

Recommender Systems

msaidzengin



- Öneri sistemleri nedir?
- Öneri algoritmaları nelerdir?
- Nerelerde kullanılır?
- Kullanılabilecek kaynaklar nelerdir?

- 1) Bu ürüne bakanlar bunlara da baktı.
- 2) Bu ürüne benzer ürünler sunlar.

Son 20-30 yıldır, Youtube, Amazon, Netflix ve daha pek çok web servisinin yükselişiyle, öneri sistemleri yaşamımızda gittikçe daha fazla yer aldı.

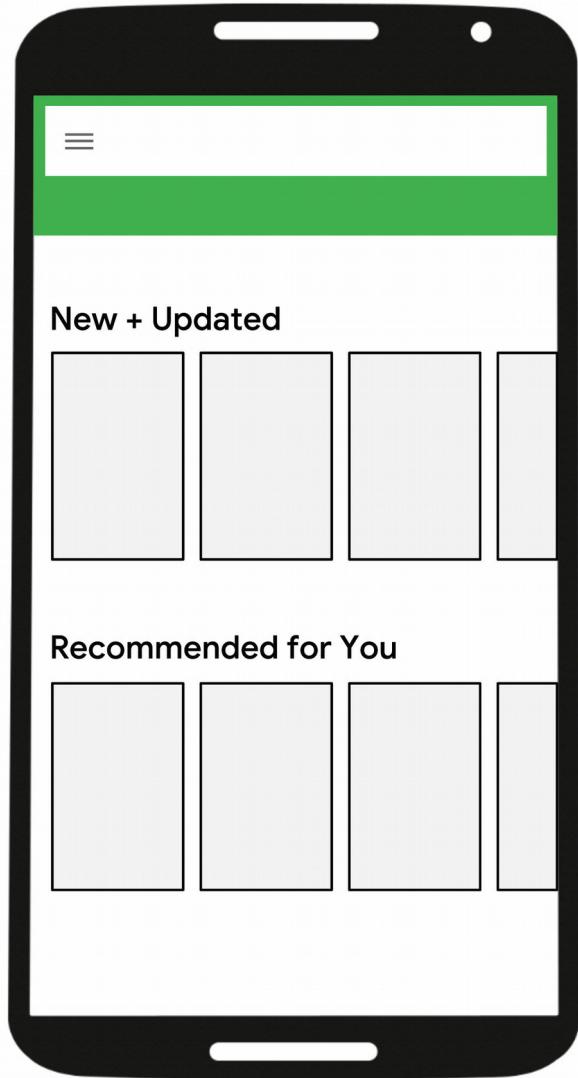
E-ticarette ürün önerileri, film-müzik önerileri, arkadaş önerileri, karşımıza çıkan reklamlar, arkadaş önerileri, search sonuçlarımız...

Netflix, Facebook, Pandora, Spotify, Amazon, Sahibinden, N11, Hepsiburada, Kitapyurdu, Google...

Özet olarak, öneri sistemleri, kullanıcılarla ilgili öğeleri önermeyi amaçlayan algoritmalardır.

(izlenecek filmler, okunacak metinler, satın alınacak ürünler veya sektörlerle bağlı olarak başka şeyler)

Tavsiye sistemleri bazı endüstrilerde kritik öneme sahiptir, çünkü verimli olduklarında büyük miktarda gelir elde edilebilir ve rakiplere bu sayede önemli ölçüde fark atılabilir.



40% of app installs on Google Play come from recommendations.
60% of watch time on YouTube comes from recommendations.



1 million dollar Netflix Prize

https://en.wikipedia.org/wiki/Netflix_Prize

Amazon.com, Inc.

