**به نام خدا**



گزارشکار پروژه‌ی سیستم توصیه‌گر فیلم

جبر خطی کاربردی

استاد راهنما: دکتر پیمان ادیبی

اعضای گروه:

پوریا طلائی

محمدکاظم هرندی

زمستان 1402

فهرست مطالب

[مسئله 3](#_Toc156415350)

[مقدمه 3](#_Toc156415351)

[اطلاعات 3](#_Toc156415352)

[ادغام اطلاعات 4](#_Toc156415353)

[پیش پردازش داده‌ها 5](#_Toc156415354)

[تجزیه مقادیر منفرد 6](#_Toc156415355)

[نمونه اجرای برنامه 8](#_Toc156415356)

[منابع 9](#_Toc156415357)

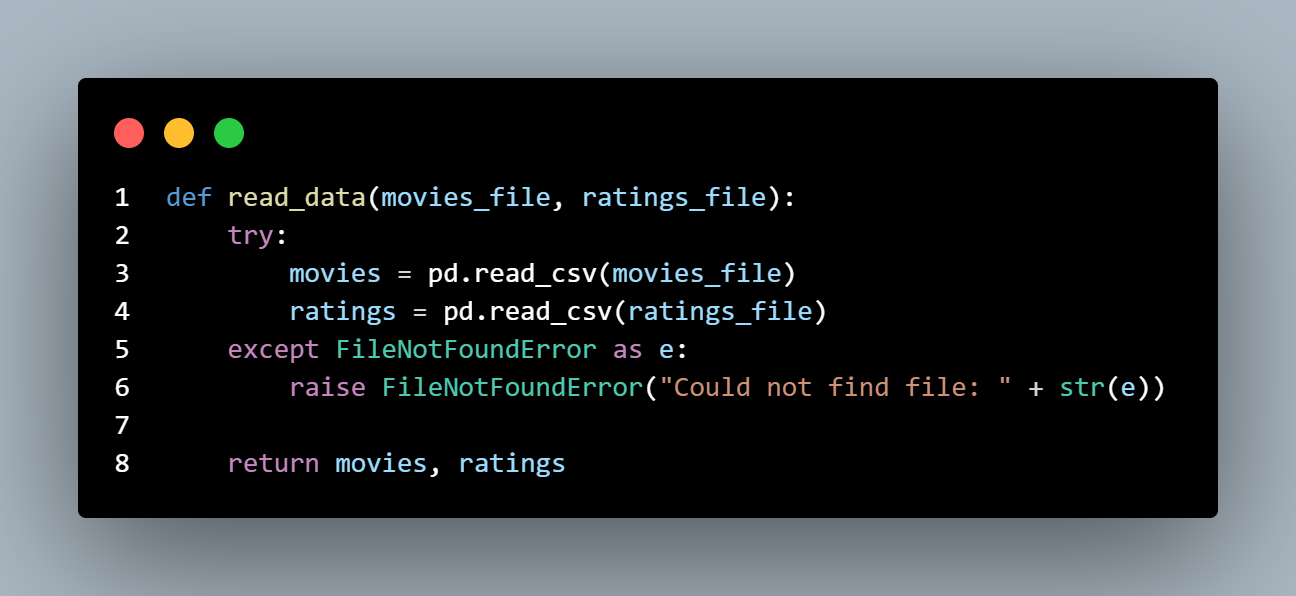
# مسئله

سامانه‌های توصیه‌گر امروزه در همه جا حاضر شده‌اند و پیشنهادهای شخصی‌سازی شده برای فیلم‌ها، موسیقی، کتاب‌ها، محصولات و موارد دیگر را به کاربران ارائه می‌کنند. این سیستم‌ها نقش مهمی در بهبود تجربه و تعامل کاربر و همچنین پیشبرد رشد کسب و کار دارند. در این پروژه قصد داریم تا با استفاده از تجزیه مقادیر منفرد[[1]](#footnote-1) یک سیستم توصیه‌گر فیلم ایجاد کنیم.

