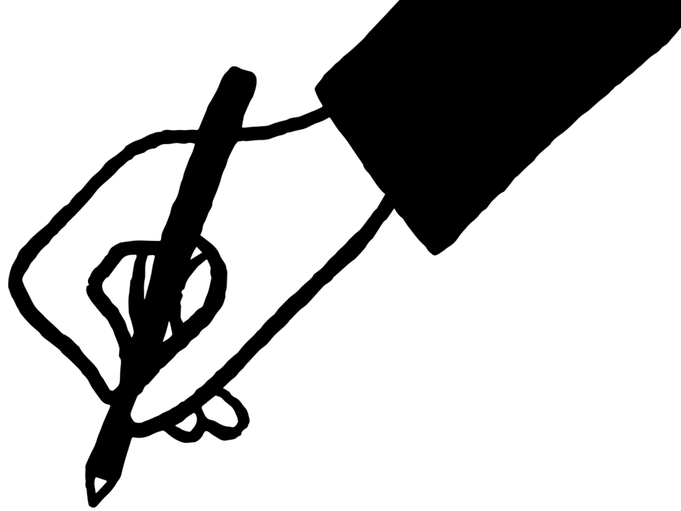
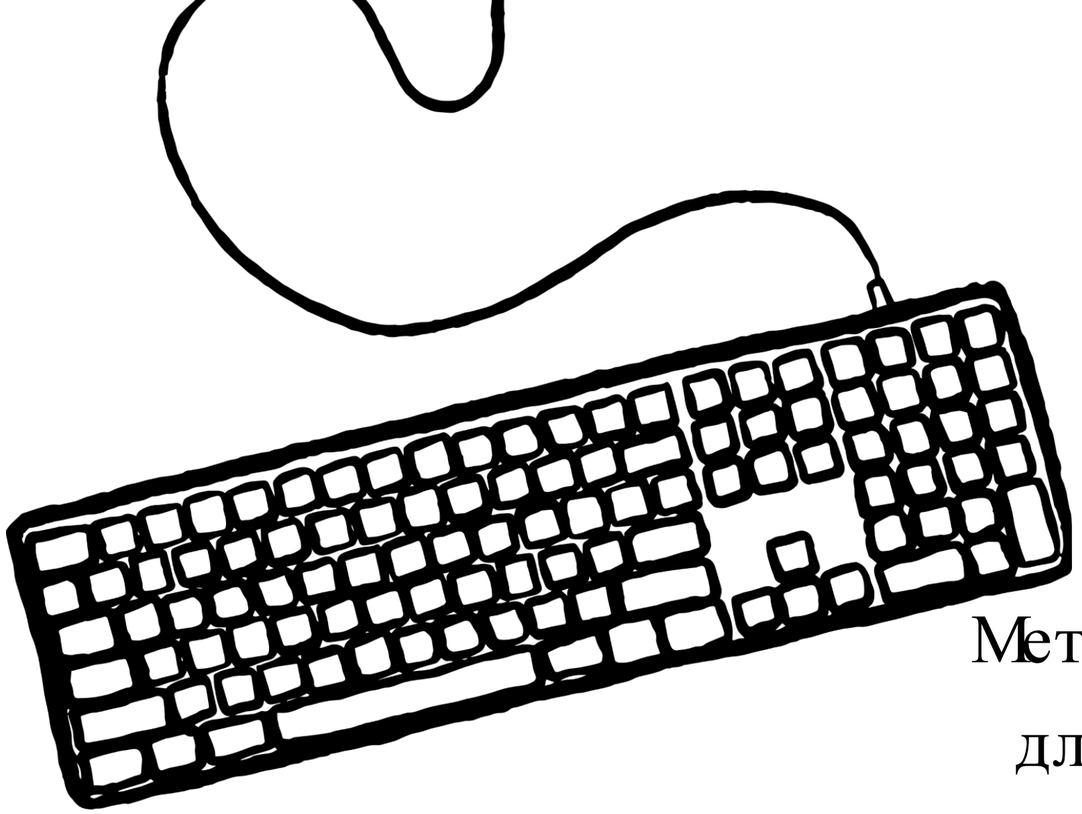


ML algorithms for recommending personalized multimedia content

based on user preferences and behavior

АЛГОРИТМИ ML ДЛЯ РЕКОМЕНДАЦІЇ ПЕРСОНАЛІЗОВАНОГО МУЛЬТИМЕДІЙНОГО КОНТЕНТУ НА ОСНОВІ ВПОДОБАНЬ І ПОВЕДІНКИ КОРИСТУВАЧІВ

науковий керівник:
Афонін Андрій
Олександрович



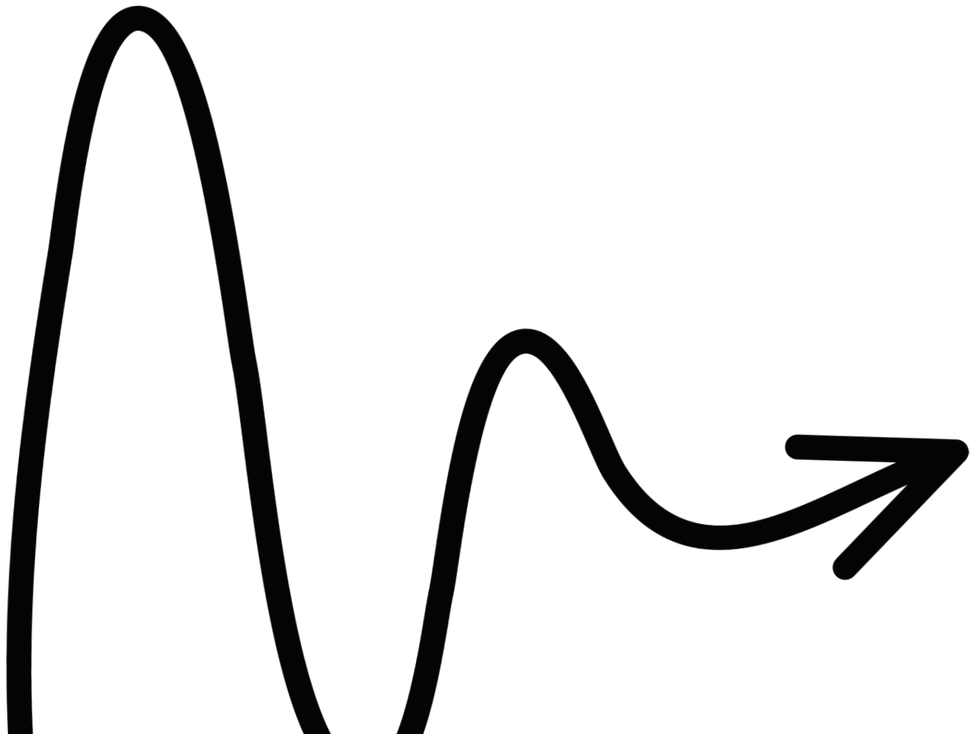
Problem Statement

Мета: представити різні алгоритми для створення рекомендаційних систем і розробити рекомендаційну систему на основі одного з цих алгоритмів

