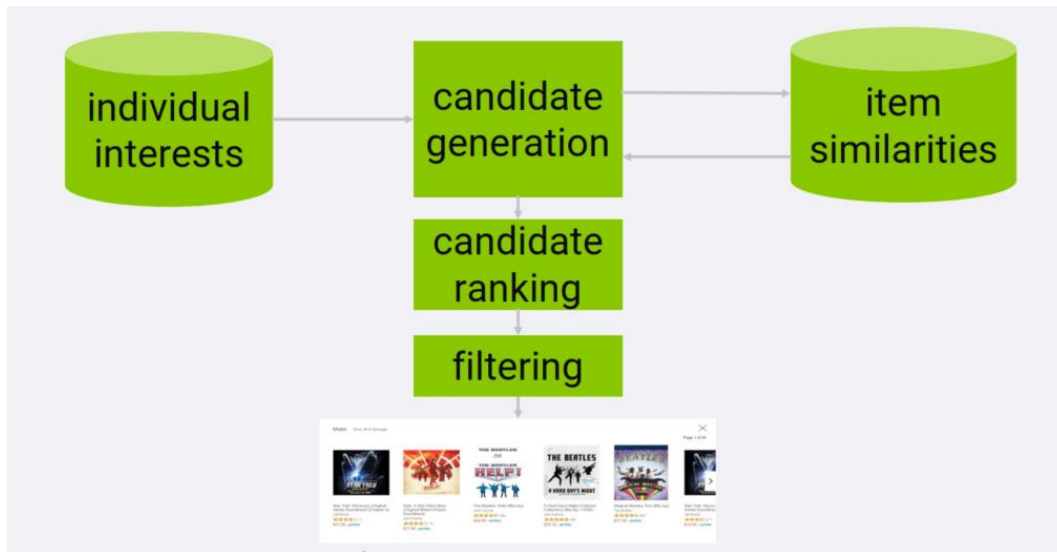
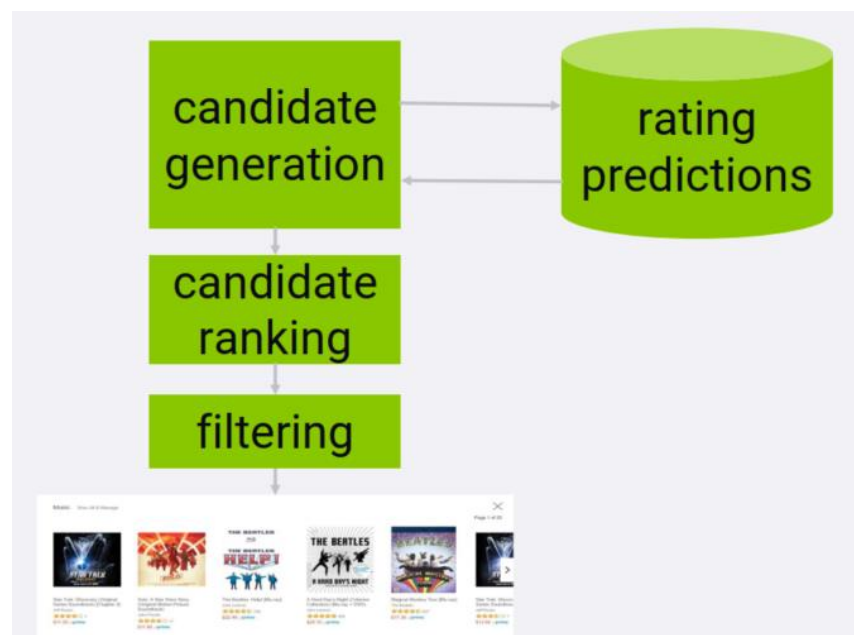


Анатомија top-N система за препоруку

-Користи се сличност између производа-



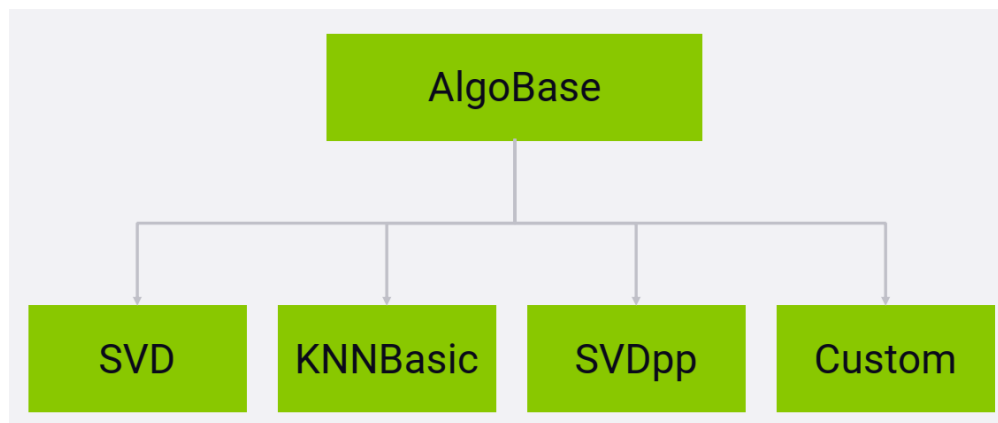
Или на други начин... користе се предикције рејтинга



Као што смо споменули, *surpriselib* је изграђена око архитектуре за предвиђања оцена сваког филма за сваког корисника и враћања најбољих (top) предикција као препорука.

Surpriselib **algorithm** base class

<https://surprise.readthedocs.io/en/stable/algobase.html>



Креирање новог алгоритма

Погледати документацију:

https://surprise.readthedocs.io/en/stable/building_custom_algo.html

Имплементација **estimate** функције:

```
class NewAlgorithm(AlgoBase):
    def __init__(self, k=40, sim_options={}):
        AlgoBase.__init__(self)

    # it's asking you to predict a rating for the user and item passed in.
    def estimate(self, user, item):
        return XX
```

Имплементација система за препорука који се базира на садржају (Content-based recommender system)

Први корак: избор атрибута који описују производе (items) тј. филмове у нашем случају

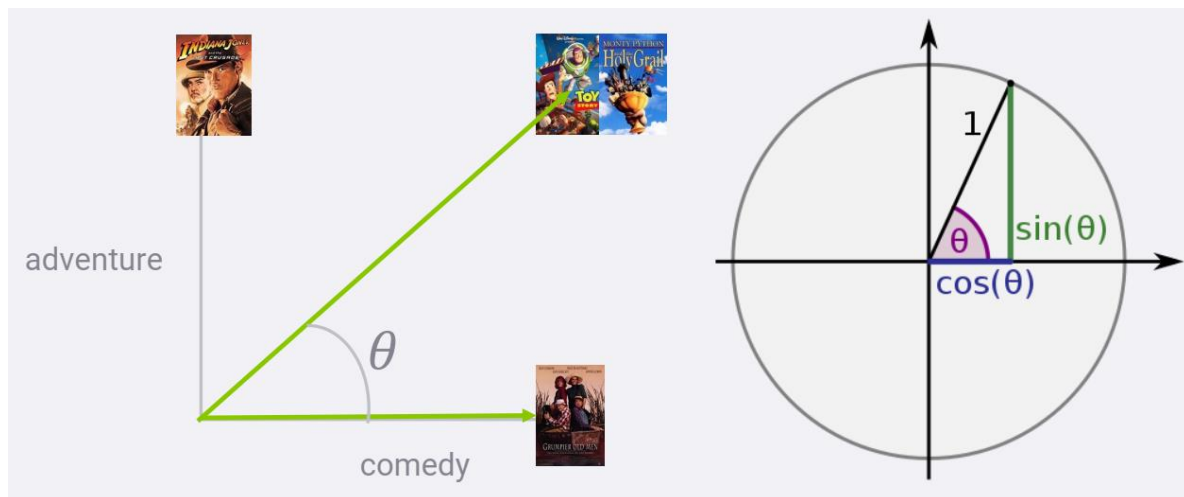
Žanrovi kojima film pripada

1808406117	The Legend of Suriyothai (2003)	Action/Adventure (у R4 скупу)
1808406133	S.W.A.T. (2003)	Crime/Gangster Thriller Action/Adventure
1808406138	Home on the Range (2004)	Comedy Musical/Performing Arts Kids/Family Animation

Action* Adventure* (у MovieLens скупу су ови жанрови одвојени) Animation* Children's* Comedy* Crime* Documentary* Drama* Fantasy* Film-Noir* Horror* Musical* Mystery* Romance* Sci-Fi* Thriller* War* Western

Косинусна сличност за 2D

Имамо само два жанра: комедија и авантура.



Косинусна сличност за вишедимензиони простор

Сви жанрови за један филм се конвертују у вектор. Пример:

1808406117: [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

1808406133: [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0]

1808406138: [0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]

1: Филмови имају један исти жанр, остали могу, а и не морају да се разликују.

Други корак: Имплементација нових метода у класи *YahooDataset*

- loadMovies()
- getGenres()
- getYears()

Трећи корак: Имплементација алгоритма за препоруку

Class ContentKNNAlgorithm (наслеђује AlgoBase класу)

1. Метода за одређивање сличности филмова на основу жанра:

computeGenreSimilarity(self, movie1, movie2, genres)

$$\text{CosSim}(x, y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}}$$

2. Метода за одређивање сличности филмова на основу године производње:

computeYearSimilarity(self, movie1, movie2, years)

$$\text{YearSim}(\text{year}_1, \text{year}_2) = e^{\frac{-|\text{year}_1 - \text{year}_2|}{10}}$$

На пример: Филм1 снимљен 2020, Филм 2 снимљен 2020. Тада функција враћа вредност 1.

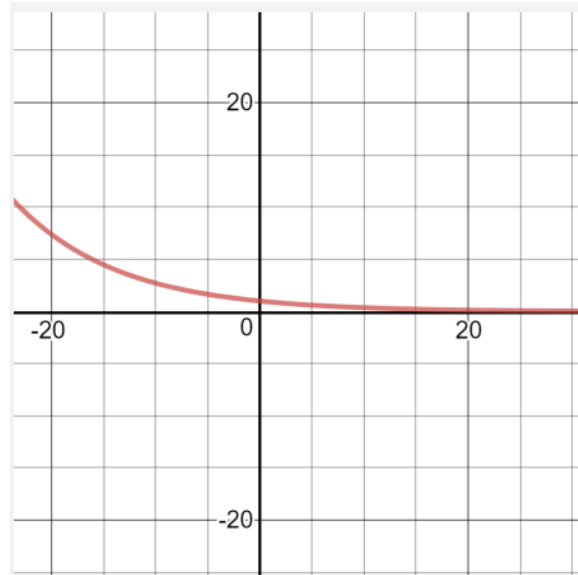


График за временску сличност.

3. Имплементација метода **fit** и **estimate**.

ОБАВЕЗНО ПРОЧИТАТИ:

- https://surprise.readthedocs.io/en/stable/basic_algorithms.html
- https://surprise.readthedocs.io/en/stable/prediction_algorithms.html#baseline-estimates-configuration