مقدمه  
در این پروژه، یک سیستم توصیه بر مبنای SVD برای پیشنهاد دادن فیلم‌ها به کاربران پیاده‌سازی شده است. این پروژه از زبان برنامه‌نویسی Python و کتابخانه‌های Pandas و NumPy برای مدیریت و پردازش داده‌ها استفاده می‌کند.

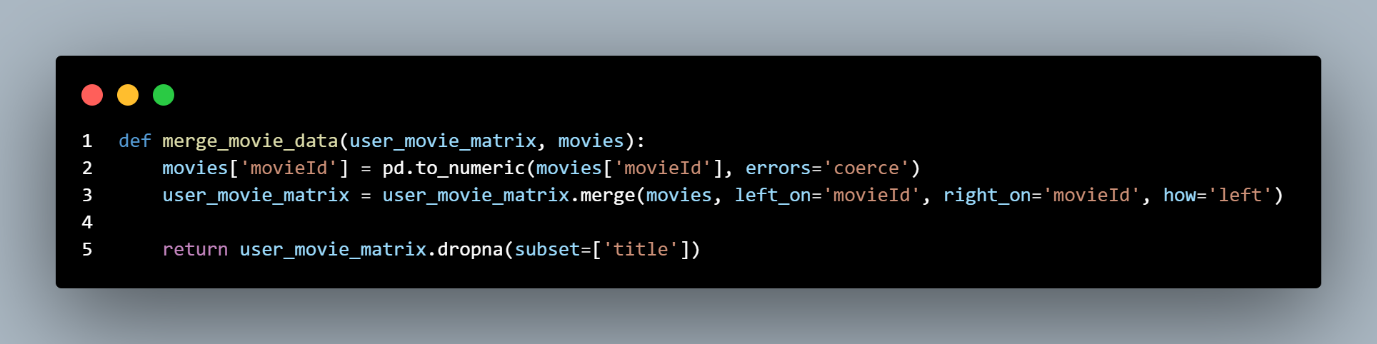
در ادامه به موارد جزئی تر و اطلاعات فنی دقیق‌تر در مورد پیاده‌سازی و عملکرد هر بخش از کد در چند مرحله اشاره خواهیم کرد.

# اطلاعات

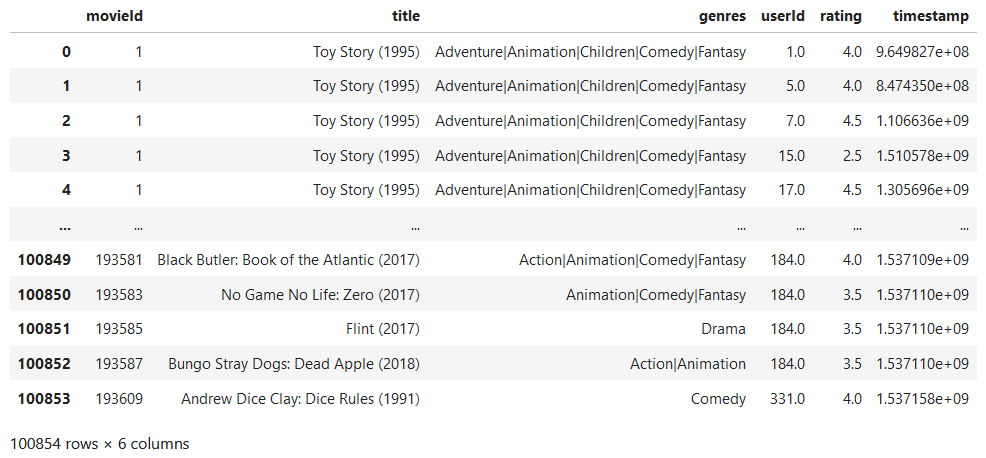
این تابع مسئول خواندن داده‌های مرتبط با فیلم‌ها و امتیازات کاربران از فایل‌های موردنظر است. درصورت وجود هر یک از فایل‌ها، از کتابخانه Pandas برای خواندن اطلاعات استفاده می‌شود. در صورتی که یک یا هر دوی این فایل‌ها یافت نشوند[[2]](#footnote-2)، یک استثناء ایجاد می‌شود و یک پیام خطا به کاربر اطلاع داده می‌شود که فایل مورد نظر پیدا نشد. در نهایت، اطلاعات مربوط به فیلم‌ها و امتیازات کاربران به عنوان خروجی تابع برگردانده می‌شوند.