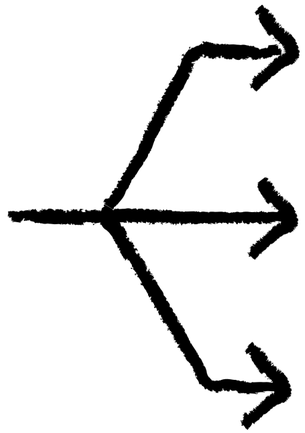
Existing Recommendation Systems

1) Netflix
2) Amazon
3) Youtube

Recommendation systems



Types of Recommendation systems

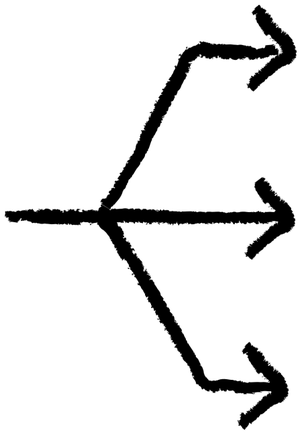


01. Collaborative Recommender system

02. Content-based recommender system

03. Demographic-based recommender system

Types of Recommendation systems

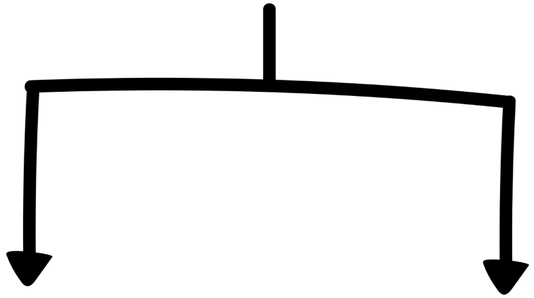


04. Utility-based recommender system

05. Knowledge-based recommender system

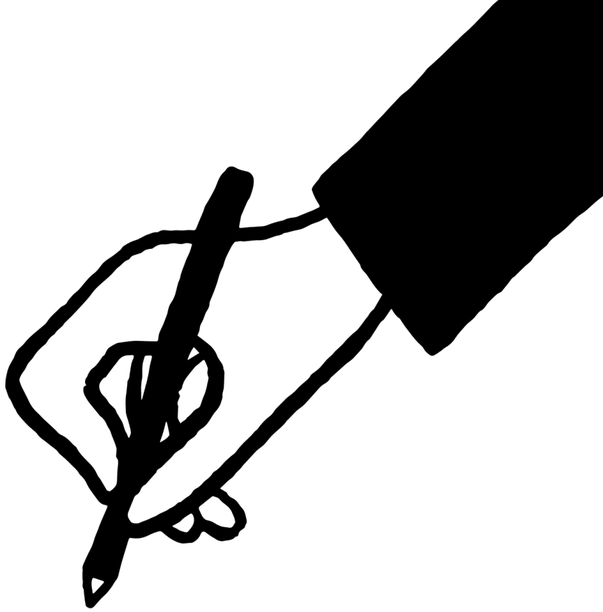
06. Hybrid recommender system

My choise

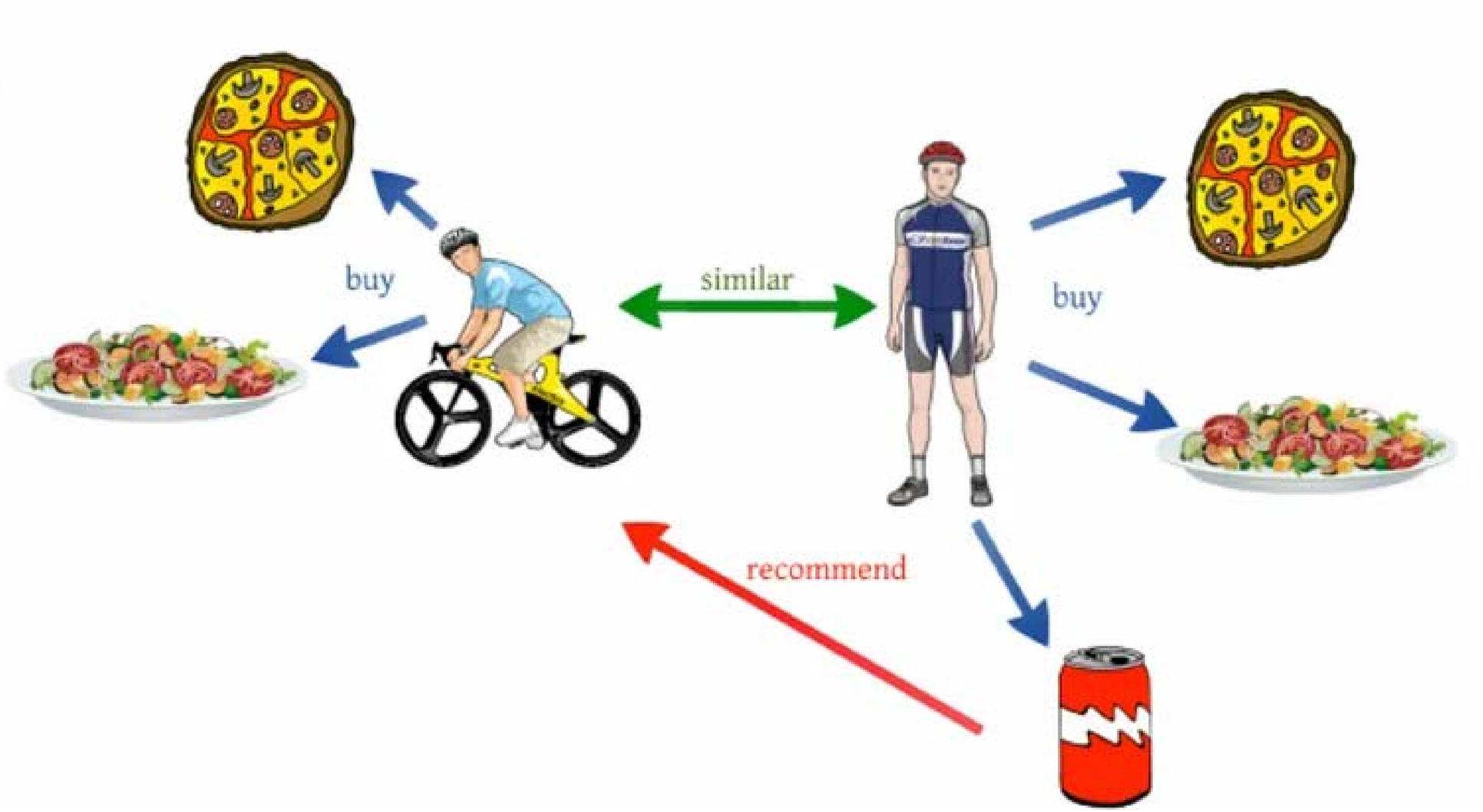


Collaborative
Recommender system

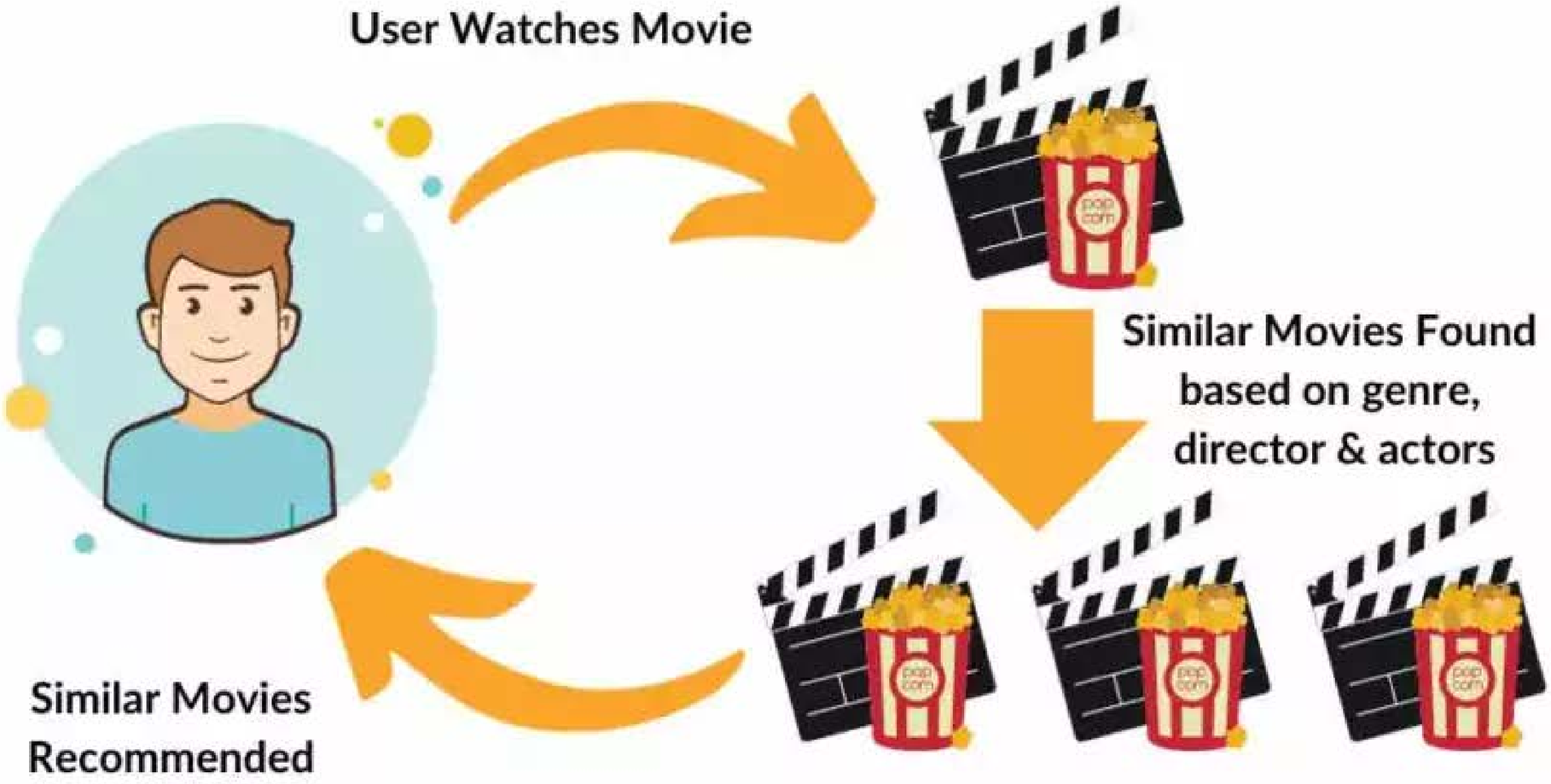
Content-based
recommender system



Collaborative Recommender system



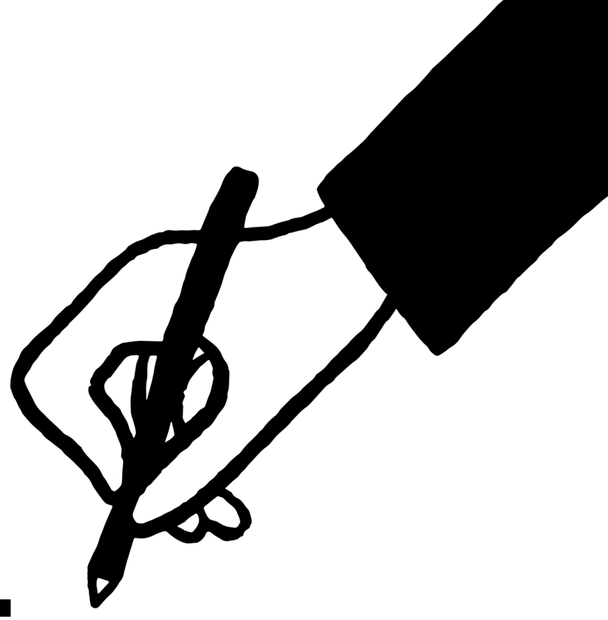
Content-based recommender system



Machine Learning algorithms

used for recommendation systems

Types of algorithms



k-найближчих сусідів

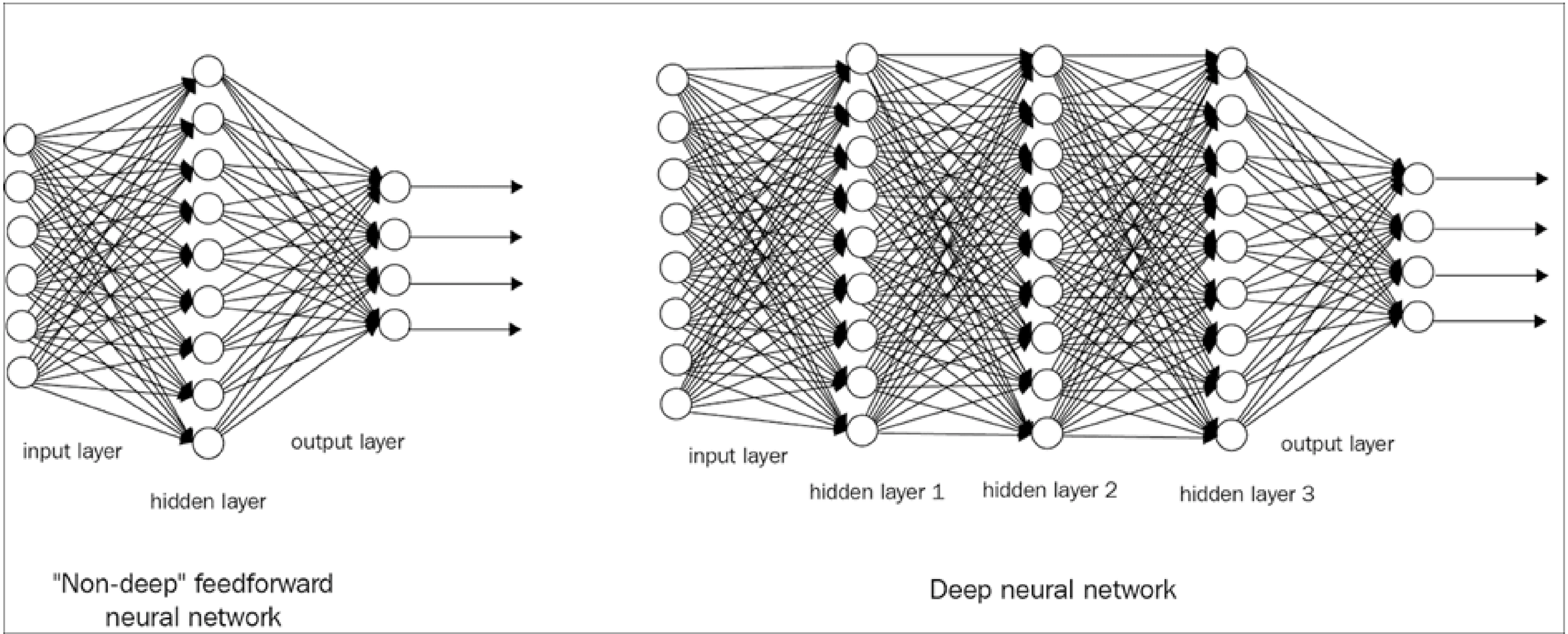
Метод з використанням факторизацією матриць

Нейронні мережі

Графові моделі

Neural networks

це обчислювальні системи, натхнені біологічними нейронними мережами, які складають мозок тварин



Matrix factorization

«Факторизація – це процес декомпозиції об’єкту (зокрема, матриці) в набір інших об’єктів (факторів), добуток яких дає початковий об’єкт»