NASDAQ: AMZN - Jun 2, 11:44 AM EDT

721.90 USD **↑2.46 (0.34%)**

1 day

5 day

1 month

3 month

1 year

5 year

max

800

600

400

200

0

2012

2013

2014

2015

2016

Open 720.97
High 722.10
Low 715.50

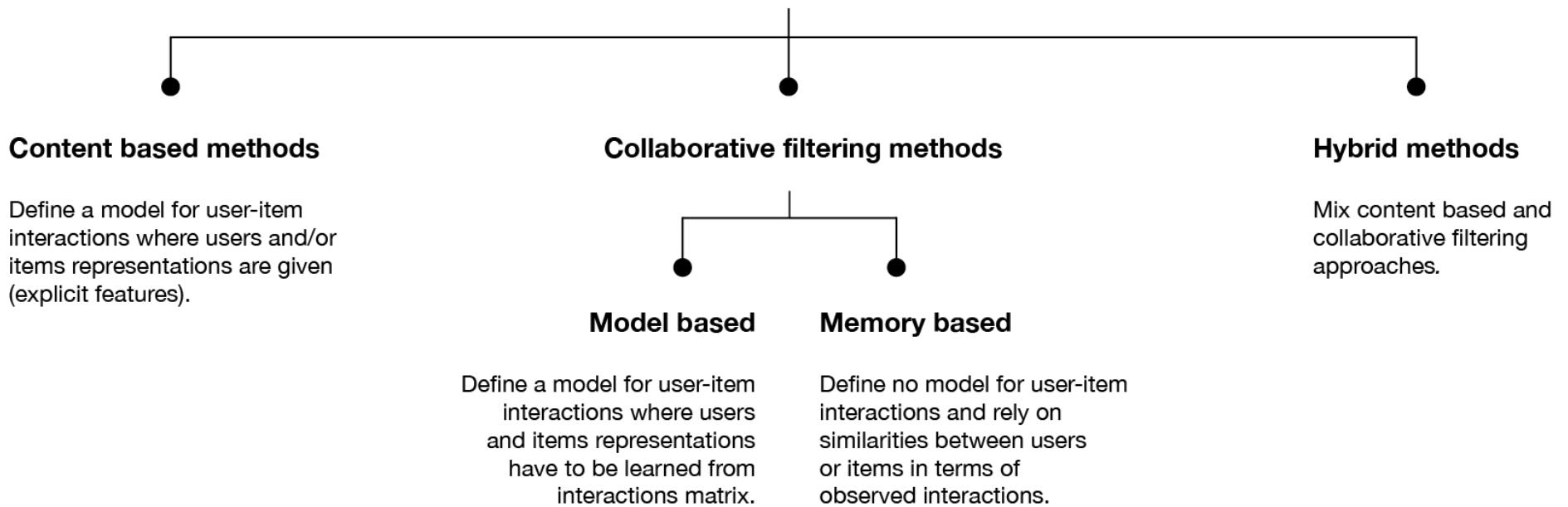
Mkt cap 339.03B
P/E ratio 297.17
Div yield -

Google Finance - Yahoo Finance - MSN Money

Disclaimer

Amazon öneri sistemleri sayesinde, ikinci mali çeyreğinde bir önceki yılın aynı dönemindeki 9.9 milyar dolar satışının, bir sonraki yıl %29'luk artışla 12.83 milyar dolar olduğunu bildirdi.

Recommender systems



- Best seller, most clicked, most viewed...
(IMDB Top 250)
- Content based
- Collaborative filtering
- Hybrit methods

Collaborative – Content Based Farkı

- Collaborative filtering başka kullanıcılarla bakarak öneriyor.
- Content based ise ürünün içeriğine bakarak öneriyor.

Type	Definition	Example
content-based filtering	Uses <i>similarity between items</i> to recommend items similar to what the user likes.	If user A watches two cute cat videos, then the system can recommend cute animal videos to that user.
collaborative filtering	Uses <i>similarities between queries and items simultaneously</i> to provide recommendations.	If user A is similar to user B, and user B likes video 1, then the system can recommend video 1 to user A (even if user A hasn't seen any videos similar to video 1).