‌ 1خواندن اطلاعات

# ادغام اطلاعات

در ابتدا، ستون 'movieId' در جدول فیلم‌ها به عدد تبدیل می‌شود. این عمل باعث مشکل در تبدیل اعداد نامعتبر به NaN می‌شود. سپس جدول فیلم‌ها را با ماتریس امتیازات کاربران ادغام می‌کند. ادغام بر اساس ستون 'movieId' انجام می‌شود و روش ادغام 'left' است، بنابراین تمام ردیف‌های موجود در ماتریس امتیازات کاربران حفظ می‌شوند و اطلاعات جدول فیلم‌ها به تعداد ممکن به ماتریس امتیازات کاربران افزوده می‌شوند. در نهایت، ردیف‌هایی که اطلاعات عنوان[[3]](#footnote-3) ندارند حذف می‌شوند و ماتریس نهایی با اطلاعات فیلم‌ها و امتیازات کاربران آماده می‌شود. این جدول به صورت زیر تبدیل می‌شود:

‌ 2ادغام اطلاعات

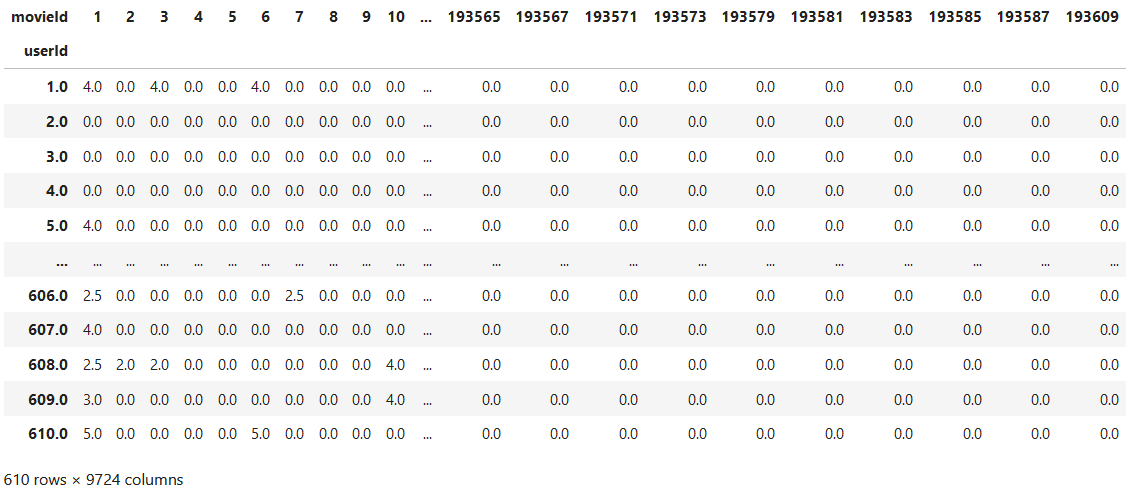


‌ 3ماتریس ادغام

# A screen shot of a computer code Description automatically generatedپیش پردازش داده‌ها

‌ 4پیش پردازش

ابتدا، داده‌های امتیازات کاربران به یک جدول خطی تبدیل می‌شوند. مقادیر خالی[[4]](#footnote-4) در جدول با مقدار صفر جایگزین می‌شوند. این به دلیل عدم ارتباط برخی از فیلم‌ها با برخی از کاربران می‌باشد. تمام ستون‌های جدول به نوع داده صحیح[[5]](#footnote-5) تبدیل می‌شوند. این تغییر باعث ایجاد ماتریس زیر برای ادامه کار می‌باشد.