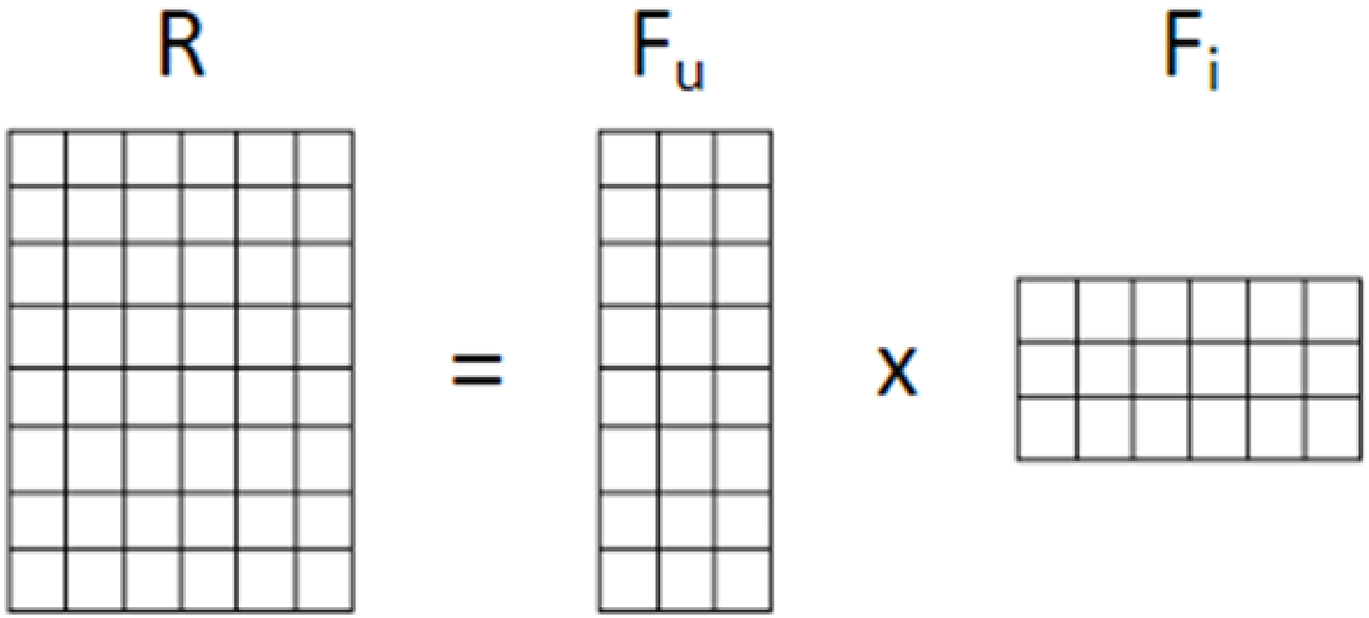
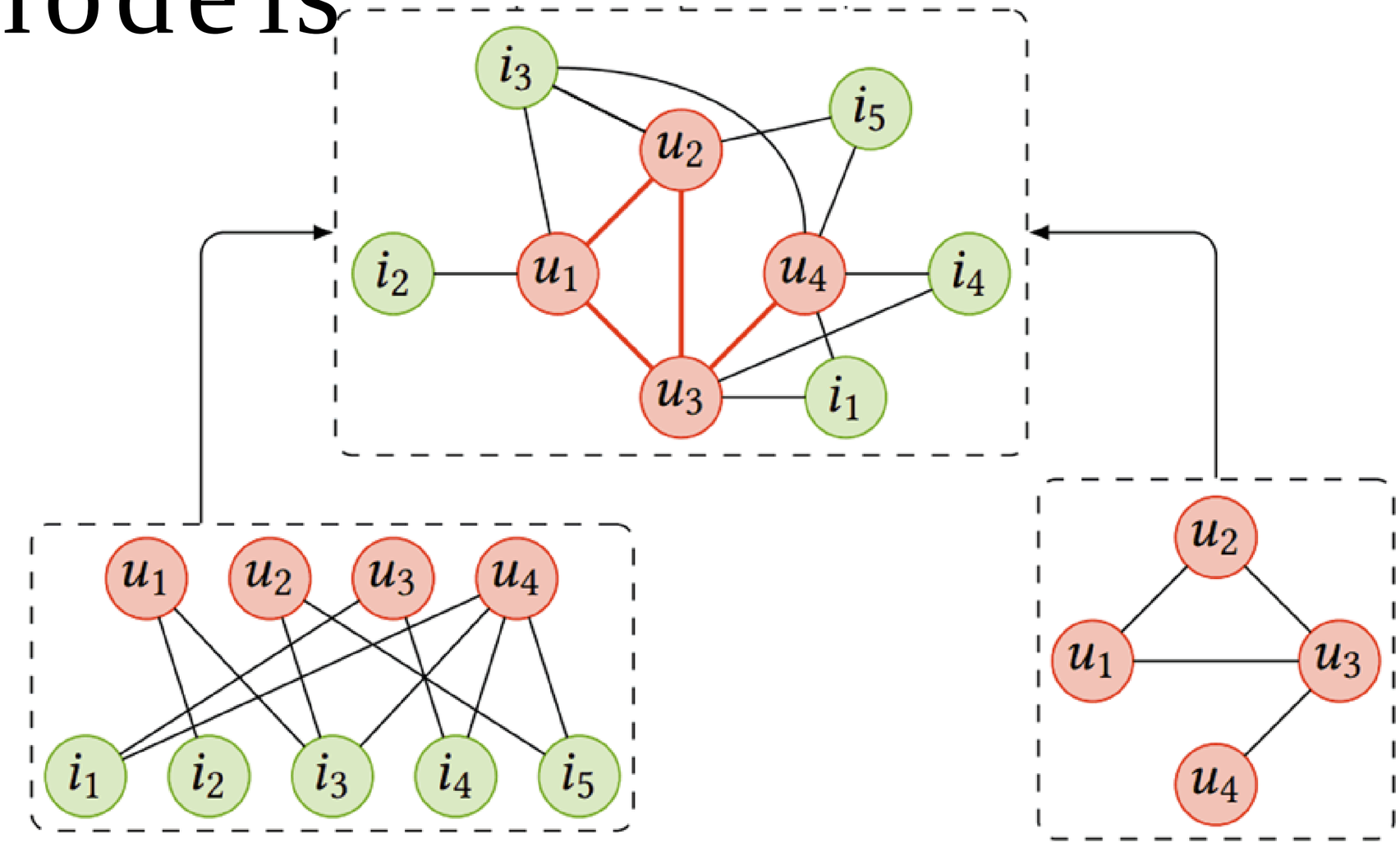


Рис. 1. Принцип факторизації матриці рейтингів

матриця прохованих факторів для користувачів Fu(n x k), де k – кількість прихованих факторів. Друга матрицю прихованих факторів об’єктів Fi(k x m).

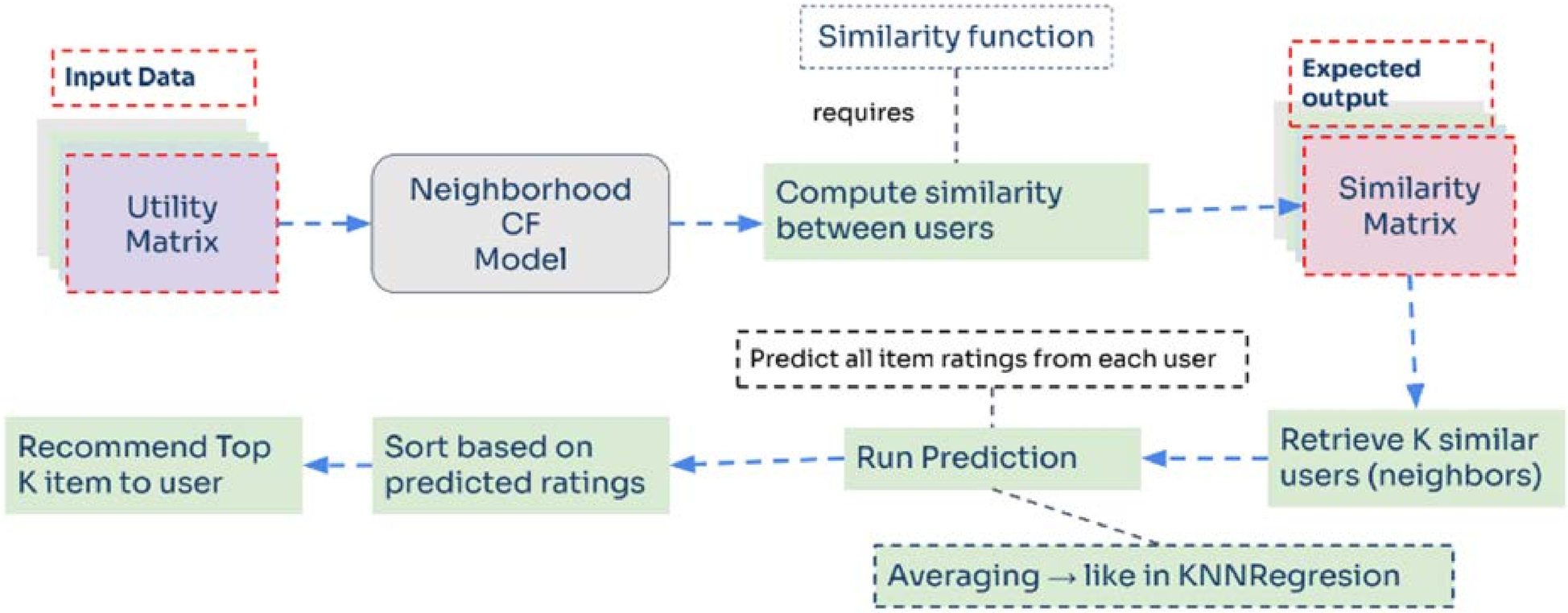
Graph models

Завдяки багатьом властивостям графів ми можемо легко вирішувати багато типів проблем з рекомендаціями за допомогою GNN.



knn

Суть цього методу доволі просто «він знаходить K-найближчі точки даних до даної точки та використовує їхні характеристики для прогнозування



Compare algorithms

Критерії	k-найближчих сусідів (kNN)	Рекомендаційні системи з використанням розкладання матриць	Нейронні мережі	Графові моделі
Простота реалізації та розуміння	Просто	Середня	Складно	Складно
Час необхідний для тренування	Не потрібен додатковий час	Середній час	Великий проміжок час залежно від к-сті даних	Великий проміжок часу(залежить від необхідної точності)
Інтерпретованість результатів	Добре	Середня	Середня	Зазвичай проста(залежить від графа)

My dataset:

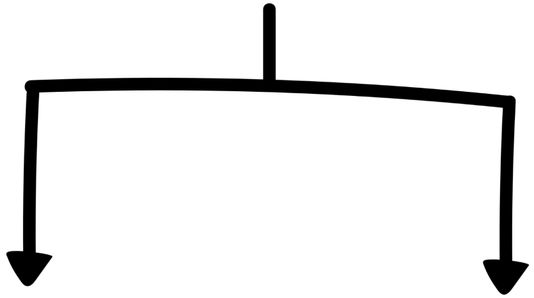
9742 фільмів, 610 користувачів

knn

Методи визначення схожості для Knn

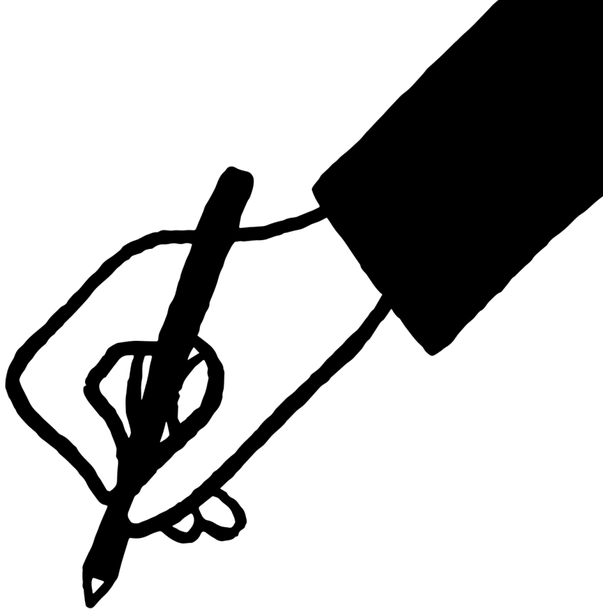
1) Euclidean distance
2) Manhattan distance
3) Minkowski Distance
4) Hamming Distance
5) Cosine similarity

