



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, искусственный интеллект и системы управления _____

КАФЕДРА _____ Системы обработки информации и управления _____

ОТЧЁТ ***К ЛАБОРАТОРНОЙ РАБОТЕ №4***

НА ТЕМУ:

Создание рекомендательной системы

Студент: Громоздов Д.Р.

Группа: ИУ5-23М

Преподаватель: Гапанюк Ю.Е.

Цель лабораторной работы: изучение разработки рекомендательных моделей.

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).

```
In [24]: import pandas as pd
        from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
        from sklearn.metrics.pairwise import cosine_similarity, euclidean_distances, manhattan_distances
```

Для создания рекомендательной модели используем набор данных для рекомендательной системы парфюма, содержащий информацию о названии, изготовителе, составе запаха, url-ссылку на изображение упаковки, а также текстовое описание, на основе которого, мы и будем составлять соответствующие рекомендации:

Загрузим датасет:

```
In [6]: df_perf_all = pd.read_csv('data/perfume.csv', sep=";", encoding = 'ansi')
        df_perf_all.head()
```

Out[6]:	Name	Brand	Description	Notes	Image URL
0	Tihota Eau de Parfum	Indult	Rapa Nui for sugar, Tihota is, quite simply, ...	Vanilla bean, musks	https://static.luckyscent.com/images/products/...
1	Sola Parfum	Di Ser	A tribute to the expanse of space extending f...	Lavender, Yuzu, Lemongrass, Magnolia, Geraniu...	https://static.luckyscent.com/images/products/...
2	Kagiroi Parfum	Di Ser	An aromatic ode to the ancient beauty of Japa...	Green yuzu, green shikuwasa, sansho seed, cor...	https://static.luckyscent.com/images/products/...
3	Velvet Fantasy Eau de Parfum	Montale	Velvet Fantasy is a solar fragrance where cit...	tangerine, pink pepper, black coffee, leat...	https://static.luckyscent.com/images/products/...
4	A Blvd. Called Sunset Eau de Parfum	A Lab on Fire	There's no way A Lab On Fire could relocate t...	Bergamot, almond, violet, jasmine, leather, s...	https://static.luckyscent.com/images/products/...

```
In [7]: df_perf_all.shape
```

Out[7]:(2191, 5)
Убедимся, что в нашем рабочем датафрейме не будет записей с отсутствующим текстовым описанием:

```
In [8]: df_perf_with_description = df_perf_all[df_perf_all['Description'].notnull()]
        df_perf_with_description = df_perf_with_description[~df_perf_with_description['Description'].str.isspace()]
```

```
In [9]: name = df_perf_with_description['Name'].values
        name[0:6]
```

```
Out[9]:array(['Tihota Eau de Parfum', 'Sola Parfum', 'Kagiroi Parfum',
              'Velvet Fantasy Eau de Parfum',
              'A Blvd. Called Sunset Eau de Parfum',
              'Freckled and Beautiful Eau de Parfum'], dtype=object)
```

```
In [11]: notes = df_perf_with_description['Notes'].values
        notes[0:5]
```

```
Out[11]:array(['Vanilla bean, musks',
              'Lavender, Yuzu, Lemongrass, Magnolia, Geranium, Jasmine, Frankincense, Myrrh',
              'Green yuzu, green shikuwasa, sansho seed, coriander, ylang-ylang, shiso, rosewood, vetiver, hinoki, cypriol, patchouli, agarwood',
              'tangerine, pink pepper, black coffee, leather, violet, jasmine, lily of the valley, heliotrope powder, vanilla, amber, sandalwood, toffee, musk, oakmoss',
              'Bergamot, almond, violet, jasmine, leather, sandalwood, vanilla, tonka'],
              dtype=object)
```

```
In [13]: descr = df_perf_with_description['Description'].values
        descr[0:3]
```

