



# پیشنهادهای محصول فروشگاه آنلاین

پروژه درس هوش مصنوعی - دکتر اکبری  
دانشگاه امیرکبیر، دانشکده ریاضی و علوم کامپیوتر  
بهار 1401

# تیم پروژه

9712045

بهناز محمدی

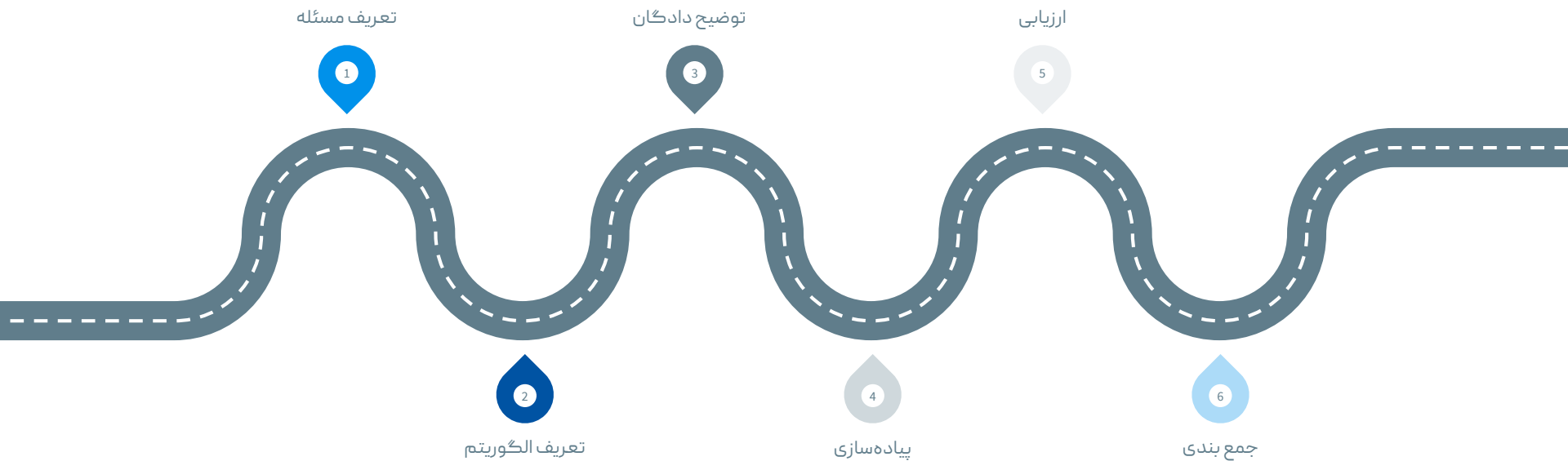
9713032 اشکان ودادی گرگری

9712039

زهرا کشاورز



# فهرست





## صورت مسئله

- یکی از چالش‌های امروزی فروشگاه‌های آنلاین دنیا، پیشنهاد محصول به خریداران خودشان است.
- روشی که ما برای حل این مسئله انتخاب کردیم، سیستم‌های توصیه‌گر (Recommender System) با الگوریتم فیلترینگ مشارکتی (Collaborative Filtering) است.
- برای پیشنهاد محصول بر اساس امتیازات خریداران، روش‌های متعددی وجود دارد که ما از میان این روش‌ها با توجه به مقایسه روش‌های گوناگون، از سیستم‌های توصیه‌گر استفاده می‌کنیم. سیستم‌های توصیه‌گر ابزارهایی هستند که به کاربر کمک می‌کنند تا گزینه مورد نظر خود را راحت‌تر پیدا کنند.



1.

# تعریف و آموزش الگوریتم

سیستم‌های توصیه‌گر



“

سیستم‌های توصیه‌گر ابزارهایی هستند که به کاربر کمک می‌کنند تا گزینه مورد نظر خود را راحت‌تر پیدا کنند (در این مسئله گزینه‌های ما همان محصولات ما هستند) و پیشنهادهایی به کاربر می‌دهد که احتمالاً مورد علاقه کاربر هستند.

## انواع سیستم‌های توصیه‌گر

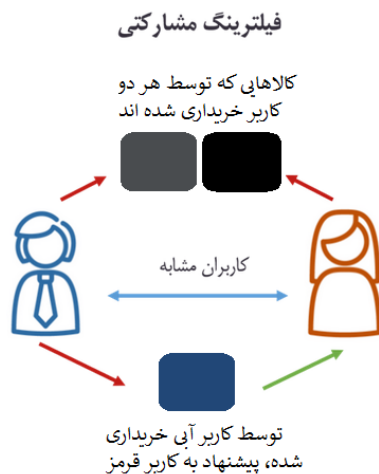
فیلترینگ محتوا محور (content-based filtering)

فیلترینگ مشارکتی (collaborative filtering)

مدل ترکیبی (hybrid)

ما برای حل این مسئله از **فیلترینگ مشارکتی** استفاده می‌کنیم.