Similarity metric

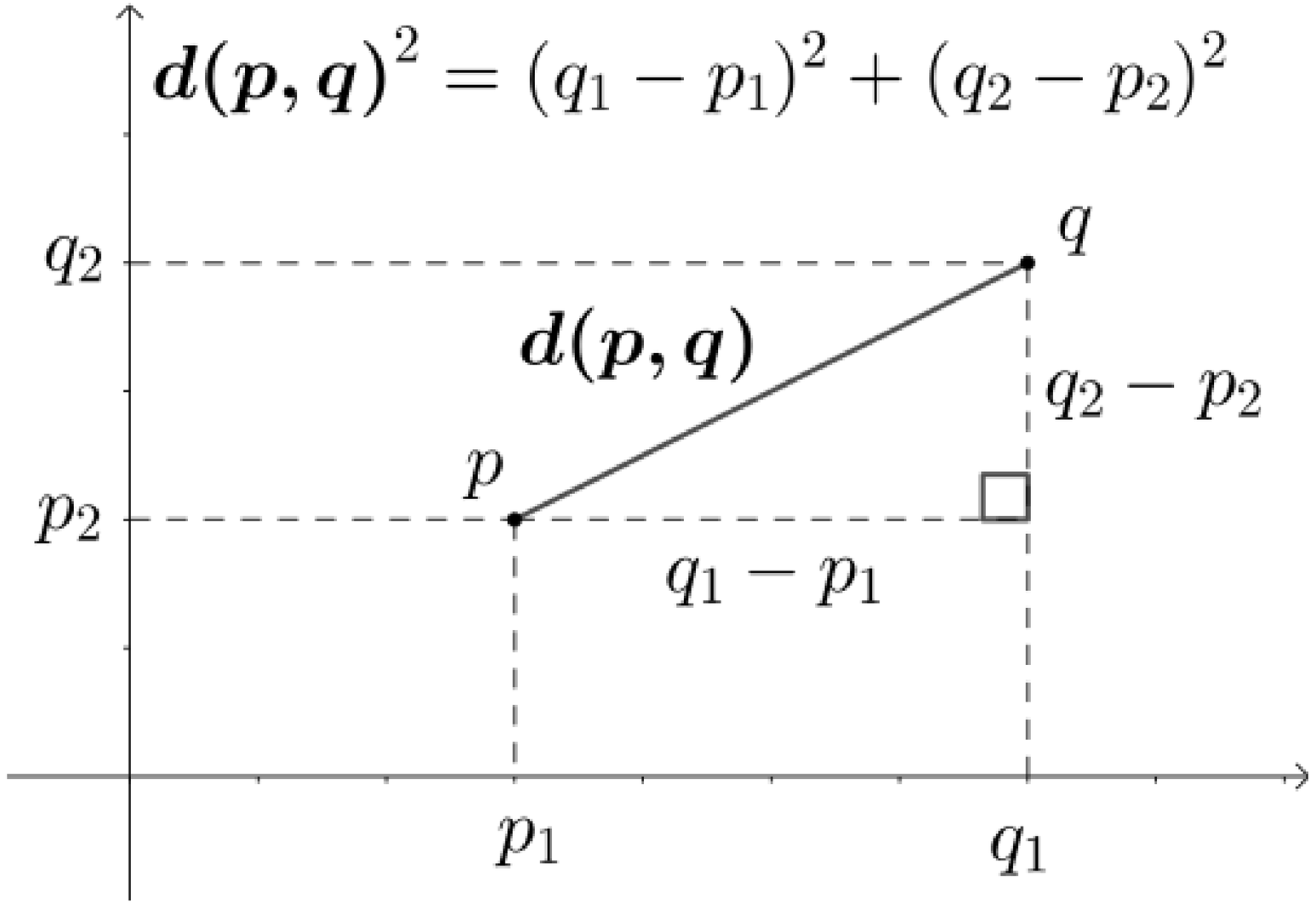


Euclidean distance

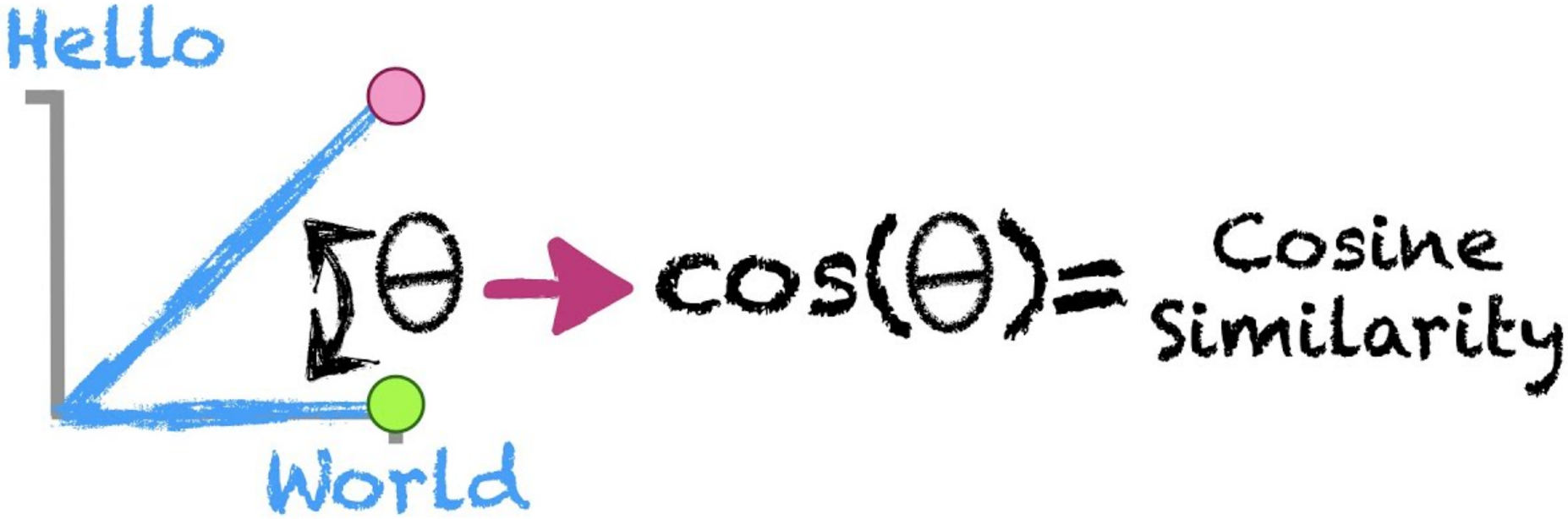
Cosine similarity



Евклідова відстань



Косинус подібності



Practical part

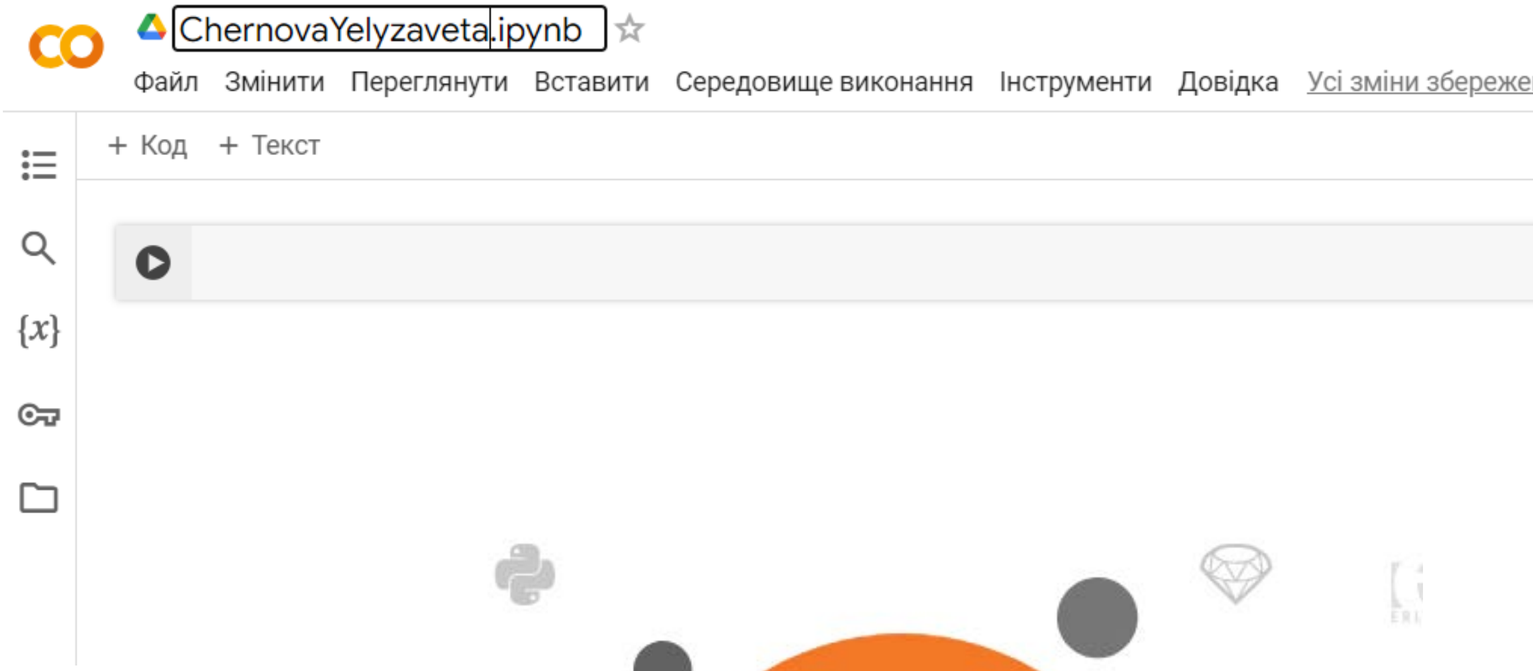
Моя задача: розробити два основних типи рекомедаційних систем(content based, collaboratory based) використовуючи алгоритм k найближчих сусідів та подібність за косинусом.

What am I using?

1)google colaboratory

2)Jupiter notebook

3)streamlit



Streamlit

What have I done?

1. content-based filtering using cosine similarity

1.1analything dataset

1.2Vectorization of information about movies

1.3cosine similarity

1.4content-based filtering

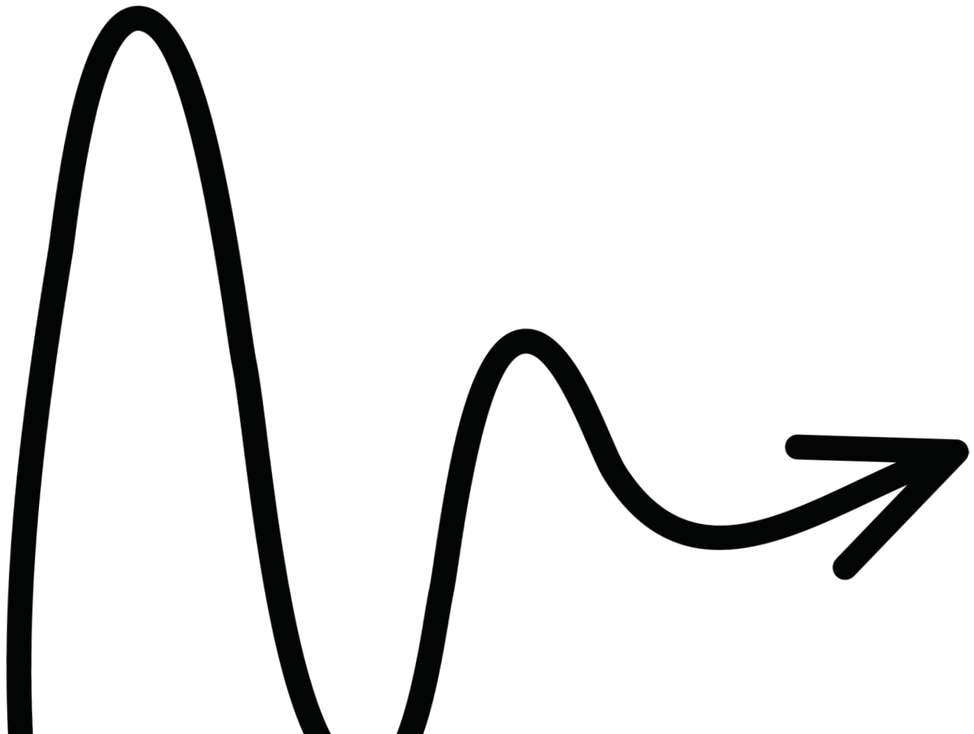
2. collaboratory-based filtering using knn

