
matrix([[ 3.78394162e-04, -4.35827216e-06, -2.92218350e-18]])
```

In [38]:

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

Out[38]:

```
array([ 3.78394162e-04, -4.35827216e-06, -2.92218350e-18])
```

In [39]:

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

Out[39]:

```
array([[ 9.99999728e-01],
       [-1.44541469e-18],
       [ 3.53594407e-33],
       [ 3.06381034e-35],
       [-4.12491330e-04],
       [ 9.99999975e-01],
       [ 0.00000000e+00],
       [-1.04994959e-03],
       [ 0.00000000e+00],
       [ 0.00000000e+00]])
```

In [40]:

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

Out[40]:

```
array([ 9.99999728e-01, -1.44541469e-18,  3.53594407e-33,  3.06381034e-35,
       -4.12491330e-04,  9.99999975e-01,  0.00000000e+00, -1.04994959e-03,
        0.00000000e+00,  0.00000000e+00])
```

In [41]:

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

Out[41]:

```
5
```

In [42]:

```
test_user
```

Out[42]:

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

In [43]:

```
# 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 ''
```

In [44]:

```
# 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
```

10 - Molino di Sant'Antimo 2010 Brunello di Montalcino - 93.0
17 - Borgogno 2015 Dolcetto d'Alba - 89.0
18 - Feudi di San Gregorio 2013 Studi Campo Aperto (Fiano di Avellino) - 93.0
20 - Ca'Romè 2012 Chiaramanti (Barbaresco) - 88.0
32 - Nicolucci 2015 Tre Rocche Sangiovese (Romagna) - 88.0
39 - Dorigo 2013 Ribolla Gialla (Colli Orientali del Friuli) - 91.0
79 - Colutta 2013 Pinot Grigio (Colli Orientali del Friuli) - 88.0
87 - Poggio Scalette 2014 Chianti Classico - 88.0
106 - Feudo Principi di Butera 2015 Nero d'Avola (Sicilia) - 88.0
116 - Canella 2014 Extra Dry (Valdobbiadene Prosecco Superiore) - 89.0
124 - Conte Collalto NV Brut (Valdobbiadene Prosecco Superiore) - 89.0
126 - Rascioni e Cecconello 2015 Maremmino (Maremma) - 88.0
137 - Castelfeder 2012 Glener Pinot Nero (Alto Adige) - 88.0
146 - La Vis 2012 L'Altro Manzoni Incrocio Manzoni (Vigneti delle Dolomiti) - 87.0
164 - Cascina Luisin 2012 Paolin (Barbaresco) - 88.0
171 - Germano Ettore 2012 del Comune di Serralunga d'Alba (Barolo) - 91.0
203 - Stemmari 2012 Nero d'Avola (Terre Siciliane) - 86.0
224 - Castelli del Grevepesa 2009 Riserva Castello di Bibbione (Chianti Classico) - 90.0
226 - Pieropan 2009 Le Colombare (Recioto di Soave) - 90.0
259 - Villa Calcinaia 2012 Chianti Classico - 89.0

In [45]:

```
# 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
```

0 - Dão Sul 2006 Berço do Infante Red (Estremadura) - 87.0
4 - Stadt Krems 2012 Steinterrassen Riesling (Kremstal) - 89.0
13 - Deco Provence - Villa Azur 2015 Rosé (Coteaux Varois en Provence) - 85.0
25 - Vignerons de Bel Air 2010 Hiver Gourmand (Morgon) - 86.0
29 - Cave du Marmandais 2011 Château Terrebert Malbec (Côtes du Marmandais) - 87.0
41 - Quinta Nova de Nossa Senhora do Carmo 2008 Referencia Grand Reserva Red (Douro) - 91.0
44 - Herdade do Perdigão 2009 Terras de Monforte Red (Alentejo) - 88.0
48 - Château L'Argilus du Roi 2011 Saint-Estèphe - 83.0
49 - Manuel Olivier 2010 Bourgogne - 85.0
53 - Château Paradis 2010 Red (Coteaux d'Aix-en-Provence) - 91.0
54 - Bernard Magrez 2011 Château du Galan (Haut-Médoc) - 90.0
55 - Quinta do Tedo 2009 Savedra Vintage (Port) - 90.0
63 - Domaine du Coudray 2015 Une Pointe d'Authenticité (Quincy) - 90.0
66 - Domaine des Cognettes 2005 Tentation Sélection Vieilles Vignes (Muscadet Sèvre et Maine) - 91.0
69 - Les Héritiers du Comte Lafon 2013 Viré-Clessé - 90.0
71 - Herdade dos Machados 2013 Santos Jorge Red (Alentejo) - 87.0
88 - Château Lafite Rothschild 2014 Carruades de Lafite (Pauillac) - 94.0
92 - Domaine Méo-Camuzet 2013 Gevrey-Chambertin - 90.0
100 - Château de Fuissé 2009 Tête de Cru (Pouilly-Fuissé) - 91.0
105 - Salomon-Undhof 2012 Hochterrassen Grüner Veltliner (Niederösterreich) - 85.0

Как видно, фильтрация на основе содержания и коллаборативная фильтрация показывают различные результаты работы в рамках рекомендательных систем