## سیستم‌های توصیه‌گر – فیلترینگ مشارکتی



فیلترینگ مشارکتی، تشابه بین کاربر و گزینه را به طور همزمان در نظر می‌گیرد. مدل‌های فیلترینگ مشارکتی می‌توانند گزینه‌ای به کاربر الف پیشنهاد دهد که بر اساس تشابه‌های بین کاربر الف و کاربر ب است.

برای مثال کاربر الف وارد سایت خرید آنلاین می‌شود، مدل فیلترینگ مشارکتی باید کالاهای مشابه کالاهایی که کاربر الف در گذشته نگاه کرده است را پیشنهاد دهد و همچنین کالاهایی که کاربران، مشابه کاربر الف خریداری کرده است.



# با یک مثال توضیح بدهیم

پیدا کردن تشابه بین کاربران  
و محصولات



## سیستم‌های توصیه‌گر – فیلترینگ مشارکتی

فرض کنیم امتیازهایی که هر کاربر به هر کالا داده است را جمع‌آوری کرده و ماتریس کاربران و کالاها را ایجاد کرده ایم (جدول زیر). در کل چهار کاربر داریم و هفت کالا و هر کاربر امتیازی بین صفر تا ۵ به هر کالا که خریداری کرده است، داده است و درایه‌هایی که امتیاز ندارند، توسط کاربر خریداری نشده اند.

| فلش مموری | هارد اکسترنال | کارت حافظه | مانیتور | پرینتر | کیس کامپیوتر | خنک کننده کامپیوتر |     |
|-----------|---------------|------------|---------|--------|--------------|--------------------|-----|
|           |               | ۱          | ۵       |        |              | ۴                  | الف |
|           |               |            |         | ۴      | ۵            | ۵                  | ب   |
|           | ۴             | ۵          | ۲       |        |              |                    | ج   |
| ۳         |               |            |         |        | ۳            |                    | د   |

## سیستم‌های توصیه‌گر – فیلترینگ مشارکتی

حال برای پیشنهاد دادن کالاها به روش فیلترینگ مشارکتی، باید بتوان روشی دقیق برای پیدا کردن تشابه بین کاربران و کالاها پیدا کرد که در مدل‌های یادگیری ماشین نیز قابل استفاده باشد. یکی از مرسوم‌ترین روش‌های پیدا کردن تشابه، مشابهت کسینوسی (Cosine similarity) است.

برای پیدا کردن تشابه بین دو کاربر، سطر هر کاربر را به عنوان یک بردار در نظر می‌گیریم و امتیازهایی که وجود ندارند را به ناچار، صفر قرار می‌دهیم. برای مثال، بردار کاربر الف به شکل زیر است:

|     |   |   |   |   |   |   |   |
|-----|---|---|---|---|---|---|---|
| الف | ۴ | ۰ | ۰ | ۵ | ۱ | ۰ | ۰ |
|-----|---|---|---|---|---|---|---|

## سیستم‌های توصیه‌گر – فیلترینگ مشارکتی

### تابع تشابه

$$\text{Similarity}(A, B) = \cos(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|}$$

کاربران الف و ب کالاهایی که مرتبط با تجهیزات جانبی کامپیوتر هستند را دوست داشتند و کاربر ج کالاهایی که مرتبط با تجهیزات ذخیره سازی ، و تشابه بین کاربران الف و ب بیشتر است.

### تشابهات بین الف باب و ج

$$\text{Similarity}(A, B) = 0.38 > \text{Similarity}(A, C) = 0.32$$

# چالش‌های وراه حل‌ها

هزینه محاسباتی زیاد و  
استفاده از SVD



$$R = P\Sigma Q^T$$

A diagram illustrating the matrix decomposition  $R = P\Sigma Q^T$ . It consists of four colored rectangles: a blue rectangle labeled  $R$  with dimensions  $m \times n$  below it; an equals sign; a yellow rectangle labeled  $P$  with dimensions  $m \times k$  below it; an orange rectangle labeled  $\Sigma$  with dimensions  $k \times k$  below it and a diagonal line across it; and a green rectangle labeled  $Q^T$  with dimensions  $k \times n$  below it.



2.