2.1analything dataset

2.2implementing knn algorithm

3. creating a web application

content-based filtering using knn



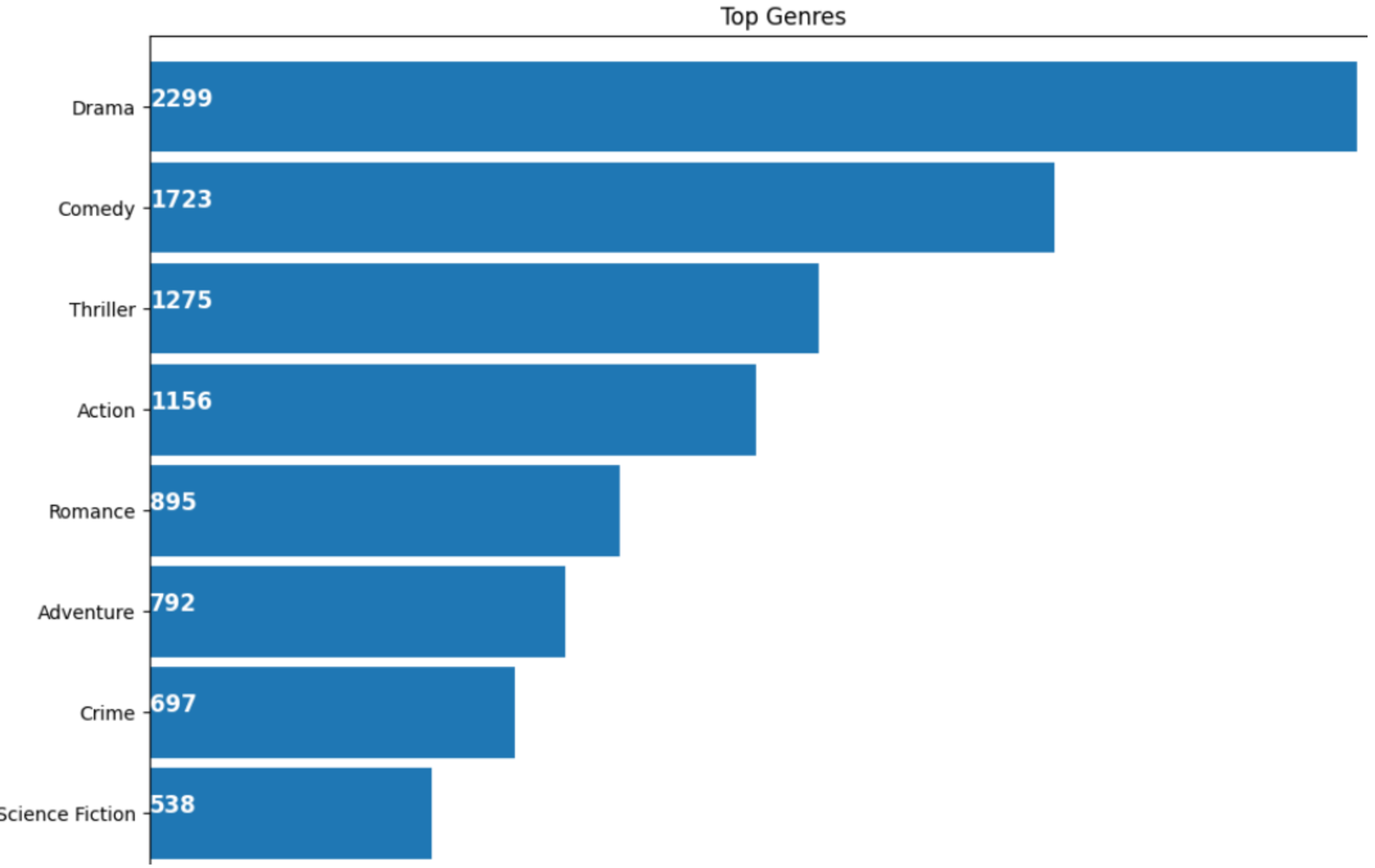
Analyt hing dat aset

Cont ent -b ased f ilt ering using cosine similarit y

	movie_id	title	overview	genres	keywords	cast	crew	tags
0	19995	Avatar	[In, the, 22nd, century,, a, paraplegic, Marin...	[Action, Adventure, Fantasy, ScienceFiction]	[cultureclash, future, spacewar, spacecolony, ...	[SamWorthington, ZoeSaldana, SigourneyWeaver]	[JamesCameron]	[In, the, 22nd, century,, a, paraplegic, Marin...
1	285	Pirates of the Caribbean: At World's End	[Captain, Barbossa,, long, believed, to, be, d...	[Adventure, Fantasy, Action]	[ocean, drugabuse, exoticisland, eastindiatrad...	[JohnnyDepp, OrlandoBloom, KeiraKnightley]	[GoreVerbinski]	[Captain, Barbossa,, long, believed, to, be, d...
2	206647	Spectre	[A, cryptic, message, from, Bond’s, past, send...	[Action, Adventure, Crime]	[spy, basedonnovel, secretagent, sequel, mi6, ...	[DanielCraig, ChristophWaltz, LéaSeydoux]	[SamMendes]	[A, cryptic, message, from, Bond’s, past, send...
3	49026	The Dark Knight Rises	[Following, the, death, of, District, Attorney...	[Action, Crime, Drama, Thriller]	[dccomics, crimefighter, terrorist, secretiden...	[ChristianBale, MichaelCaine, GaryOldman]	[ChristopherNolan]	[Following, the, death, of, District, Attorney...
4	49529	John Carter	[John, Carter, is, a, war-weary,	[Action, Adventure,	[basedonnovel, mars, medallion	[TaylorKitsch, LynnCollins,	[AndrewStanton]	[John, Carter, is, a, war-weary,