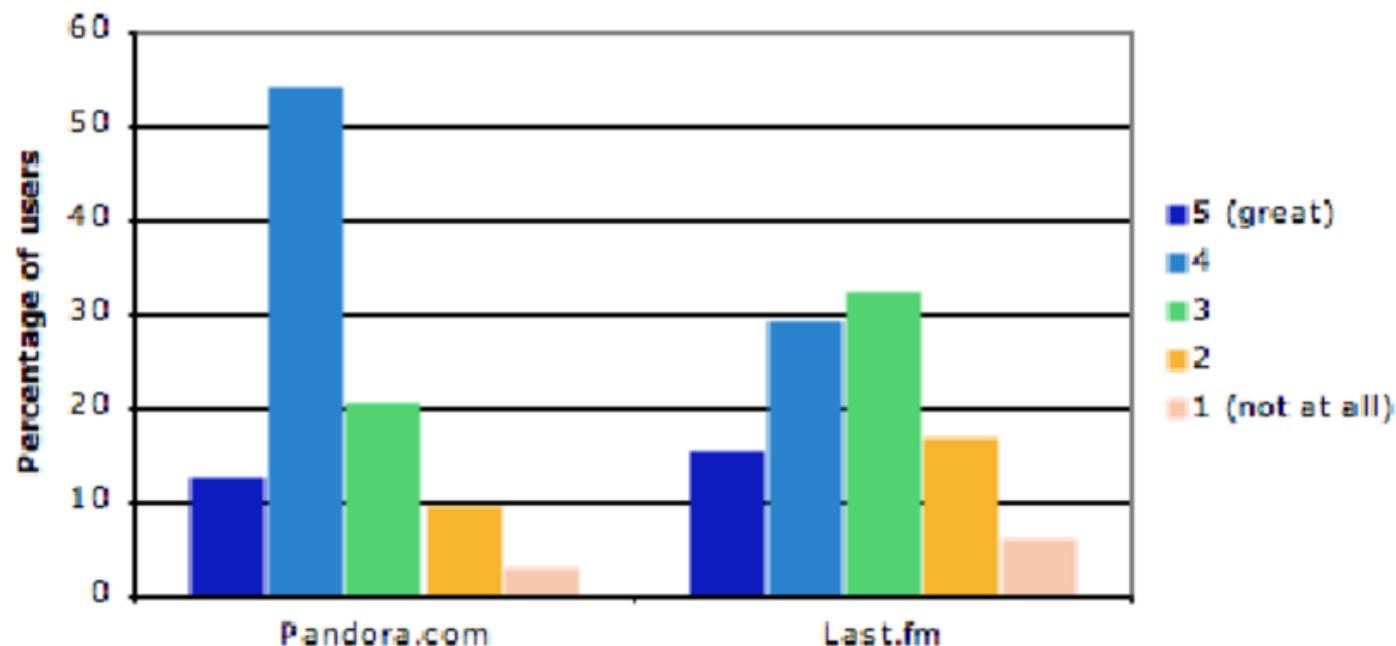
- 2007'de yazılan User Technology Adoption Issues in Recommender Systems isimli makalede iki müzik sitesi öneri sistemleri bakımından karşılaştırıldı.
- Pandora (content-based recommender)
- Last.fm (rating-based recommender (collaborative))
- Temel amaç, kullanıcıların öneri teknolojisini ilk kez benimsemelerini ve ilgili sistemlere yönelik öznel algılarını araştırmaktı.
- Sonuç olarak bu tür web sitelerinin popüler olma yolundaki ilk adımlarının, basit bir arayüz tasarıımı, kayıt olmanın kolaylığı ve önerilen öğelerin kalitesi (doğruluk, yenilik ve zevklilik) olduğu görüldü.

- Last.fm, kullanıcının düzenli olarak dinlediği müzikleri ve bireysel şarkıları kaydederek ve diğer kullanıcıların dinleme davranışlarıyla karşılaştırarak önerilen şarkıları oluşturur.
- Last.fm, kullanıcının kütüphanesinde görünmeyen parçaları çalar ve bu parçalar genellikle benzer ilgi alanlarına sahip diğer kullanıcılar tarafından çalınan parçalardır.
- Bu yaklaşım kullanıcıların davranışlarını artırdığından, collaborative filtering tekniğinin bir örneğidir.

- Pandora, bir şarkının veya sanatçının özelliklerini benzer özelliklere sahip müzikler olan bir öneri sistemi hazırlamak için kullanır. (Müzik Genom Projesi tarafından sağlanan 400 niteliğin alt kümesi)
- Kullanıcı geri bildirimine göre, öneri sonuçlarını iyileştirmek, bir kullanıcı bir şarkıyı "beğenmediğinde" belirli nitelikleri vurgulamak ve bir kullanıcı bir şarkıyı "sevdiğinde" diğer nitelikleri vurgulamak için kullanılır.
- Bu, content based bir yaklaşım örneğidir.

- Her tür sistemin güçlü ve zayıf yönleri vardır.
- Bu örnekte Last.fm'e bir kullanıcıya doğru önerilerde bulunmak için büyük miktarda bilgi gerekmektedir.
- Bu, soğuk başlangıç (cold start) probleminin bir örneğidir ve collaborative filtering sistemlerinde yaygındır.
- Pandora'nın öneri sisteminin başlatılması için çok az bilgiye ihtiyaç duymasına rağmen, kapsam bakımından çok daha sınırlıdır. (Örneğin sadece benzer şeyleri önerebilir.)

How enjoyable were the recommended songs?

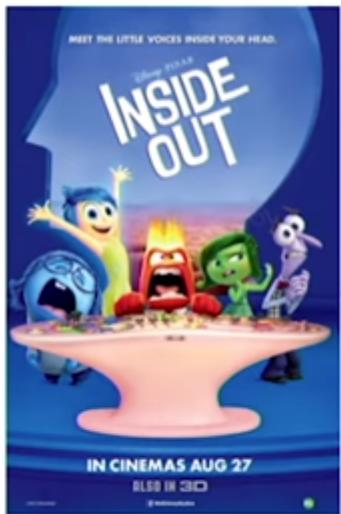


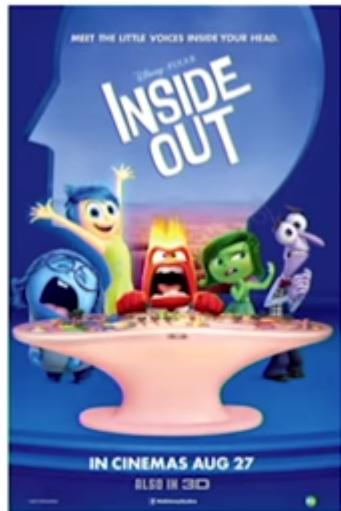
- Öneri sistemlerinin ML'den farkı kullanıcıya özel olması.
- Accuracy gibi başarımı ölçen değerler kesin olarak ölçülemiyor.
- En iyi öneriyi yapsan da kullanıcının o anki durumuna bağlı olabilir.
- Tüm kullanıcıların beğendiği, tıkladığı bilgilere bakarak öneri yapılabilir.
- Veya sadece senin geçmişine bakarak, ileride beğenebileceğin şeyleri önerebilir.

- Öneri sistemleri, kullanıcıların başka türlü bulamadıkları öğeleri keşfetmelerine yardımcı oldukları için arama algoritmalarına yararlı bir alternatiftir.
- Öneri sistemleri genellikle geleneksel olmayan verileri indeksleyen arama motorları kullanılarak uygulanır.

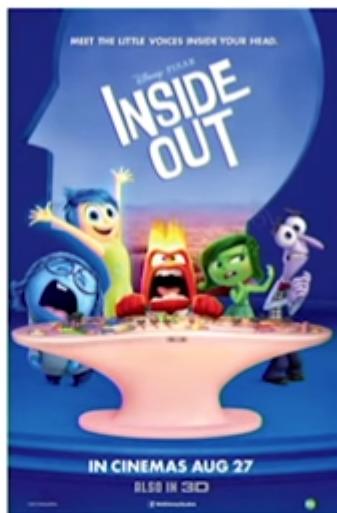
Content Based

- Ürünlerin feature benzerliğine göre çalışır.
- İki ürün arasındaki similarities'e göre öneri yapar.
- Kullanıcı bilgilerine ihtiyaç duymaz.
- Kullanıcının geçmiş derecelendirmelere ve diğer kullanıcıların tercihlerine göre bir öğeye vereceği notu veya tercihi tahmin etmeye çalışır.

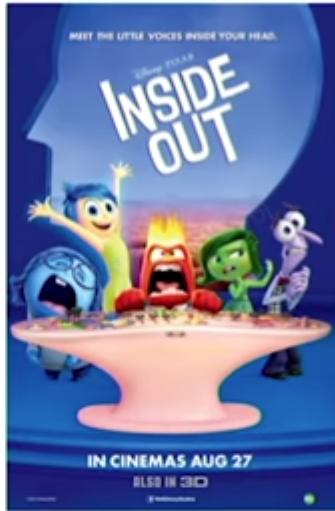




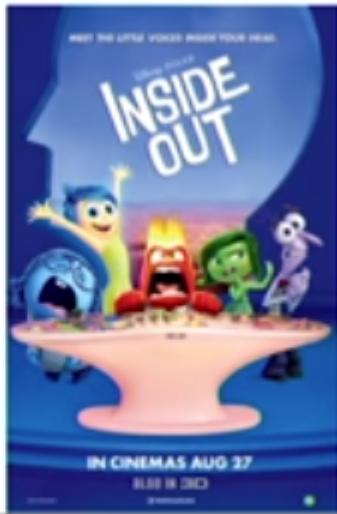
Animated	Yes	Yes	No	No
-----------------	-----	-----	----	----



Animated	Yes	Yes	No	No
Marvel	No	No	Yes	Yes



Animated	Yes	Yes	No	No
Marvel	No	No	Yes	Yes
Super Villain	No	Yes	Yes	Yes
Bechdel Test?	Pass	Fail	Pass	Fail
Parks and Rec Alum	Yes (Amy Poehler)	Yes (John Hamm)	No	Yes (Paul Rudd)



Animated	Yes	Yes	No	No
Marvel	No	No	Yes	Yes
Super Villain	No	Yes	Yes	Yes
Bechdel Test?	Pass	Fail	Pass	Fail
Parks and Rec Alum	Yes (Amy Poehler)	Yes (John Hamm)	No	Yes (Paul Rudd)



- Content based algoritma,
 - Daha önce izlenen filme göre yeni film öneriyor.
 - Satın alınan ürünne göre yeni ürün öneriyor.
 - Tıklanan kitaba göre yeni kitaplar öneriyor.
 - Görüntülenen bir CVE için yeni CVE'ler önerebiliyor.

- Bunu yaparken CVE'lerin feature benzerliklerine bakıyor. Bu benzerlikler True False (0-1) olabilir.
- Bir CVE'den diğer CVE'lere toplam benzerlik puanı tutulabilir ve bu puan sıralanıp en yüksek puan olanlar önerilebilir.
- True False değerler yerine virgülü sayılar ile benzerlik bulunup bu da puana katılabılır. Text2vec ile description benzerliği hesaba katılabılır.
- CVEörneğinde base score, description, product eşleşmeleri özellik olarak kullanılabilir. Hepsinin ayrı bir katsayısı olabilir.
- Her CVE için diğer tüm CVE'lere yakınlık hesaplanıp puanlar sıralanır. Top n CVE ise öneri olarak kullanılabilir.

- Content Based metodlarda tüm kullanıcıların verisine ihtiyaç yoktur. Sadece ürün benzerlikleri ve öneri yapılacak kullanıcının daha önce beğendiği-tıkladığı ürünler yeterlidir.
- Collaborative filteringde ise kullanıcı bilgileri gereklidir. Kullanıcıların beğenilerine, tıklamalarına göre çalışır.

- Collaborative filtering yönteminde ne kadar çok veri varsa o kadar başarılı öneriler yapılabilir.
- Netflix, Amazon, Google gibi devler en iyi öneri yapan sistemlere sahiptir. (Ürün, reklam, film önerisi açısından)
- Örneğin mükemmel çalışan bir arama motoru algoritması keşfedilse bile, kullanıcı sayısı 1 milyar olmadığı sürece Google'dan daha iyi çalışması imkansız.