‌ 5 ماتریس نهایی

# تجزیه مقادیر منفرد

‌ 6محاسبه‌ی مقادیر منفرد

در این تابع، تجزیه به ماتریسSVD بر روی ماتریس ورودی انجام می‌شود. در ادامه توضیح دقیقتری به مراحل و عملکرد هر قسمت از کد آمده است:

**محاسبه ماتریس کوواریانس:** این مرحله با ضرب ترانهاده ماتریس ورودی در خودش، ماتریس کوواریانس را ایجاد می‌کند. این ماتریس مهمی در مراحل بعدی تجزیه به ماتریس SVD است.

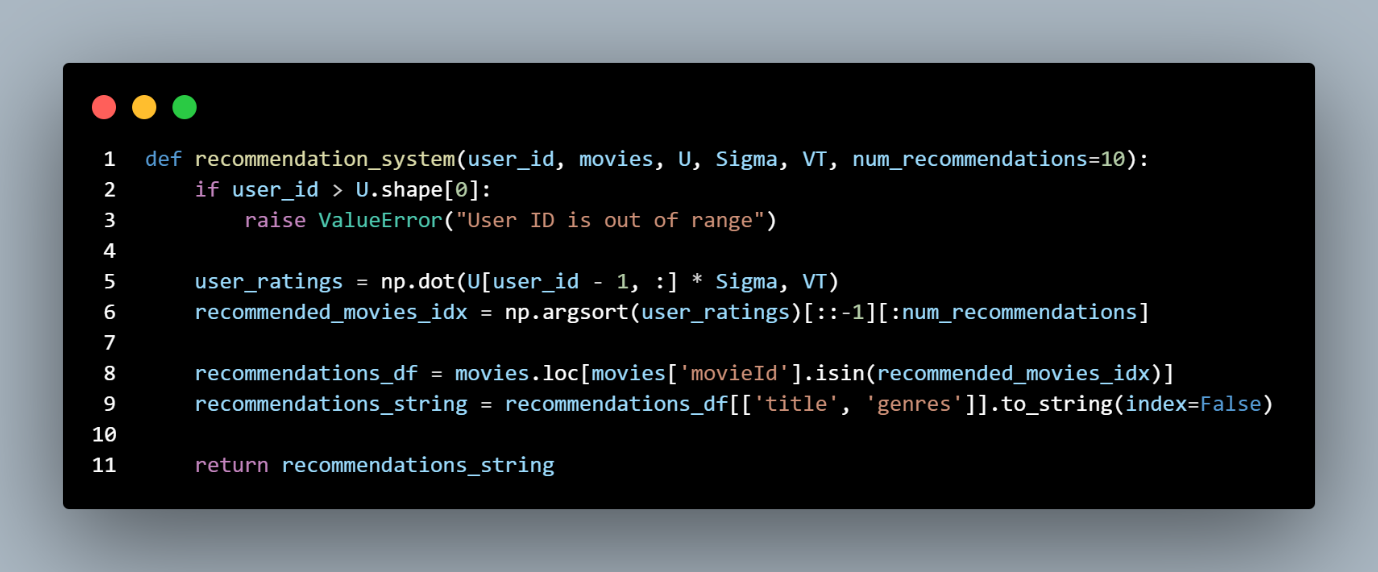
**محاسبه مقادیر و وکتورهای ویژه:** این بخش با استفاده از تابع eigh مقادیر و وکتورهای ویژه ماتریس کوواریانس را محاسبه می‌کند.

**مرتب‌سازی مقادیر و وکتورهای ویژه:** مقادیر و وکتورهای ویژه بر اساس مقدار مرتب شده مرتب می‌شوند.

**محاسبه مقادیر واحد تک‌تکرار:** مقادیر واحد تک‌تکرار با محاسبه مقادیر مثبت و جذر می‌شوند.

**محاسبه ماتریس‌های U و VT:** ماتریس‌های U و VT به عنوان ماتریس‌های تک‌تکرار حاصل از تجزیه SVD محاسبه می‌شوند.

تصحیح **مقادیر NaN و حذف مقادیر کوچک:** در این بخش، مقادیر NaN در ماتریس‌های U و VT به صفر تبدیل می‌شوند و مقادیر واحد تک‌تکرار کمتر از یک حد تعیین شده توسط toleranc حذف می‌شوند.