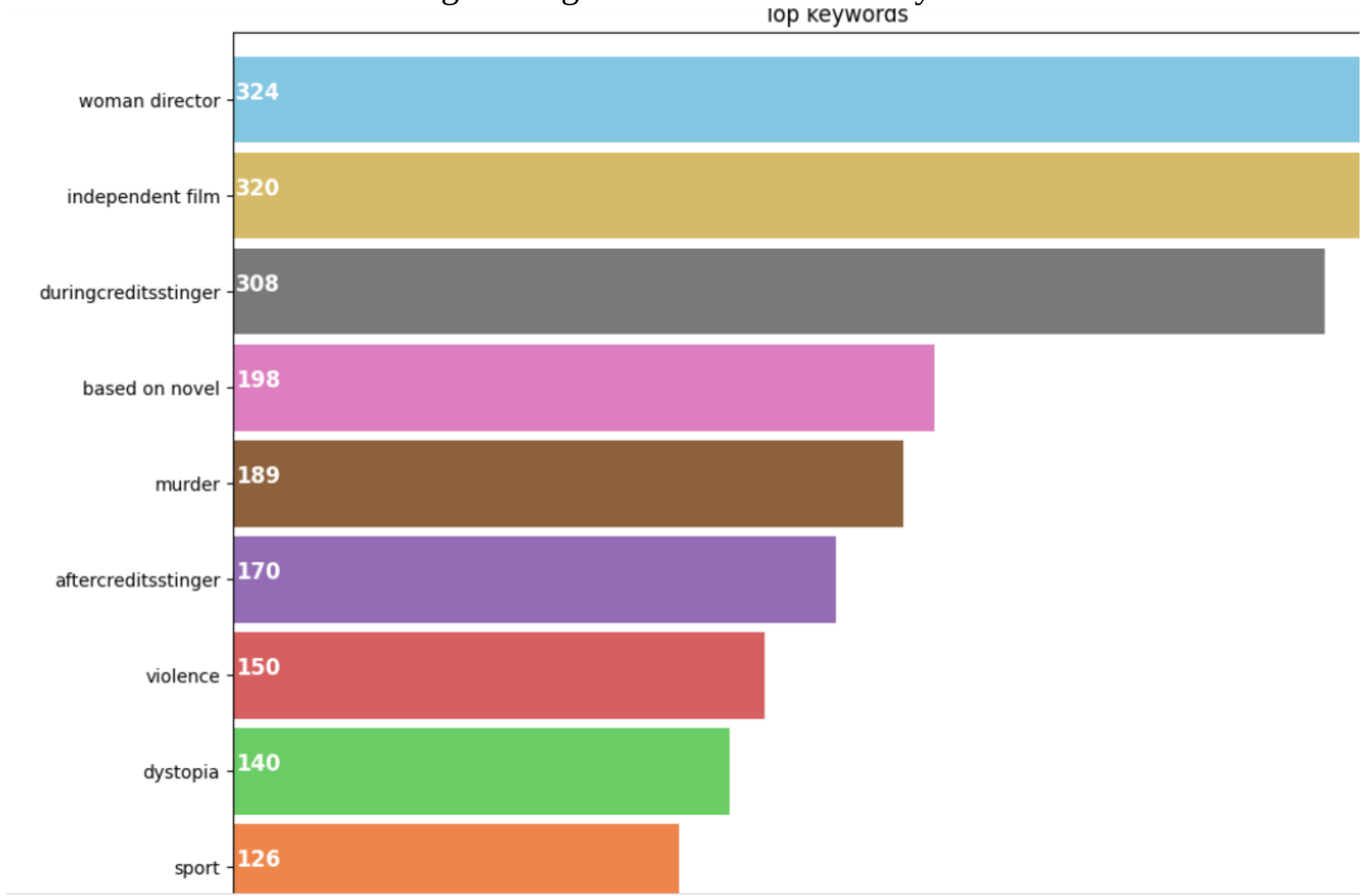
Analyt hing dat aset

Cont ent -b as ed f ilt er ing us ing cos ine s imil ar it y



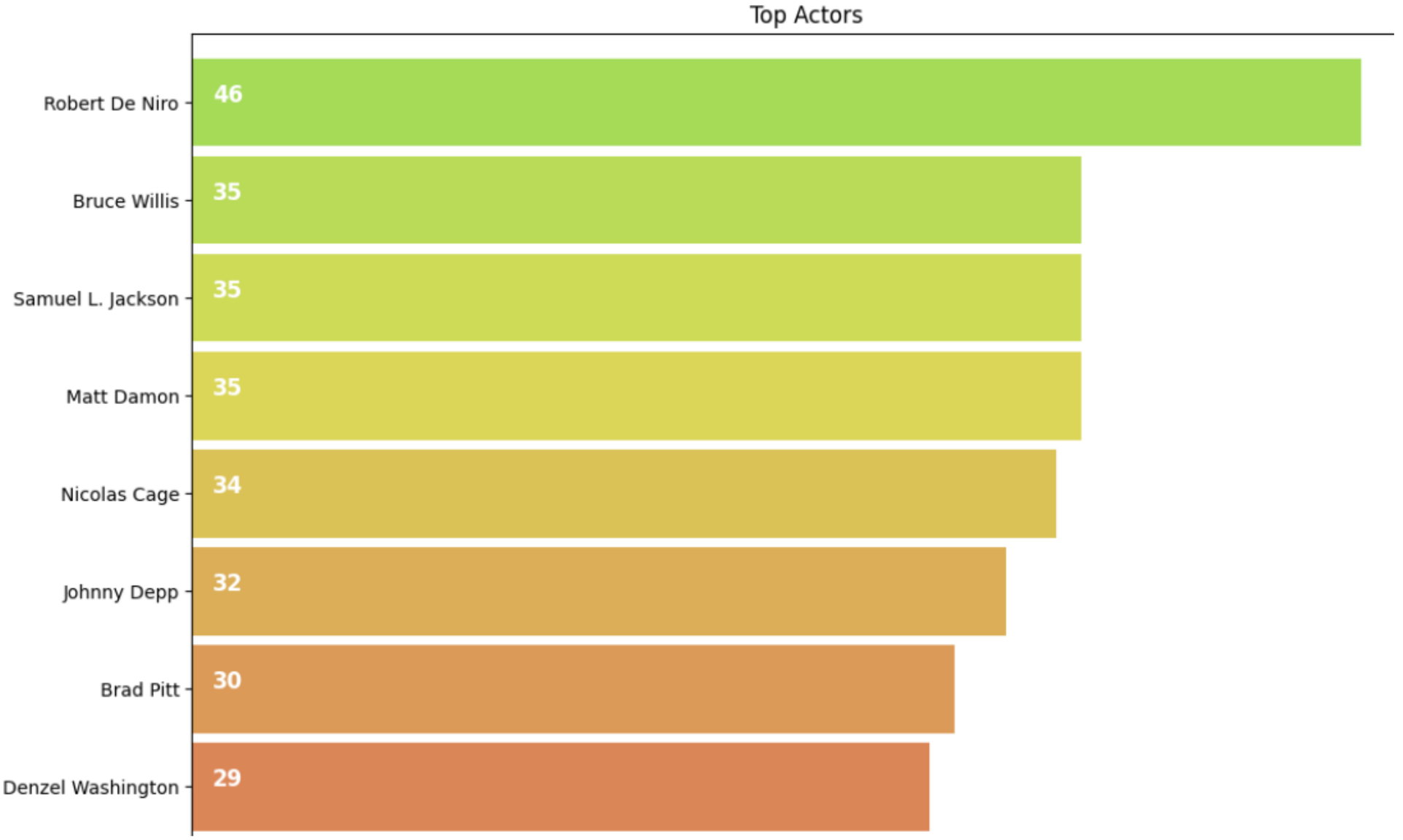
Analyt hing dat aset

Cont ent -b a s e d f ilt e r ing u s ing c o s ine s imila r it y



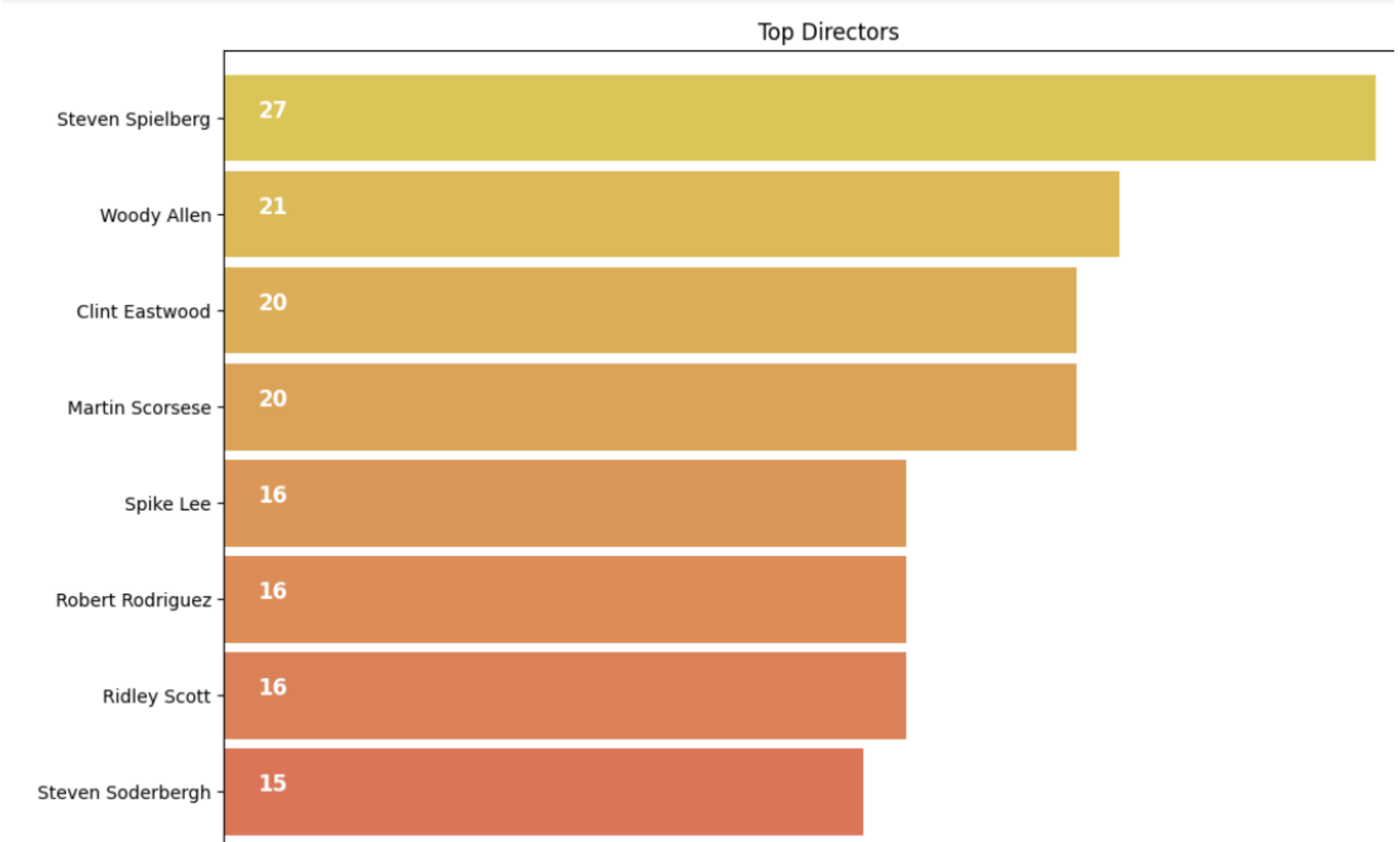
Analyt hing dat aset

Cont ent -b a s e d f ilt e r ing u s ing c o s ine s imila r it y



Analyt hing dat aset

Cont ent -b a s e d f ilt e r ing u s ing c o s ine s imila r it y



Vectorization of information about movies

Content-based filtering using cosine similarity

1)split the data into tokens

2)represent data as vectors

Vectorization of information about movies

split the data into tokens

```
['In',
 'the',
 '22nd',
 'century,',
 'a',
 'paraplegic',
 'Marine',
 'is',
 'dispatched',
 'to',
 'the',
 'moon',
 'Pandora',
 'on',
 'a',
 'unique',
 'mission,',
 'but',
 'becomes',
 'torn',
 'between',
 'following',
 'orders',
 'and',
 'protecting',
```

Vectorization of information about movies

represent data as vectors

```
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 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, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

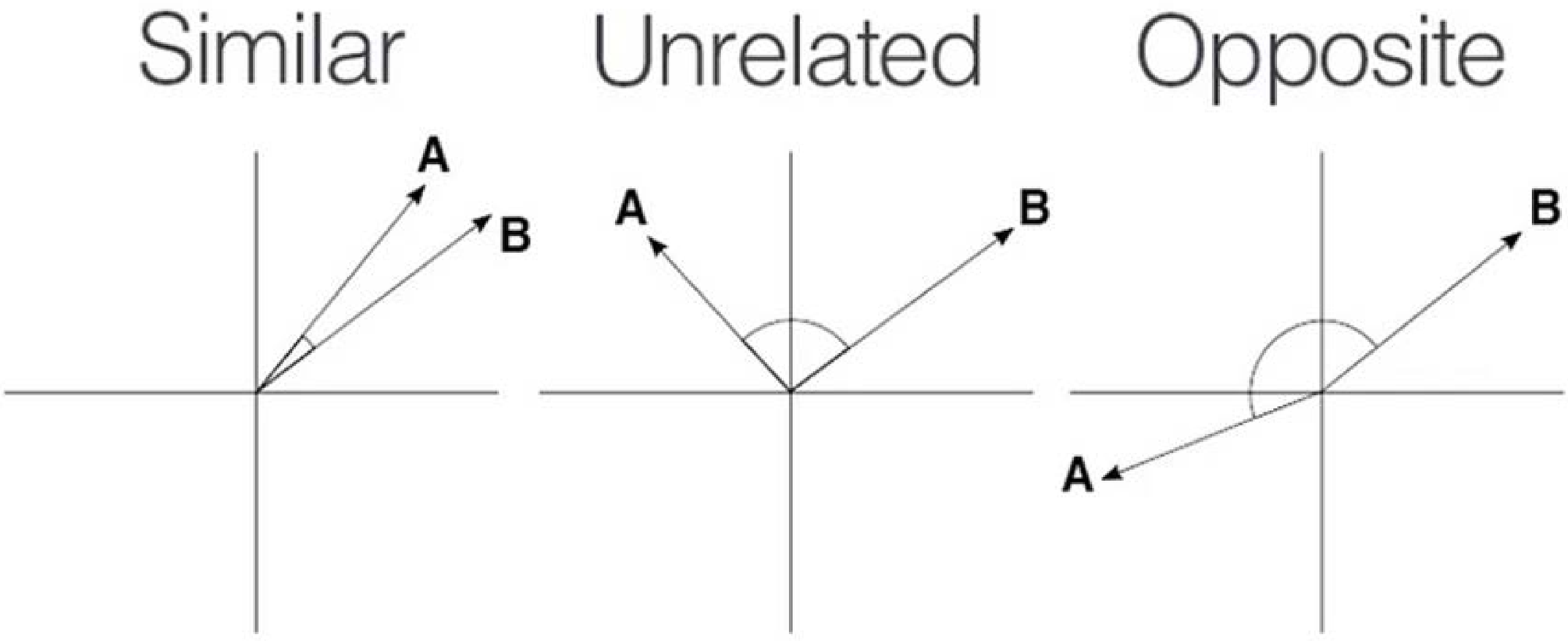
Cosine similarity

Content-based filtering using cosine similarity

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}$$

Cosine similarity

Content-based filtering using cosine similarity



Cosine similarity

Content-based filtering using cosine similarity

```
array([1.        , 0.08346223, 0.0860309 , 0.07347184, 0.18929941,
       0.10838875, 0.04024218, 0.1467348 , 0.05923489, 0.09673017,
       0.10259784, 0.0946497 , 0.09037128, 0.04499213, 0.12824729,
       0.06282809, 0.07894737, 0.13977654, 0.09493291, 0.0830813 ,
       0.0580381 , 0.1096817 , 0.06622662, 0.08740748, 0.05333807,
       0.05101628, 0.15389675, 0.18693292, 0.11654331, 0.06503325,
       0.06684848, 0.15907119, 0.08520286, 0.09733285, 0.        ,
       0.09933993, 0.17316974, 0.07894737, 0.08111071, 0.08226127,
       0.07694838, 0.16563698, 0.        , 0.09086217, 0.0338255 ,
       0.08240856, 0.13910372, 0.19672237, 0.08447772, 0.05827165,
       0.1129565 , 0.08048436, 0.14843121, 0.0429735 , 0.03121953,
       0.07038153, 0.13834289, 0.02418254, 0.06748819, 0.07254763,
       0.01836796, 0.23179316, 0.08377078, 0.08175191, 0.02961744,
       0.01952916, 0.02249606, 0.16452255, 0.12049505, 0.02649065,
       0.07600419, 0.0733674 , 0.13710212, 0.03311331, 0.20246457,
       0.10390487, 0.07175473, 0.01439482, 0.05066946, 0.06231097,
       0.05046473, 0.11020775, 0.07600419, 0.17996851, 0.12361285,
       0.11412706, 0.07098728, 0.11128298, 0.09037128, 0.05757929,
       0.06131393, 0.16692447, 0.06231097, 0.07098728, 0.18731716,
       0.14048787, 0.07694838, 0.06952347, 0.07509393, 0.02666904,
       0.06765101, 0.07694838, 0.12462195, 0.07987231, 0.05564149,
       0.10375717, 0.06231097, 0.06180642, 0.14509525, 0.06488857,
       0.04499213, 0.15018785, 0.07694838, 0.0433555 , 0.04588315,
       0.04374786, 0.04836508, 0.05923489, 0.03364316, 0.0433555 ,
       0.        , 0.11128298, 0.11654331, 0.11248032, 0.06748819,
       0.07843305, 0.13334519, 0.14601069, 0.        , 0.09269795,
       0.0270369 , 0.10526316, 0.04499213, 0.02823912, 0.12977714,
       0.        , 0.02341465, 0.03012376, 0.12824729, 0.12072655,
```

Content-based filtering

Content-based filtering using cosine similarity

```
recommend('Spider-Man 2')
```

```
Spider-Man 3
Spider-Man
The Amazing Spider-Man
Iron Man 2
Superman
```

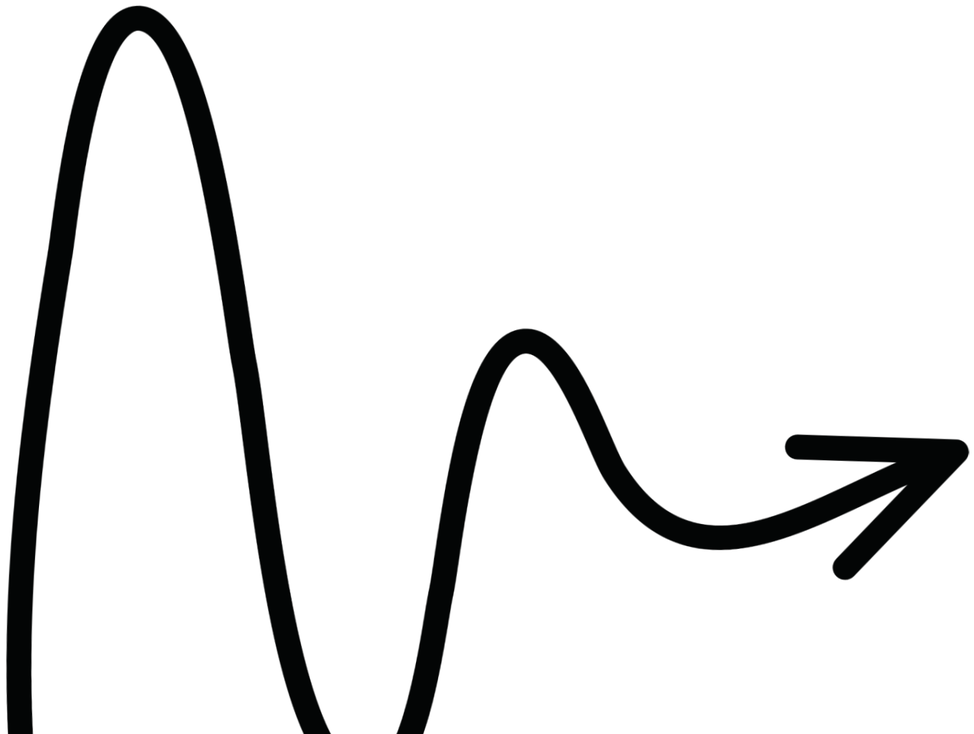
```
recommend('Tangled')
```