Collaborative Filtering

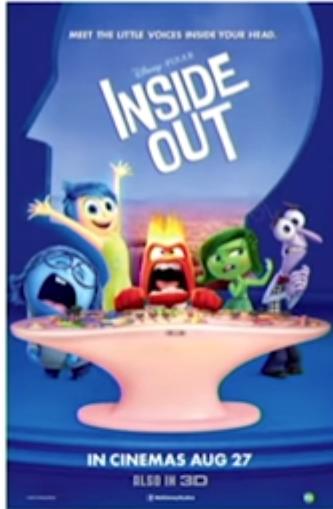
- Collaborative filtering yöntemlerini yöneten ana fikir, geçmiş kullanıcı-öge etkileşimlerinin benzer kullanıcıları veya benzer öğeleri tespit etmek ve tahminler yapmaktadır.
- Eğer bir kişi belirli bir eşyayı sevdiyse, ona benzer bir eşyayı da sevecektir.
- user - user
- user - item
- matrix factorization



Jason	Yes	Yes	Yes	Yes
Andi	No	Yes	No	Yes
Sarah	Yes	No	Yes	No
Sam	No	No	Yes	Yes



Jason	Yes	Yes	Yes	Yes
Andi	No	Yes	No	Yes
Sarah	Yes	No	Yes	No
Sam	No	No	Yes	Yes
Scaz	Yes	??	??	??



Jason	Yes	Yes	Yes	Yes
Andi	No	Yes	No	Yes
Sarah	Yes	No	Yes	No
Sam	No	No	Yes	Yes
Scaz	Yes	??	??	??

1
vote

2
votes

1
vote

- Günümüzde ise Amazon, Facebook, Netflix gibi tüm şirketler hybrit recommender system kullanıyor.
- Yani hem kullanıcı bilgileri, hem kullanıcı beğenileri, hem onlara benzer kullanıcıların beğenileri, hem geçmişte görüntülediği şeyler, görüntülediği şeylere benzer şeyler vs...

Content-based Filtering Advantages & Disadvantages

Advantages

- The model doesn't need any data about other users, since the recommendations are specific to this user. This makes it easier to scale to a large number of users.
- The model can capture the specific interests of a user, and can recommend niche items that very few other users are interested in.

Disadvantages

- Since the feature representation of the items are hand-engineered to some extent, this technique requires a lot of domain knowledge. Therefore, the model can only be as good as the hand-engineered features.
- The model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.

Collaborative Filtering Advantages & Disadvantages

Advantages

 No domain knowledge necessary

 Serendipity

The model can help users discover new interests.

 Great starting point

Disadvantages

 Cannot handle fresh items

 Hard to include side features for query/item

Lots of Data and Data Sources

User-generated

- “Community meta-data”
- e.g., tags, reviews



Interaction Data

- Listening logs/shared listening histories
- Feedback (“thumbs”)
- Purchases



Curated collections

- Playlists, radio channels
- CD album compilations



Referanslar

- <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>
- https://en.wikipedia.org/wiki/Recommender_system
- https://github.com/grahamjenson/list_of_recommender_systems
- <https://github.com/merveyapnaz/Tavsiye-Sistemleri-Recommender-System>
- <https://www.datacamp.com/community/tutorials/recommender-systems-python>
- https://github.com/rounakbanik/movies/blob/master/movies_recommender.ipynb
- <https://www.kaggle.com/rounakbanik/the-movies-dataset>
- <https://medium.com/datadriveninvestor/how-to-built-a-recommender-system-rs-616c988d64b2>
- <https://towardsdatascience.com/recommender-systems-in-practice-cef9033bb23a>
- <https://medium.com/recombee-blog/machine-learning-for-recommender-systems-part-1-algorithms-evaluation-and-cold-start-6f696683d0ed>
- <https://developers.google.com/machine-learning/recommendation>
- <https://ai.google/research/pubs/pub45530>
- <https://www.slideshare.net/FabienGouyon/music-recommendation-2018-116102609>
- https://en.wikipedia.org/wiki/Netflix_Prize
- <https://www.youtube.com/watch?v=Eeg1DEeWUjA>
- <https://www.kaggle.com/rounakbanik/movie-recommender-systems>