

Практическая работа №15

Content-based filtering

Цель работы: разработка и внедрение системы рекомендаций на основе контентного подхода с использованием алгоритмов машинного обучения и обработки данных.

Ход работы

Content-Based Filtering (Контентная фильтрация) – это метод рекомендаций, основанный на анализе характеристик объектов и предпочтений пользователя. В контентно-ориентированных рекомендательных системах каждому объекту (например, фильму, книге или товару) присваиваются определённые характеристики, и система предлагает пользователю те объекты, которые схожи с теми, что ему уже понравились.

В данной практической мы будем работать с той же таблицей, что и в предыдущей. Для начала мы удалим лишних столбцы, обработку жанров, кодирование жанров в бинарный формат.

Мы будем использовать метод однократного кодирования чтобы преобразовать список жанров в вектор, где каждый столбец соответствует одному возможному значению функции. Это кодирование необходимо для подачи категориальных данных. В этом случае мы сохраняем каждый отдельный жанр в столбцах, которые содержат либо 1, либо 0. 1 показывает, что фильм имеет этот жанр, а 0 - что нет:

movieId	title	genres	year	Adventure	Animation	Children	Comedy	Fantasy	Romance	Horror	Mystery	Sci-Fi	IMAX	Documentary	War	Musical	Western	Film-Noir	(no genres listed)
0	1	Toy Story (1995)	[Adventure, Animation, Children, Comedy, Fantasy]	1995	1.0	1.0	1.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	2	Jumanji (1995)	[Adventure, Children, Fantasy]	1995	1.0	0.0	1.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	3	Grumpier Old Men (1995)	[Comedy, Romance]	1995	0.0	0.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	4	Waiting to Exhale (1995)	[Comedy, Drama, Romance]	1995	0.0	0.0	0.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	5	Father of the Bride Part II (1995)	[Comedy]	1995	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Рисунок 1 – таблица фильмов с жанрами

Метод **Content-Based Filtering** позволяет выяснить, какие аспекты товара нравятся пользователю, а затем рекомендовать товары, которые отражают эти аспекты. В нашем случае мы попытаемся определить любимые жанры на основе представленных фильмов и оценок.

Создадим пользователя, который будет рекомендовать фильмы другим пользователям, добавим идентификатор фильма для ввода пользователем. После завершения ввода мы извлекаем идентификатор входного фильма из фрейма данных `movies` и добавим его в него:

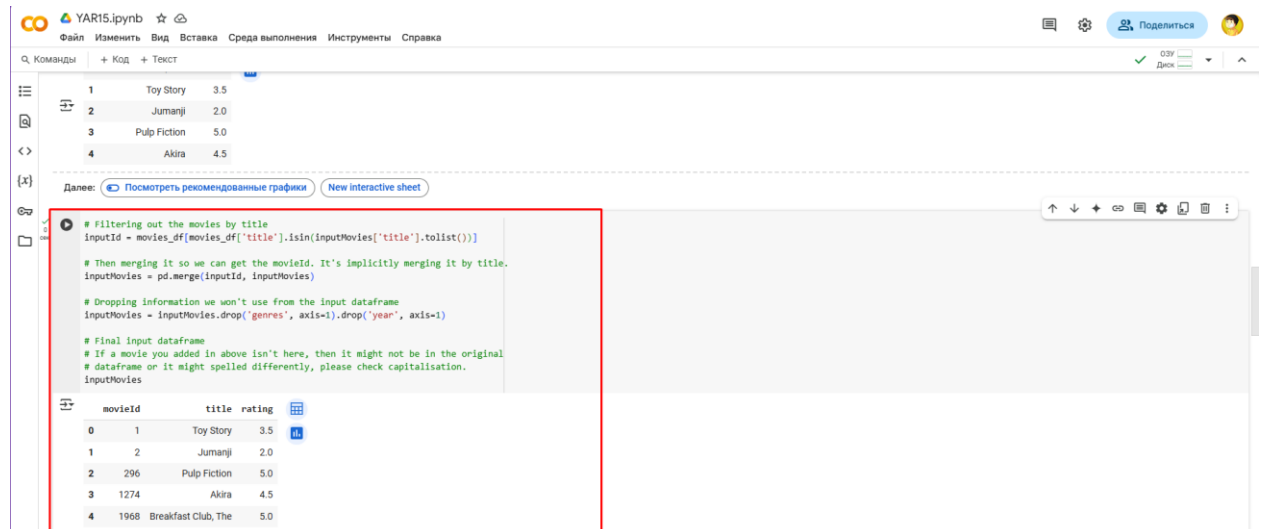


Рисунок 1.1 – рекомендованные пользователем фильмы

Далее распределяем каждый жанр в категории. Мы можем сделать это, используя входные отзывы и умножив их на входную таблицу жанров, а затем суммируя полученную таблицу по столбцам:

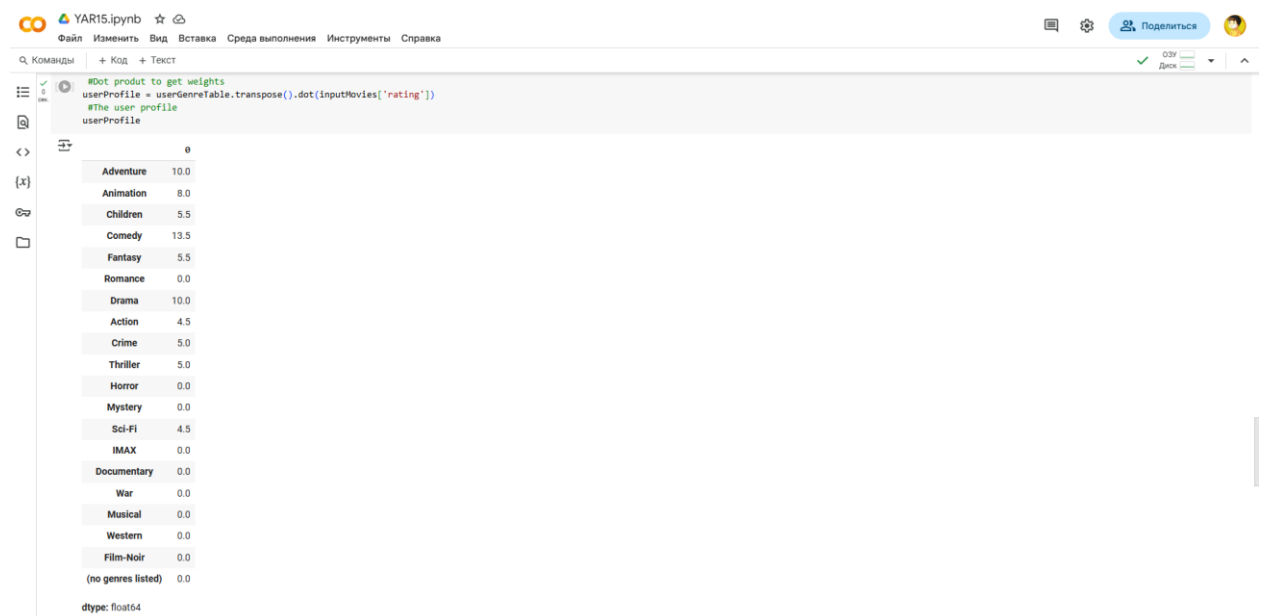
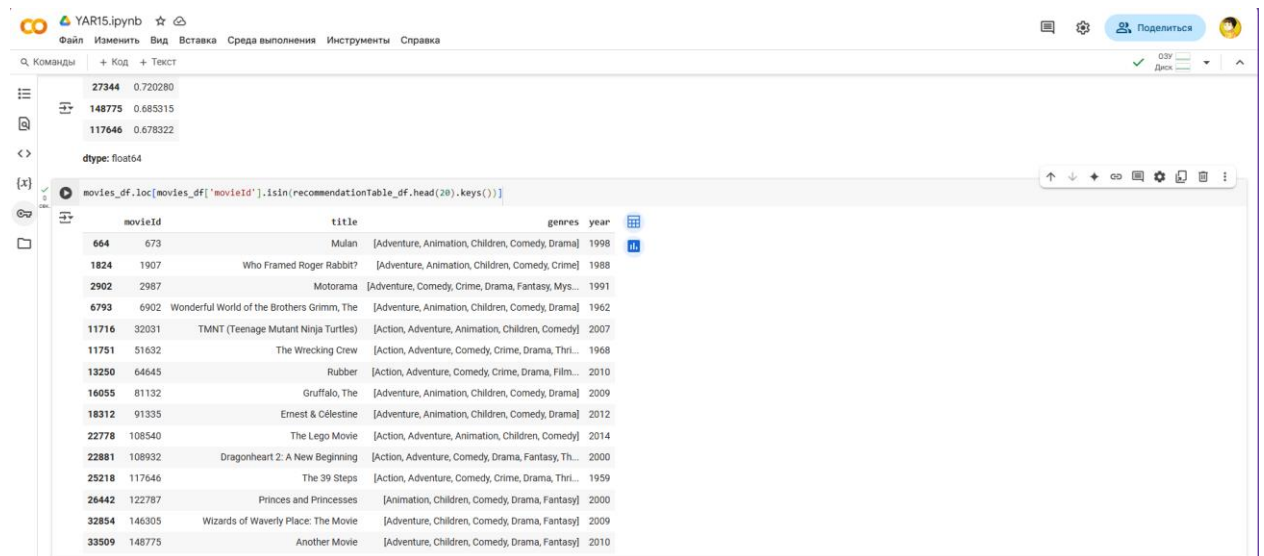