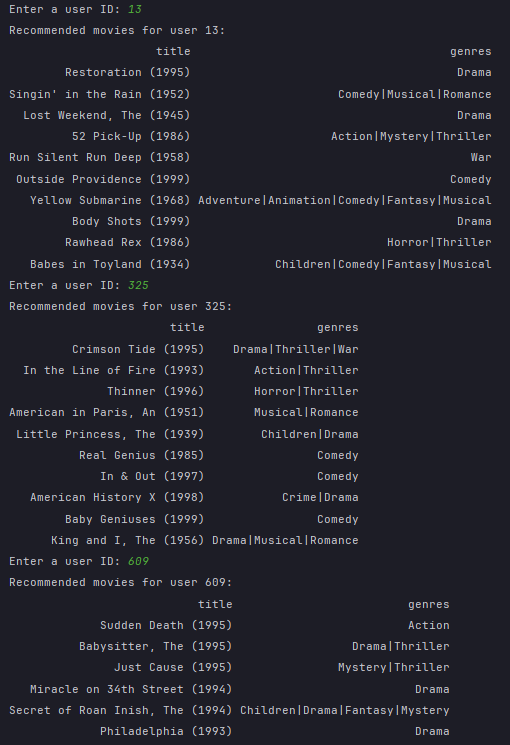
**خروجی:** ماتریس‌های U و VT همراه با مقادیر واحد تک‌تکرار به عنوان خروجی تابع ارائه می‌شوند. این ماتریس‌ها قابل استفاده در تحلیل داده‌ها و تولید پیش‌بینی‌ها بر اساس تجزیه SVD هستند.  
  
سیستم توصیه

‌ 7سیستم توصیه

با استفاده از ماتریس‌های U، Sigma و VT محاسبه امتیازهای تخمینی کاربر برای تمام فیلم‌ها انجام می‌شود. سپس امتیازهای کاربر بر اساس مقدار مرتب می‌شوند و تعداد مشخصی از فیلم‌های با بالاترین امتیازات برای پیشنهاد به کاربر انتخاب می‌شوند. درنهایت اطلاعات فیلم‌های پیشنهادی را از جدول اصلی فیلم‌ها با استفاده از شناسه فیلم‌ها استخراج می‌کند و اطلاعات فیلم‌های پیشنهادی را به یک رشته مناسب برای چاپ تبدیل می‌کند.

# نمونه اجرای برنامه

همانطور که می‌دانیم محاسبت برای بدست پیدا کردن مقادیرSVD کمی بالاست، پس باید برای انجام محاسبات کمی صبر کنیم. در ادامه با توجه به اطلاعاتی که داشتیم، به 610 کاربر موجود می‌توانیم فیلم هایی برای مشاهده توصیه کنیم. مواردی از آنها را در خروجی زیر می توانید مشاهده کنید.



‌ 8خروجی برنامه

# منابع

* machinelearningmastery.com/using-singular-value-decomposition-to-build-a-recommender-system
* analyticsindiamag.com/singular-value-decomposition-svd-application-recommender-system
* www.geeksforgeeks.org/singular-value-decomposition-svd/?ref=gcse
* www.geeksforgeeks.org/compute-the-factor-of-a-given-array-by-singular-value-decomposition-using-numpy/?ref=gcse
* black box
* bard.google.com
* chat.openai.com

1. Singular Value Decomposition [↑](#footnote-ref-1)
2. FileNotFoundError [↑](#footnote-ref-2)
3. title [↑](#footnote-ref-3)
4. NaN [↑](#footnote-ref-4)
5. integer [↑](#footnote-ref-5)