

این پروژه مربوط به طرح کارآموزی دیتاشیپ است.

لطفاً برای حمایت بیشتر از ما شبکه‌های اجتماعی ما را دنبال کنید.

- [Website](#)
- [Youtube](#)
- [Github](#)
- [Linkedin](#)
- Telegram Channel: @data\_hub\_ir
- Instagram Page: @data\_hub\_ir
- Telegram Group: @data\_jobs

---

## طراحی سیستم توصیه دهنده

این تمرین هم جنبه تحلیل و هم پیاده‌سازی دارد و نیازمند آشنایی با پایتون و یادگیری ماشین است. هدف تمرین هر چه بیشتر قدرت تجزیه و تحلیل و البته مهارت کدنویسی است.

توجه: توصیه می‌شود برای اجرا و پیاده‌سازی از [colab](#) استفاده کنید.  
ابتدا به آدرس زیر مراجعه کرده و داده‌ها را دانلود کنید.

<https://github.com/mohamad-dehghani/persian-movie-info>

ابتدا داده‌ها را فراخوانی کرده و سپس باهدف شناخت بهتر داده‌ها به سراغ `eda` بروید و سعی کنید به کمک نمودارهای مناسب دید خوبی نسبت به پراکندگی داده‌ها به دست آورید؛ نمودارهای مختلف مثل توزیع فیلم‌ها بر اساس سال تولید و امتیاز `imdb`.

سپس داده‌های نال بررسی کرده و حتماً روش مناسبی برای پرکردن آن‌ها استفاده کنید.

در ادامه وابستگی بین ویژگی‌های مختلف به کمک نمودارهایی مثل `heatmap` ترسیم شود.

در قسمت بعدی به بحث پیش‌پردازش داده‌ها پرداخته و برای هر پیش‌پردازی که انجام می‌دهید در حد کوتاه کامنت بنویسید. مثلاً ستون بازیگر، گاهی فارسی بوده و گاهی انگلیسی گه باید برای آن تمهیدی اندیشید.

هدف اصلی پروژه طراحی یک سیستم توصیه دهنده است. سپس به سراغ مهندسی ویژگی رفته و مهم‌ترین ویژگی‌ها را انتخاب کنید؛ باید تلاش کنید از حداکثر ویژگی‌ها برای پیاده‌سازی آن استفاده کنید.

نکته: استفاده از ستون description و استفاده از پردازش زبان طبیعی، نمره اضافی ویژه دارد.

در ادامه به سراغ یک الگوریتم دلخواه مثل SVD<sup>1</sup> رفته و سیستم توصیه‌گر را پیاده‌سازی کنید. حتماً باید دلیل انتخاب الگوریتم بیان شود.

نکته: استفاده از یادگیری عمیق و شبکه عصبی نمره اضافی دارد.

در نهایت به کمک معیارهای رایج، خروجی‌ها را به‌خوبی تحلیل کنید و راه‌هایی که برای بهبود بیشتر داده‌ها به ذهنتان می‌رسد را در قالب کامنت بنویسید.

برای قسمت نهایی، حتماً باید یک  $UI^2$  مناسب (مثلاً با Streamlit) و ترجیحاً شکل طراحی کنید به صورتی که ابتدا لیست ژانرها نمایش داده شود و امکان انتخاب یکی از آن‌ها باشد. سپس متناسب با ژانر انتخاب شده، لیست فیلم‌های آن ژانر نمایش داده شود و کاربر یکی از فیلم‌ها را از لیست انتخاب کند و در نهایت شبیه‌ترین فیلم‌ها به آن را نمایش دهد.

مثلاً کاربر از بین ژانرها، "درام" را انتخاب می‌کند سپس ما لیست فیلم‌های درام به او نمایش می‌دهیم. حالا از بین این لیست، یک فیلم توسط کاربر انتخاب می‌شود مثلاً فیلم "آپاندیس". در نهایت ما باید شبیه‌ترین فیلم به فیلم "آپاندیس" را نمایش دهیم.

- کدها ترجیحاً به زبان پایتون و به کمک sklearn باشد.
- مرتب و منظم نوشتن کدها و همچنین تکه‌تکه نوشتن امتیاز بالایی دارد.
- وجود کامنت‌های کامل و جامع توصیه می‌شود.
- مثل همیشه استفاده از کدهای آماده قابل قبول نیست.

---

<sup>1</sup> singular value decomposition

<sup>2</sup> user interface