```
Out of Inferno
The Princess and the Frog
Home on the Range
Animals United
Toy Story 3
```

```
recommend('Toy Story')
```

```
Toy Story 2
Toy Story 3
The Adventures of Elmo in Grouchland
Should've Been Romeo
Harry Potter and the Philosopher's Stone
```

collaboratory-based filtering using knn



analyzing dataset

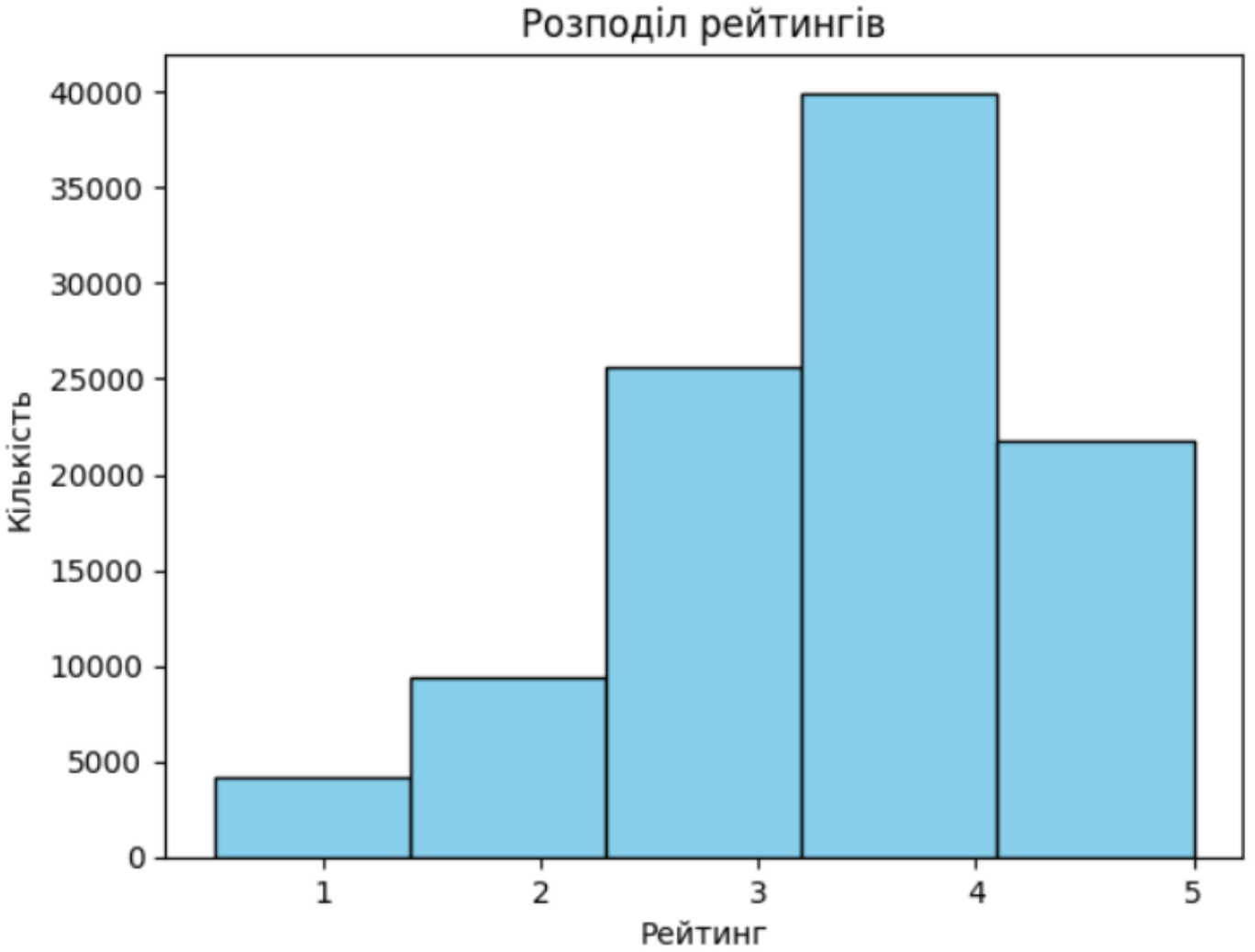
collaboratory-based filtering using knn

userId	movieId	rating	timestamp
1	1	4.0	964982703
1	3	4.0	964981247
1	6	4.0	964982224
1	47	5.0	964983815
1	50	5.0	964982931
1	70	3.0	964982400
1	101	5.0	964980868
1	110	4.0	964982176
1	151	5.0	964984041
1	157	5.0	964984100

movieId	title	genres
4051	Horrors of Spider Island (Ein Toter Hing im Netz) (1960)	Horror|Sci-Fi
4238	Along Came a Spider (2001)	Action|Crime|Mystery|Thriller
5349	Spider-Man (2002)	Action|Adventure|Sci-Fi|Thriller
5356	Giant Spider Invasion, The (1975)	Horror|Sci-Fi
6197	Spider (2002)	Drama|Mystery
6786	Kiss of the Spider Woman (1985)	Drama
8636	Spider-Man 2 (2004)	Action|Adventure|Sci-Fi|IMAX
52722	Spider-Man 3 (2007)	Action|Adventure|Sci-Fi|Thriller|IMAX
58105	Spiderwick Chronicles, The (2008)	Adventure|Children|Drama|Fantasy|IMAX

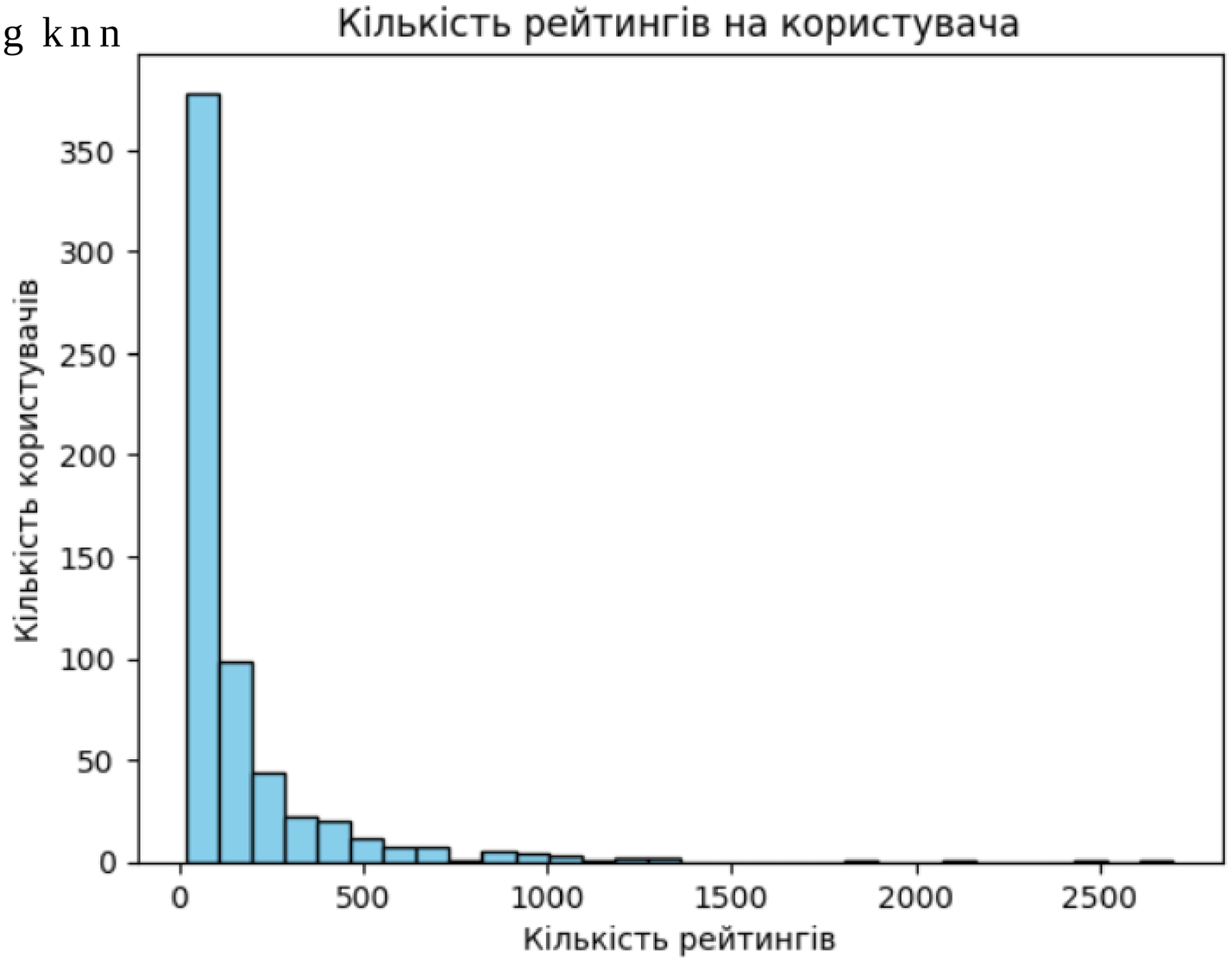
analyzing dataset

collaboratory-based filtering using knn



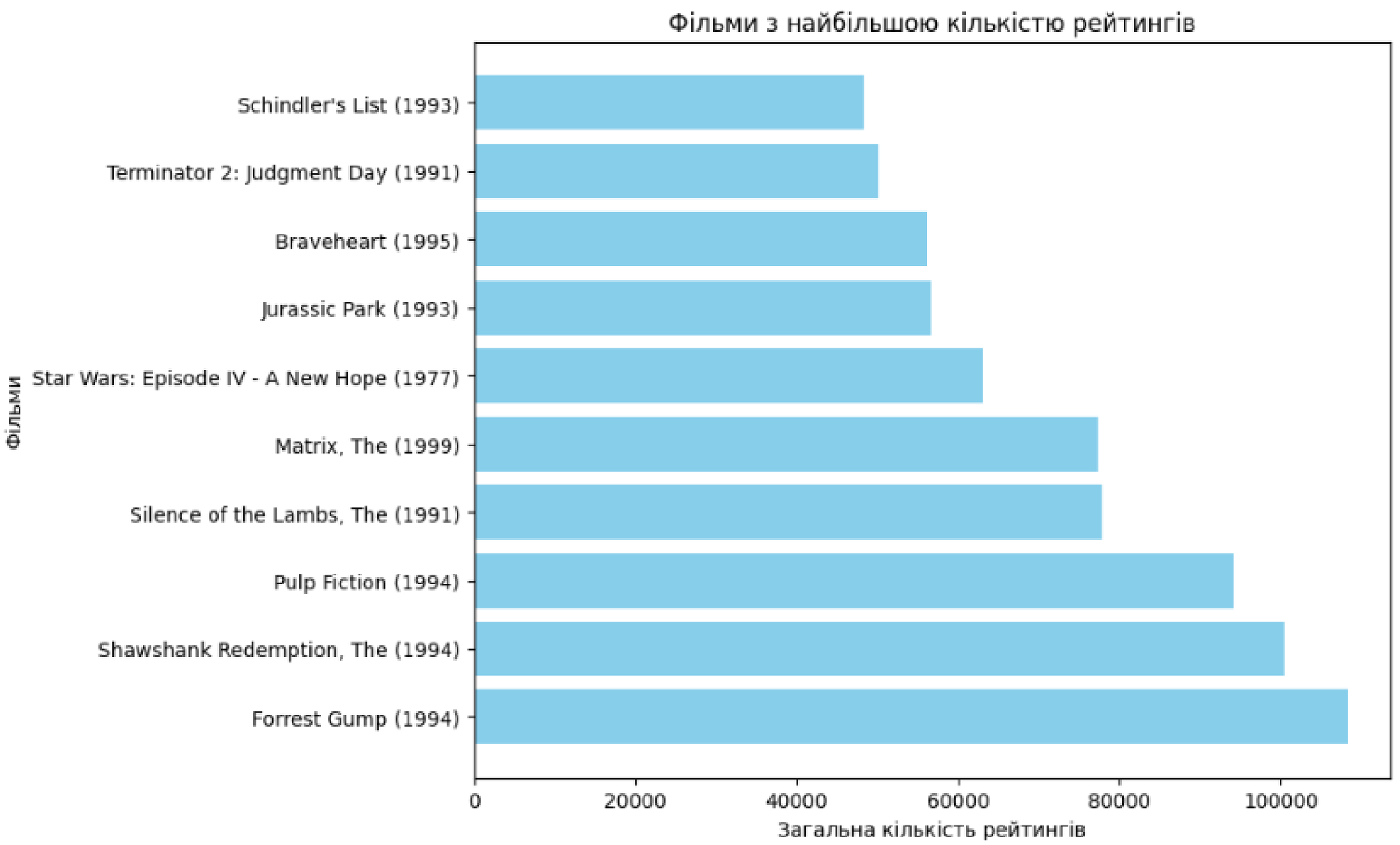
analyzing dataset

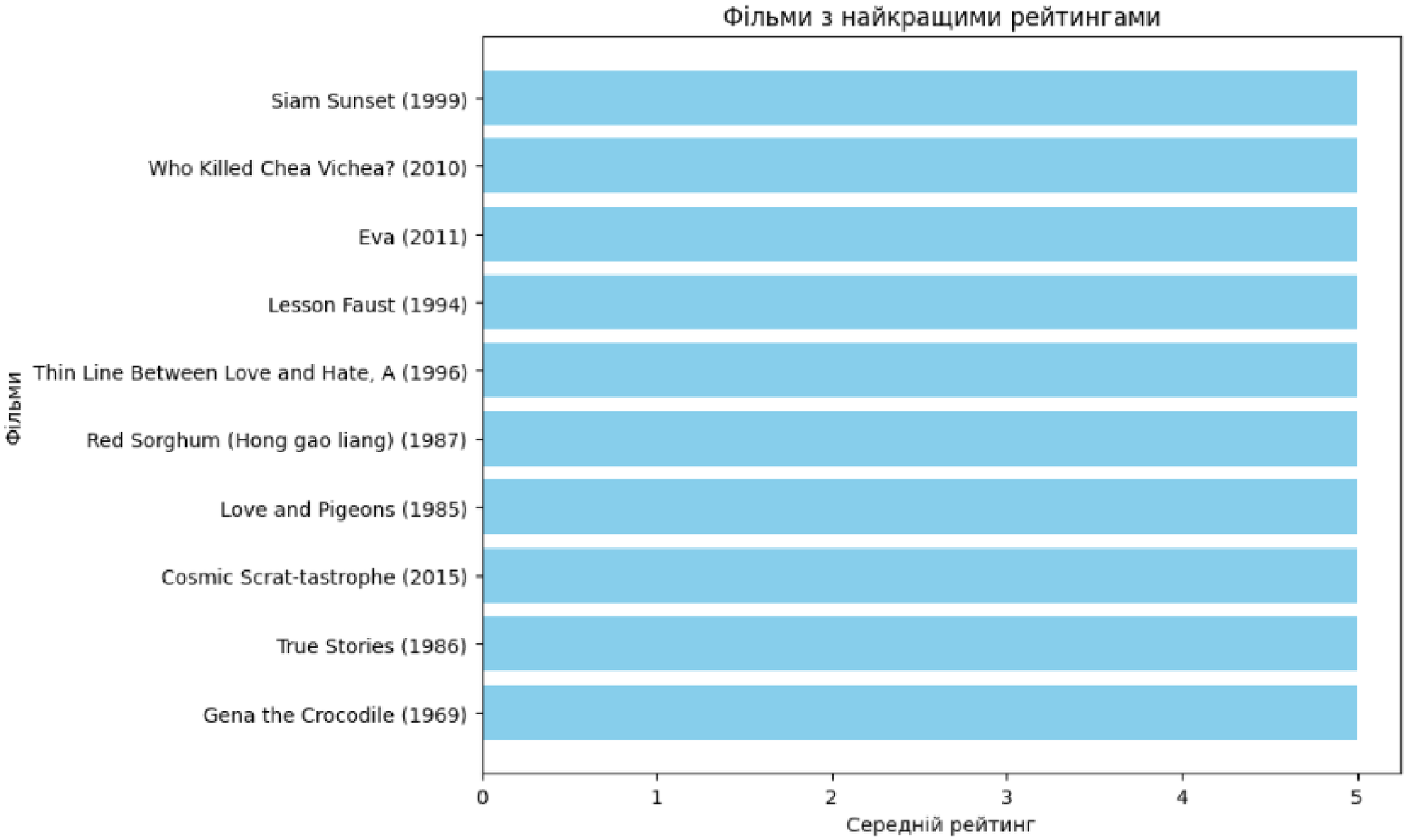
collaboratory-based filtering using knn



analyzing dataset

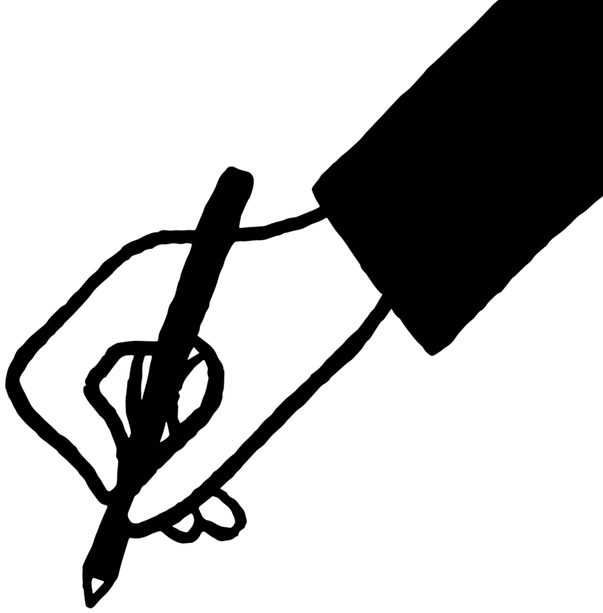
collaboratory-based filtering using knn


Фільми з найбільшою кількістю рейтингів
Schindler's List (1993)
Terminator 2: Judgment Day (1991)
Braveheart (1995)
Jurassic Park (1993)
Star Wars: Episode IV - A New Hope (1977)
Matrix, The (1999)
Silence of the Lambs, The (1991)
Pulp Fiction (1994)
Shawshank Redemption, The (1994)
Forrest Gump (1994)
Фільми
0
20000
40000
60000
80000
100000
Загальна кількість рейтингів


Фільми з найкращими рейтингами
Siam Sunset (1999)
Who Killed Chea Vichea? (2010)
Eva (2011)
Lesson Faust (1994)
Thin Line Between Love and Hate, A (1996)
Red Sorghum (Hong gao liang) (1987)
Love and Pigeons (1985)
Cosmic Scrat-tastrophe (2015)
True Stories (1986)
Gena the Crocodile (1969)
Фільми
0
1
2
3
4
5
Середній рейтинг


Фільми з найменшими рейтингами
Rust and Bone (De rouille et d'os) (2012)
The Emoji Movie (2017)
The Butterfly Effect 3: Revelations (2009)
Follow Me, Boys! (1966)
The Beast of Hollow Mountain (1956)
Dead of Night (1945)
Maria Bamford: The Special Special Special! (2012)
Haunted House 2, A (2014)
Midnight Chronicles (2009)
Baby Boy (2001)
Фільми
0.0
0.1
0.2
0.3
0.4
0.5
Середній рейтинг

Similarity metric



Euclidean distance

between (x_1, y_1) and (x_2, y_2) = $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

Cosine similarity

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|}$$

Ranking Evaluation

collaboratory-based filtering using knn

Metric

Hit Ratio @ 10:

$$HR = \frac{|U_{hit}^{L}|}{|U_{all}|}$$

where $|U_{hit}^{L}|$ is the number of users for which the correct answer is included in the top L recommendation list, $|U_{all}|$ is the total number of users in the test dataset.

Implementing knn

collaboratory-based filtering using knn algorithm

```
from scipy.sparse import csr_matrix

movie_features_df_matrix = csr_matrix(movie_features_df.values)

from sklearn.neighbors import NearestNeighbors



model_knn = NearestNeighbors(metric = 'euclidean', algorithm = 'brute')
model_knn.fit(movie_features_df_matrix)
```

```
NearestNeighbors
NearestNeighbors(algorithm='brute', metric='euclidean')
```

Евклідова відстань

```
Hit Ratio @ 10: 0.66
```

Implementing knn

collaboratory-based filtering using knn algorithm

```
from scipy.sparse import csr_matrix

movie_features_df_matrix = csr_matrix(movie_features_df.values)

from sklearn.neighbors import NearestNeighbors


model_knn = NearestNeighbors(metric = 'cosine', algorithm = 'brute')
model_knn.fit(movie_features_df_matrix)
```

```
NearestNeighbors
NearestNeighbors(algorithm='brute', metric='cosine')
```

Косинус подібності

Hit Ratio @ 10: 0.91

Recommendations for Toy Story (1995):

1: Toy Story 2 (1999),simularity is 57% with distance of 0.427398681640625:
2: Jurassic Park (1993),simularity is 56% with distance of 0.4343631863594055:
3: Independence Day (a.k.a. ID4) (1996),simularity is 56% with distance of 0.435738205909729:
4: Star Wars: Episode IV - A New Hope (1977),simularity is 55% with distance of 0.4426117539405823:
5: Forrest Gump (1994),simularity is 54% with distance of 0.45290398597717285:

Recommendations for Spider-Man 2 (2004):

1: Spider-Man (2002),simularity is 73% with distance of 0.26784181594848633:
2: X2: X-Men United (2003),simularity is 70% with distance of 0.29243040084838867:
3: Incredibles, The (2004),simularity is 64% with distance of 0.356227576732635 5:
4: X-Men: The Last Stand (2006),simularity is 62% with distance of 0.3754291534423828:
5: Pirates of the Caribbean: The Curse of the Black Pearl (2003),simularity is 60% with distance of 0.3916928172111511:

Те що подобається користувачу

Apollo 13 (1995)
Babe (1995)
Clerks (1994)
Client, The (1994)
Desperado (1995)
Рекомендації для користувача

1: Star Wars: Episode VI - Return of the Jedi (1983), with the distance 0.32925695180892944:
2: Back to the Future Part II (1989), with the distance 0.3434367775917053:
3: Star Wars: Episode V - The Empire Strikes Back (1980), with the distance 0.3456980586051941:
4: Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981), with the distance 0.35584235191345215
5: Back to the Future Part III (1990), with the distance 0.3617665767669678:

Movies Recomendation System Using Machine Learning

Type or select a movie to get recomendation

Harry Potter and the Half-Blood Prince

Show recommendation

Harry Potter an | Harry Potter an | Harry Potter an | Harry Potter an | Harry Potter an

Type or select a movie to get recomendation

Titanic

Show recommendation

The Notebook | Under the Same | Ghost Ship | The Bounty | Pirates of the

Movies Recomendation System Using Machine Learning

Type or select a movie to get recomendation

Spider-Man 3

Show recommendation

Spider-Man 2 | Spider-Man | The Amazing Spi | The Amazing Spi | Arachnophobia

Type or select a movie to get recomendation

Despicable Me 2

Show recommendation

Minions | Over the Hedge | Despicable Me | The Lego Movie | The Croods

Thank you!

for your attention