# دادگان (Dataset)

آمازون و دانشگاه استنفورد

## دادگان (Dataset)

این مخزن داده شامل چندین دادگان است. برای این مسئله ما از مجموعه وسایل الکترونیکی (electronics) و فقط امتیازدهی شده‌اند را استفاده می‌کنیم. این دادگان فقط شامل نام کاربرها، کالاها، امتیازدهی و زمان ثبت امتیاز است. هر سطر متعلق به یک کاربر و یک کالا است که امتیاز کاربر و زمان خرید را به ما نشان می‌دهد.

|   | userId         | productId  | ratings | timestamp  |
|---|----------------|------------|---------|------------|
| 0 | AKM1MP6P0OYPR  | 0132793040 | 5.0     | 1365811200 |
| 1 | A2CX7LUOHB2NDG | 0321732944 | 5.0     | 1341100800 |
| 2 | A2NWSAGRHCP8N5 | 0439886341 | 1.0     | 1367193600 |
| 3 | A2WNBOD3WNDNKT | 0439886341 | 3.0     | 1374451200 |
| 4 | A1GI0U4ZRJA8WN | 0439886341 | 1.0     | 1334707200 |

# راه دانلود دادگان

گوگل سرچ کنید و دومین  
لینک را باز کنید!







3.

# پیاده سازی

6 بخش مهم دارد.

## 6 بخش به طور مجزا



### بارگزاری کتابخانه‌ها

باتوجه به مسئله که باید با داده‌های فراوانی در طی پیاده‌سازی کار کنیم، کتابخانه‌های مربوط به خواندن داده، کار با داده و ... را اضافه کردیم.



### خواندن دیتاست

مقدماتی‌ترین چالش در مسائلی که با داده‌های بزرگ کار می‌کنند، گرفتن و خواندن داده‌ها و ذخیره آن‌ها در یک ماتریس (آرایه) است.



### دقیق‌تر کردن داده‌ها

در مجموع هر کاربر چند دفعه رای داده است و روی افرادی که بیشتر رای داده‌اند، تمرکز کنیم.



### تقسیم داده‌ها

ساخت مجموعه train و test از روی دادگان با نسبت 70:30.



### فیلترینگ مشارکتی

روش اصلی ما پیاده‌سازی این بخش است.



### ارزیابی

برای ارزیابی از روش کمترین مربعات و اختلاف میان مقدار پیشنهاد شده و مقدار واقعی استفاده کردیم.



4.

# تحلیل خروجی ارزیابی

با کمک کمترین مربعات

## جمع بندی و نتیجه گیری

کافی است نزدیک بودن مقدار واقعی و تخمینی را نشان بدهیم.

در خروجی زیر ۵ مقدار اول را نشان می‌دهیم که نزدیک بودن مقدار واقعی و تخمینی را می‌بینید و در پایین جدول مقدار کمترین مربعات حاصل از تمام کاربرانی که انتخاب کردیم را مشاهده می‌کنید که بسیار به صفر نزدیک است و دلالت بر درستی الگوریتم و دقیق بودن دارد.

|            | Avg_REAL_ratings | Avg_PREDICATE_ratings | item_index |
|------------|------------------|-----------------------|------------|
| productId  |                  |                       |            |
| 1400599997 | 0.090909         | 0.088513              | 0          |
| B00000DM9M | 0.454545         | 0.449816              | 1          |
| B00000J061 | 0.454545         | 0.558292              | 2          |
| B00000J08C | 0.454545         | 0.449816              | 3          |
| B00000J0A2 | 0.363636         | 0.354053              | 4          |

RMSE = 0.05854



5.

## جمع بندی

با توجه به خروجی ارزیابی ...

## جمع بندی و نتیجه گیری

ما برای حل این پروژه از روش های گوناگونی سعی کردیم استفاده کنیم و از مسائل متفاوت جستجو و ماشین لرنینگ سعی کردیم کمک بگیریم که بهترین راه حل میان تمام راه حل های موجود استفاده از توصیه گر مشارکتی در میان منابع مختلف بیشتر از سایر منابع پیشنهاد شده بود.

به طور کلی برای مسائل پیشنهاد محصول، کالا، فیلم، آهنگ و ... الگوریتم های توصیه گر مشارکتی بسیار استفاده می شوند و با توجه به خروجی ارزیابی نیز می توان به این نتیجه رسید که این امر بدون دلیل نیست.



# خسته نباشید

@behnazmhm

| [behnazmhm78@gmail.com](mailto:behnazmhm78@gmail.com)

@ashkanvg

| [vedadi.ashkan@gmail.com](mailto:vedadi.ashkan@gmail.com)

@zahra\_kvr

| [zahrakeshavarz124@gmail.com](mailto:zahrakeshavarz124@gmail.com)