Рисунок 1.2 – распределение жанров по категориям и их количество

Теперь рассчитаем средневзвешенную оценку каждого фильма на основе входного профиля и порекомендовать 20 лучших фильмов:



The screenshot shows a Jupyter Notebook interface with a pandas DataFrame containing 20 movie recommendations. The DataFrame has columns: movieid, title, genres, and year. The movies listed are:

movieid	title	genres	year
664	Mulan	[Adventure, Animation, Children, Comedy, Drama]	1998
1824	Who Framed Roger Rabbit?	[Adventure, Animation, Children, Comedy, Crime]	1988
2902	Motorama	[Adventure, Comedy, Crime, Drama, Fantasy, Mys...]	1991
6793	Wonderful World of the Brothers Grimm, The	[Adventure, Animation, Children, Comedy, Drama]	1962
11716	TMNT (Teenage Mutant Ninja Turtles)	[Action, Adventure, Animation, Children, Comedy]	2007
11751	The Wrecking Crew	[Action, Adventure, Comedy, Crime, Drama, Thri...]	1968
13250	Rubber	[Action, Adventure, Comedy, Crime, Drama, Film...]	2010
16055	Gruffalo, The	[Adventure, Animation, Children, Comedy, Drama]	2009
18312	Ernest & Célestine	[Adventure, Animation, Children, Comedy, Drama]	2012
22778	The Lego Movie	[Action, Adventure, Animation, Children, Comedy]	2014
22881	Dragonheart 2: A New Beginning	[Action, Adventure, Comedy, Drama, Fantasy, Th...]	2000
25218	The 39 Steps	[Action, Adventure, Comedy, Crime, Drama, Thri...]	1959
26442	Princes and Princesses	[Animation, Children, Comedy, Drama, Fantasy]	2000
32854	Wizards of Waverly Place: The Movie	[Adventure, Children, Comedy, Drama, Fantasy]	2009
33509	Another Movie	[Adventure, Children, Comedy, Drama, Fantasy]	2010

Рисунок 1.3 – рейтинг фильмов

Вывод: в результате практической работы мы познакомились с методом контентной фильтрации. Научились создавать таблицу рейтинга фильмов, имея в своем распоряжении входной профиль и полный список фильмов и их жанров.