

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 ИУ5-23М
Дин Но

Москва — 2021 г.

1.Цель лабораторной работы:

изучение разработки рекомендательных моделей.

2. Введение в систему рекомендаций

Основываясь на цене звездных отелей в различных городах мира и другой информации, система рекомендует отели и номера в соответствии с требованиями клиентов.

3. Конкретный код

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from keras.models import Sequential
from keras.layers.core import Dense,Activation,Dropout
from keras.layers import LSTM
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
import seaborn as sns
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from xgboost import XGBClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from keras.callbacks import EarlyStopping
import math
import os
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.metrics.pairwise import linear_kernel, cosine_similarity
from nltk.stem.snowball import SnowballStemmer
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import wordnet
from nltk.stem import WordNetLemmatizer

hotel_details=pd.read_csv('drive/MyDrive/hotel/Hotel_details.csv',delimiter=',')
hotel_rooms=pd.read_csv('drive/MyDrive/hotel/Hotel_Room_attributes.csv',delimiter=',')
hotel_cost=pd.read_csv('drive/MyDrive/hotel/hotels_RoomPrice.csv',delimiter=',')

hotel_details.head()
```

	id	hotelid	hotelname	address	city	country	zipcode	propertytype	starrating	latitude	longitude	Source	url	curr
0	46406	1771651	Meditaran Bungalow Galeb	Vukovarska 7	Omisi	Croatia	21310.0	Holiday parks	4	43.440124	16.682505	2	https://www.booking.com/hotel/hr/bungalow-luxu...	EUR
1	46407	177167	Hotel Polonia	Plac Teatralny 5	Torun	Poland	NaN	Hotels	3	53.012329	18.603800	5	https://www.agoda.com/en-gb/hotel-polonia/hote...	EUR
2	46408	1771675	Rifugio Sass Bece	Belvedere del Pordo,1	Canazei	Italy	38032.0	Hotels	3	46.477920	11.813350	2	http://www.booking.com/hotel/it/rifugio-sass-b...	EUR
3	46409	177168	Madalena Hotel	Mykonos	Mykonos	Greece	84600.0	Hotels	3	37.452316	25.329849	5	https://www.agoda.com/en-gb/madalena-hotel/hot...	EUR
4	46410	1771718	Pension Morenfeld	Mair im Korn Strasse 2	Lagundo	Italy	39022.0	Hotels	3	46.682780	11.131736	2	http://www.booking.com/hotel/it/pension-morenf...	EUR

```
hotel_rooms.head()
```

	id	hotelcode	roomamenities	roomtype	ratedescription
0	50677497	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15 m²/161 ft², Shower, 1 king bed
1	50672149	8328096	Air conditioning: ;Closet: ;Fireplace: ;Free W...	Vacation Home	Shower, Kitchenette, 2 bedrooms, 1 double bed ...
2	50643430	8323442	Air conditioning: ;Closet: ;Dishwasher: ;Firep...	Vacation Home	Shower, Kitchenette, 2 bedrooms, 1 double bed ...
3	50650317	7975	Air conditioning: ;Clothes rack: ;Coffee/tea m...	Standard Triple Room	Room size: 20 m²/215 ft², Shower, 3 single beds
4	50650318	7975	Air conditioning: ;Clothes rack: ;Coffee/tea m...	Standard Triple Room	Room size: 20 m²/215 ft², Shower, 3 single beds

```
del hotel_details['id']
del hotel_rooms['id']
del hotel_details['zipcode']

hotel_details=hotel_details.dropna()
hotel_rooms=hotel_rooms.dropna()

hotel_details.drop_duplicates(subset='hotelid',keep=False,inplace=True)
hotel=pd.merge(hotel_rooms,hotel_details,left_on='hotelcode',right_on='hotelid',how='inner')
hotel.columns

Index(['hotelcode', 'roomamenities', 'roomtype', 'ratedescription', 'hotelid',
      'hotelname', 'address', 'city', 'country', 'propertytype', 'starrating',
      'latitude', 'longitude', 'Source', 'url', 'curr'],
      dtype='object')
```

```
del hotel['hotelid']
del hotel['url']
del hotel['curr']
del hotel['Source']
hotel.columns

Index(['hotelcode', 'roomamenities', 'roomtype', 'ratedescription',
      'hotelname', 'address', 'city', 'country', 'propertytype', 'starrating',
      'latitude', 'longitude'],
      dtype='object')
```

Рекомендательная система, основанная только на городе

#Recommender system based only on City and ratings about the hotel							
<pre>def citybased(city): hotel['city']=hotel['city'].str.lower() citybase=hotel[hotel['city']==city.lower()] citybase=citybase.sort_values(by='starrating',ascending=False) citybase.drop_duplicates(subset='hotelcode',keep='first',inplace=True) if(citybase.empty==0): hname=citybase[['hotelname','starrating','address','roomamenities','ratedescription']] return hname.head() else: print('No Hotels Available') print('Top 5 hotels') citybased('London')</pre>							
Top 5 hotels							
	hotelname	starrating	address	roomamenities	ratedescription		
48985	The Colonnade London Hotel	4	2 Warrington Crescent	Additional toilet; ;Air conditioning; ;Alarm c...	Room size: 28 m²/301 ft², Balcony/terrace, Non...		
58663	South Point Suites – London Bridge	4	Bermondsey Street	Air conditioning; ;Alarm clock; ;Blackout curt...	Room size: 24 m²/258 ft², Street view, Non-smo...		
106943	Doubletree by Hilton London Chelsea	4	Imperial Road Imperial Wharf	Air conditioning; ;Alarm clock; ;Blackout curt...	Room size: 20 m²/215 ft², Street view, Non-smo...		
57452	Sir Christopher Wren Hotel & Spa	4	Unknown	Air conditioning; ;Alarm clock; ;Bathrobes; ;B...	Room size: 25 m²/269 ft², Non-smoking, Shower ...		
92479	The Manor Hotel Heathrow	4	Village Green, Datchet,	Additional bathroom; ;Additional toilet; ;Air ...	Shower and bathtub, 1 double bed		

Рекомендательная система, основанная только на городе и рейтингах отеля

<pre>def pop_citybased(city,number): hotel['city']=hotel['city'].str.lower() popbased=hotel[hotel['city']==city.lower()] popbased=popbased[popbased['guests_no']==number].sort_values(by='starrating',ascending=False) popbased.drop_duplicates(subset='hotelcode',keep='first',inplace=True) if popbased.empty==True: print('Sorry No Hotels Available\n tune your constraints') else: return popbased[['hotelname','roomtype','guests_no','starrating','address','roomamenities','ratedescription']].head(10) pop_citybased('London',4)</pre>							
--	--	--	--	--	--	--	--

	hotelname	roomtype	guests_no	starrating	address	roomamenities	ratedescription
1440	Holiday Inn London – Heathrow T5	Family Room	4	4	Old Bath Road	Air conditioning; ;Alarm clock; ;Blackout curt...	Room size: 25 m²/269 ft², Non-smoking, Separat...
53496	Mercure London Hyde Park Hotel	Family Room	4	4	8–14 Talbot Square	Air conditioning; ;Desk; ;Free Wi-Fi in all ro...	Room size: 31 m²/334 ft², Non-smoking, 1 singl...
60734	Hallmark Hotel London Chigwell Prince Regent	Executive Family Room	4	4	Manor Road	Air conditioning; ;Carpeting; ;Closet; ;Clothe...	Non-smoking, Shower and bathtub, 1 double bed
107246	Radisson Blu Edwardian Kenilworth – Bloomsbury	Family Room 2 adults + 2 children	4	4	97 Great Russell Street	Air conditioning; ;Alarm clock; ;Bathrobes; ;C...	Room size: 20 m²/215 ft², Non-smoking, Shower ...
55913	Acorn Lodge Gatwick	Family Stay & Park 8 Days	4	4	79 Massetts Road, Horley	Air conditioning; ;Carpeting; ;Closet; ;Coffee...	Shower, 1 double bed and 1 bunk bed
64171	Gatwick Cambridge Hotel	Family Room	4	4	19 Bonehurst Road, Horley	Air conditioning; ;Alarm clock; ;Carpeting; ;C...	Room size: 15 m²/161 ft², Non-smoking, Shower ...
96611	Trumbles Gatwick B&B	Family Room (2 Adults and 2 Children) – Number 6	4	4	Stan Hill	Air conditioning; ;Alarm clock; ;Carpeting; ;C...	Room size: 18 m²/194 ft², Shower and bathtub, ...
97400	Heathrow/Windsor Marriott Hotel	Deluxe Room, Guest room, 2 Double, Family Room	4	4	Ditton Road, Langley	Air conditioning; ;Alarm clock; ;Carpeting; ;C...	Room size: 30 m²/323 ft², Non-smoking, Shower ...
65811	Atrium Hotel Heathrow	Family 1	4	4	Great south-west road Feltham	Additional toilet; ;Air conditioning; ;Alarm c...	Room size: 28 m²/301 ft², Shower, 2 single bed...
112259	The Level at Melia White House	The Level Family Room	4	4	Albany Street	Air conditioning; ;Bathrobes; ;Blackout curtai...	Shower and bathtub, Executive lounge access, 1...

Рекомендации, основанные на требованиях и особых потребностях

```
#Requirement And special needs based Recommender

import nltk
nltk.download('stopwords')
nltk.download('wordnet')

hotel['roomamenities']=hotel['roomamenities'].str.replace(':',',',')
def requirementbased(city,number,features):
    hotel['city']=hotel['city'].str.lower()
    hotel['roomamenities']=hotel['roomamenities'].str.lower()
    features=features.lower()
    features_tokens=word_tokenize(features)
    sw = stopwords.words('english')
    lemm = WordNetLemmatizer()
    fl_set = {w for w in features_tokens if not w in sw}
    f_set=set()
    for se in fl_set:
        f_set.add(lemm.lemmatize(se))
    reqbased=hotel[hotel['city']==city.lower()]
    reqbased=reqbased[reqbased['guests_no']==number]
    reqbased=reqbased.set_index(np.arange(reqbased.shape[0]))
    l1=[];l2=[];cos=[];
    #print(reqbased['roomamenities'])
    for i in range(reqbased.shape[0]):
        temp_tokens=word_tokenize(reqbased['roomamenities'][i])
        templ_set={w for w in temp_tokens if not w in sw}
        temp_set=set()
        for se in templ_set:
            temp_set.add(lemm.lemmatize(se))
        rvector = temp_set.intersection(f_set)
        #print(rvector)
        cos.append(len(rvector))
    reqbased['similarity']=cos
    reqbased=reqbased.sort_values(by='similarity',ascending=False)
    reqbased.drop_duplicates(subset='hotelcode',keep='first',inplace=True)
    return reqbased[['hotelname','roomtype','guests_no','starrating','address','roomamenities','ratedescription','similarity']].head(10)
```

requirementbased('London',1,'I need a extra toilet and room should be completely air conditioned.')								
	hotelname	roomtype	guests_no	starrating	address	roomamenities	ratedescription	similarity
309	The Colonnade London Hotel	Studio	1	4	2 Warrington Crescent	additional toilet,air conditioning,alarm clock...	Room size: 28 m²/301 ft², Balcony/terrace, Non...	3
927	The Cleveland Hotel	Studio Suite	1	4	39–40 Cleveland Square	air conditioning,carpeting,closet,clothes rack...	Room size: 15 m²/161 ft², Non-smoking, Shower,...	3
286	Best Western Palm Hotel	Standard Room With Three Single Beds	1	3	64–76 Hendon Way	additional toilet,air conditioning,carpeting,c...	Room size: 28 m²/301 ft², Bathtub, 3 single beds	3
77	Little Foxes Hotel	Single – 1 Person	1	3	Charlwood Road, Crawley, West sussex	additional toilet,air conditioning,alarm clock...	Room size: 15 m²/161 ft², Country/region view,...	3
1092	Holiday Inn London – Heathrow Ariel	2 Single Beds Non-smoking	1	3	118 Bath Road	additional toilet,air conditioning,carpeting,c...	Room size: 25 m²/269 ft², Non-smoking, Shower ...	3
88	Belgrave Hotel Oval	Single Room	1	4	9–13 Clapham Road	additional toilet,air conditioning,carpeting,c...	Room size: 10 m²/108 ft², Non-smoking, Shower,...	3
793	Holiday Inn Express Park Royal	2 Single Beds Non-smoking	1	3	Victoria RoadNorth Acton	air conditioning,alarm clock,carpeting,closet,...	Room size: 25 m²/269 ft², Non-smoking, Shower,...	3
407	South Point Suites – London Bridge	Deluxe Studio Suite	1	4	Bermondsey Street	additional bathroom,additional toilet,air cond...	Room size: 45 m²/484 ft², 2 bathrooms, Shower,...	3
816	Simply Rooms & Suites Hotel	Deluxe Studio Suite	1	4	21 Avonmore Road	additional toilet,air conditioning,cleaning pr...	Room size: 20 m²/215 ft², City view, Non-smoki...	3
729	Hilton London Olympia Hotel	Single Hilton Guestroom	1	4	380 Kensington High Street	additional toilet,air conditioning,alarm clock...	Room size: 15 m²/161 ft², Non-smoking, Separat...	3

Список литературы

[1] Гапанюк Ю. Е. Лабораторная работа «Подготовка обучающей и тестовой выборки,

кросс-валидация и подбор гиперпараметров на примере метода ближайших соседей»

[Электронный ресурс] // GitHub. — 2019. — Режим доступа:

[https://github.com/](https://github.com/ugapanyuk/ml_course/wiki/LAB_KNN)

[ugapanyuk/ml_course/wiki/LAB_KNN](https://github.com/ugapanyuk/ml_course/wiki/LAB_KNN) (дата обращения: 05.04.2019).

[2] Team The IPython Development. IPython 7.3.0 Documentation [Electronic resource] //

Read the Docs. — 2019. — Access mode: <https://ipython.readthedocs.io/en/stable/> (online; accessed: 20.02.2019).

[3] Waskom M. seaborn 0.9.0 documentation [Electronic resource] // PyData. — 2018. —

Access mode: <https://seaborn.pydata.org/> (online; accessed: 20.02.2019).

[4] pandas 0.24.1 documentation [Electronic resource] // PyData. — 2019. — Access mode:

<http://pandas.pydata.org/pandas-docs/stable/> (online; accessed: 20.02.2019).

[5] dronio. Solar Radiation Prediction [Electronic resource] // Kaggle. — 2017. — Access

mode: <https://www.kaggle.com/dronio/SolarEnergy> (online; accessed: 18.02.2019).

[6] Chrétien M. Convert datetime.time to seconds [Electronic resource] // Stack Overflow.

— 2017. — Access mode: <https://stackoverflow.com/a/44823381> (online; accessed: 20.02.2019).

[7] scikit-learn 0.20.3 documentation [Electronic resource]. — 2019. — Access mode: <https://scikit-learn.org/>

(online; accessed: 05.04.2019).