Out[13]:array(["Rapa Nui for sugar, Tihota is, quite simply, The One. The One that will call to you every moment you're not smelling it, The One that you've only had hints of in other vanilla perfumes, The One that lasts and lasts, The One that has perfectly captured the essence of the pure nature of fresh vanilla beans and has harnessed it beyond your wildest dreams. It inspires obsessive devotion with its fragrance of smooth, vanilla bean pods dipped in honeyed water and left to steep. The result is the pure magic, an unbridled vanilla, sweet, raw and achingly desirable. Of course, Tihota isn't the first perfume to focus on the dark sweetness of vanilla... perfumery is filled with vanillas, but this smells like the dream of a master perfumer who was obsessed with vanilla and was finally given free rein to worship the note with no apologies. It is, without a sliver of a doubt, the finest pure vanilla we've ever had the honor of carrying, and it's so beautiful it pains us to put the bottle down.",
 'A tribute to the expanse of space extending from the sky to the flower and fruit laden earth below, Sola, the newest offering from DI SER and perfumer Yasuyuki Shinohara, showcases the natural flora of Hokkaido, Japan with locally sourced and sustainably harvested ingredients from DI SER's own gardens. Starting with aromatic notes of freshly harvested Hokkaido lavender, lemongrass and yuzu, the uplifting scent moves to a floral middle of Japanese rose, magnolia, geranium and jasmine before settling on a spiritual bed of frankincense and myrrh. The spiritual and meditative scent evokes an imagery of a peaceful universe in complete harmony with the bounties of nature and of peace and contentment with its generous offerings.',
 'An aromatic ode to the ancient beauty of Japan, Kagiroi represents the changes in the shade of the dawn sky over the mountains of Hokkaido, Japan. As the dawn sky transitions through shades of purple, yellow, orange and red, this scent that showcases many ingredients used for the first time in perfumery transitions from an aromatic slightly spicy top to an earthy and woody base, representing the ebb and flow in the stages of life. The perfume opens with notes of Japanese herbs and citruses, moves to a slightly floral spicy middle and settles into a deep woody base accentuated by the use of the Hinoki cedar that dot the mountain landscapes. A deeply personal and spiritual scent, Kagiroi is a scent of shifting aromatic complexities representing transitions in the journey of life.Parfum Edition: We are honored and delighted to offer the rare and precious Parfum edition of Kagiroi, which adds to the composition a precious specimen of Japanese agarwood, which adds rich, spiritual dimensionality and a truly impressive and long-lasting drydown to this already complex and wonderful scent.'],
 dtype=object)

Векторизуем описания с помощью Tf-Idf Vectorizer

```
In [16]: tfidf = TfidfVectorizer()
        descr_matrix = tfidf.fit_transform(descr)
        descr_matrix
```

```
Out[16]:<2191x18753 sparse matrix of type '<class 'numpy.float64'>'
        with 216102 stored elements in Compressed Sparse Row format>
```

И с помощью CountVectorizer:

```
In [42]: countv = CountVectorizer()
descr_matrix_co = countv.fit_transform(descr)
descr_matrix_co

Out[42]: <2191x18753 sparse matrix of type '<class 'numpy.int64'>'
with 216102 stored elements in Compressed Sparse Row format>
```

```
In [30]: class SimpleKNNRecommender:
```

```
def __init__(self, X_matrix, X_names, X_notes, X_descr):
    """
    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(
        {'Perfume Name': pd.Series(X_names, dtype='str'),
         'Notes': pd.Series(X_notes, dtype='str'),
         'Description': pd.Series(X_descr, 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
```

Выберем тестовый образец, на основе которого мы будем давать рекомендации:

```
In [35]: test_perfume_name = 1000
name[test_perfume_name]

Out[35]: 'California Snow Eau de Parfum'
Зададим его матрицу:

In [37]: test_perfume_matrix = descr_matrix[test_perfume_name]
test_perfume_matrix

Out[37]: <1x18753 sparse matrix of type '<class 'numpy.float64'>'
with 146 stored elements in Compressed Sparse Row format>

In [39]: skr1 = SimpleKNNRecommender(descr_matrix, name, notes, descr)

In [40]: test = df_perf_with_description.iloc[test_perfume_name]
test
```

Out[40]:Name California Snow Eau de Parfum
Brand A Lab on Fire
Description California Snow crackles with the electricity...
Notes Sage, tea, chamomile, coumarin, narcissus, ro...
Image URL https://static.luckyscent.com/images/products/...
Name: 1000, dtype: object

Делаем рекомендацию на основании описания векторизованного Tf-Idf и косинусного расстояния:

In [41]: rec1 = skr1.recommend_for_single_object(15, test_perfume_matrix)
rec1

Out[41]:	Perfume Name	Notes	Description	Dist
1000	California Snow Eau de Parfum	Sage, tea, chamomile, coumarin, narcissus, ro...	California Snow crackles with the electricity...	1000000.000000
1900	1 Parfum Extrait	Fresh tangerine blossom, cassis, neroli bigar...	One of the world's most exclusive fragrances,...	201793.045093
1895	Arso Eau de Parfum	Leather, incense, pine resin, cedar leaves	The sharp, evocative scent of wood smoke - t...	201572.053201
1921	De Bachmakov Eau de Parfum	Cedar wood, bergamot, shiso leaves, nutmeg, c...	In a tribute to his Russian ancestry and to c...	183499.803113
1220	Clementine California Cologne Absolue	Clementine, mandarin, juniper berries, star a...	In terms of sheer, natural joy, how many sens...	181620.621872
845	Woody Mood Eau de Parfum	bergamot, ginger, clary sage, saffron, sequoi...	With Woody Mood, perfumer Betrand Duchaufour ...	174702.171909
1496	Gypsy Water Hair Perfume	Bergamot, lemon, pepper, juniper berries, inc...	BYREDO presents a hair perfume collection tha...	174182.876741
1640	Gypsy Water Eau de Parfum	Bergamot, lemon, pepper, juniper berries, inc...	This entrancing shape-shifter is one of the m...	174153.486327
469	Oliver Eau de Parfum	bitter orange, mojito, grapefruit, bergamot, ...	Oliver is an original take on the aromatic he...	172561.033154
1151	Sole di Positano Eau de Parfum	Calabrian bergamot, bitter orange, mandarin, ...	There is nothing like the thrill of a good ci...	171114.096042
1576	Russian Tea Eau de Parfum	Mint, black pepper, raspberry, black tea, mag...	Out there, the snow-covered Nevsky Prospekt, ...	170739.967660
1049	Sunny Side Up Eau de Parfum	Amyris, vanilla, sandalwood, iris butter, jas...	Leave it to iconoclast Juliette Has A Gun to ...	169980.941333
1635	Terralba Eau de Parfum	Clary Sage, Lemon, Green Tangerine, Myrtle, T...	In front of you, the sea. Behind, mediterrane...	168400.570028
1518	Mississippi Medicine Eau de Parfum	Red cedar, aldehydes, frankincense distillate...	Tribal pyres burning in the forests, with smo...	166695.197137
312	The Time Eau de Parfum	Bergamot, wormwood, chamomile, oolong tea, ir...	What a beautiful, refreshing tea scent! Fans ...	166510.753634

In [43]: test_perfume_matrix_co = descr_matrix_co[test_perfume_name]
test_perfume_matrix_co

Out[43]:<1x18753 sparse matrix of type '<class 'numpy.int64'>' with 146 stored elements in Compressed Sparse Row format>

In [44]: skr2 = SimpleKNNRecommender(descr_matrix_co, name, notes, descr)

Делаем рекомендации по описаниям векторизованным CountVectorizer и на основе Евклидова расстояния:

In [46]: rec2 = skr2.recommend_for_single_object(15, test_perfume_matrix_co, cos_flag = False)
rec2

Out[46]:	Perfume Name	Notes	Description	Dist
1895	Arso Eau de Parfum	Leather, incense, pine resin, cedar leaves	The sharp, evocative scent of wood smoke - t...	1.846619e+07
182	L'Etre Aime - Homme Eau de Parfum	bergamot, lavender, ginger, basil, cardamom, ...	Yvon Mouchel, dedicated the 7th addition to t...	1.902630e+07
1664	Hinoki Eau de Toilette	Cypress, turpentine, camphor, cedar, thyme, p...	Hinoki takes its name from the hinoki cypress...	1.920937e+07
1437	Ambre Gris Eau de Parfum	Davana, rose, geranium, sandalwood, cedar, am...	Ambergris, the legendarily rare ingredient pr...	1.928730e+07
820	And The World Is Yours Extrait de Parfum	Neroli, cumin, orange blossom absolute, rose,...	Few photographs capture the sumptuousness of ...	1.931321e+07
1170	Brooklyn Eau de Parfum	Bergamot, squeezed lemon, orange juice, incen...	The beating creative heart of the East Coast,...	1.933908e+07
20	Gris Chanel Eau de Parfum	Fig, Black tea, Cardamom essence, Absolute of...	Quai Saint-Bernard, near the Jardin des Plant...	1.936492e+07
565	Oeilleres Parfum Extrait	Eucalyptus, broom, chamomile, lavender, cumin...	"I want to echo the flowers I photograph witho...	1.941649e+07
2071	Tubereuse Couture Eau de Parfum	kalamanzi oil, green jasmine shoots, ylang-yl...	Created in the spirit of High Fashion, Tubere...	1.944222e+07
609	Arancia di Sicilia Eau de Parfum	Blood orange (brown extraction & sfuma torchi...	In perfumery the essential oil from bitter or...	1.944222e+07
349	L'Air du desert marocain Eau de Toilette ...	Coriander, Petitgrain (Bitter orange), Lemon,...	'Imagine finding peace in a room, lying on th...	1.949359e+07
1124	Rosa Greta Eau de Parfum	Lychee, white tea, rosebud, damascus rose inf...	A hidden paradise away from prying eyes - a p...	1.954482e+07
257	L'etre Aime - Femme Eau de Parfum	neroli, bergamot, nectarine, jasmine, lily, t...	At the heart of our lives is a perfume we wil...	1.957039e+07
242	Mohur Eau de Parfum	Cardamom, coriander, ambrette, carrot, black ...	The name refers to the most valuable gold coi...	1.957039e+07
1518	Mississippi Medicine Eau de Parfum	Red cedar, aldehydes, frankincense distillate...	Tribal pyres burning in the forests, with smo...	1.959592e+07

In []:

