

## Задание 2.1.

# Тушканова Ольга Николаевна

Санкт-Петербург  
2020



# Входные данные

---

data.csv

- 40 users
- 30 movies
- $\{-1, 1, 2, 3, 4, 5\}$ ; -1 – **нет оценки**

	Movie 1	Movie 2	Movie 3	Movie ...	Movie 30
User 1	5	-1	4	...	2
User 2	1	2	-1	...	4
User 3	2	5	-1	...	1
User ...	...	...	...	...	...
User 40	-1	3	2	...	4

# Входные данные

---

context\_day.csv

- {Mon, Tue, Wed, Thu, Fri, Sat, Sun}

	Movie 1	Movie 2	Movie 3	Movie ...	Movie 30
User 1	Sun	-1	Fri	...	Wed
User 2	Sat	Fri	-1	...	Tue
User 3	Thu	Tue	-1	...	Thu
User ...	...	...	...	...	...
User 40	-1	Fri	Mon	...	Sun

# Входные данные

---

context\_place.csv

- { h , c , v }
- “h” – дома, “c” – в кинотеатре, “v” – в гостях

	Movie 1	Movie 2	Movie 3	Movie ...	Movie 30
User 1	h	-1	v	...	c
User 2	h	h	-1	...	h
User 3	c	c	-1	...	h
User ...	...	...	...	...	...
User 40	-1	h	v	...	c

# Задание

---

1. Для заданного пользователя (совпадает с вашим номером в таблице) рассчитать оценки для всех фильмов, которые он не оценил.
2. Порекомендовать заданному пользователю 1 фильм, который он посмотрел бы в выходной дома, если такой фильм есть в данных.

# Примечания

- Для расчета оценок использовать подход user-based коллаборативной фильтрации, метод kNN, **k = 4**.
- Для расчета оценки  $\hat{r}_{ui}$  пользователя  $u$  и для фильма  $i$  использовать формулу

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v=1}^k sim_{vu} \cdot (r_{vi} - \bar{r}_v)}{\sum_{v=1}^k |sim_{vu}|}$$

где  $\hat{r}_{ui}$  - рассчитываемая оценка,  
 $\bar{r}_u$  - средняя оценка у пользователя  $u$ ,  
 $\bar{r}_v$  - средняя оценка у пользователя  $v$ ,  
 $r_{vi}$  - оценка пользователя  $v$  для фильма  $i$ ,  
 $sim_{vu}$  - значение метрики сходства для пользователей  $u$  и  $v$

# Примечание 1

---

- При расчете сходства использовать только те фильмы, для которых у обоих пользователей есть оценка.
- В качестве метрики сходства использовать метрику косинуса:

$$sim_{u,v} = \frac{\sum_{i=1}^m u_i v_i}{\sqrt{\sum_{i=1}^m u_i^2} \sqrt{\sum_{i=1}^m v_i^2}}$$

где  $u_i$  - оценка пользователя  $u$  для фильма  $i$ ,  
 $v_i$  - оценка пользователя  $v$  для фильма  $i$ ,  
 $m$  - количество фильмов, для которых у обоих пользователей есть оценка

## Примечание 2

- При выработке контекстных рекомендаций (з. 2) использовать свой подход.
- При всех подсчетах используйте округление до **3го знака** после запятой.



# Выходные данные

---

- Выход алгоритма оформить в виде json-файла.

Формат json

```
{  
  "user": 3,  
  "1": {  
    "movie 1": 2.6,  
    "movie 2": 3.3,  
    "movie 3": 1.5  
  },  
  "2": {  
    "movie 7": 4.4  
  }  
}
```

# Как сдать?

---

0. Не списывать.
1. Создать репозиторий на GitHub. Указать ссылку на него в Google-таблице.
2. Выгрузить код своего рекомендательного алгоритма.
3. Json-файл выгрузить на курс в «Сдать задание №2.1. Рекомендательные системы: сделай сам».
4. Получить мой комментарий.
5. Подойти ко мне на практике и обсудить код.
6. После 27.10 баллы снижаются на 30%.

# Баллы

---

- 0-10 за первое подзадание.
- 0-10 за второе подзадание.

**Спасибо за внимание!**