



کاوش داده گان انبوه

دستیاران آموزشی

مهدی صادقی (mahdisadeghi@ut.ac.ir)

شیوا پارسا راد (shiva.parsarad@ut.ac.ir)

علیرضا جعفری (Alirezajafari@ut.ac.ir)

ایمان برازنده (Barazandeh.iman@ut.ac.ir)

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده علوم و فنون

ترم اول سال تحصیلی - پاییز ۱۳۹۹

مجموعه تمرین هفتم : سیستم توصیه گر

مهلت تحویل این تمرین سه شنبه ۵ اسفند ساعت ۲۳:۵۹ می باشد.

(طراحی و پیاده سازی)

دیتاست تمرین شامل دو فایل است. فایل matrix.txt شامل یک ماتریس کاربر - آیتم R است. تعداد کاربرها ۹۹۸۵ نفر و آیتم ها شامل ۵۶۳ برنامه تلویزیونی است. اگر کاربر آیتم را پسندیده باشد $R_{ij} = 1$ و در غیر اینصورت صفر است. فایل items.txt شامل لیست برنامه ها با همان ترتیب ستون های R است.

می خواهیم به کاربر ۵۰۰ (سطر ۴۹۹) تعدادی آیتم برای مشاهده پیشنهاد بدهیم. فرض کنید او هیچکدام از ۱۰۰ آیتم اول را مشاهده نکرده و می خواهیم ده آیتم از آنها را به او پیشنهاد بدهیم. این کار را بر اساس رفتار او در مشاهده دیگر آیتم ها انجام می دهیم.

الف) ابتدا ده آیتم پر طرفدار از این ۱۰۰ آیتم را پیدا کرده و به او پیشنهاد بدهید.

یکی از موارد زیر (ب یا ج) را انجام دهید و نتایج را مقایسه و تحلیل کنید.

ب) با استفاده از روش user-user collaborative filtering ، ده آیتم به او پیشنهاد دهید.

ج) با استفاده از روش item-item collaborative filtering ، ده آیتم به او پیشنهاد دهید.



کاوش داده‌گان انبوه

دستیاران آموزشی

مهدی صادقی (mahdisadeghi@ut.ac.ir)

شیوا پارسا راد (shiva.parsarad@ut.ac.ir)

علیرضا جعفری (Alirezaiafari@ut.ac.ir)

ایمان برازنده (Barazandeh.iman@ut.ac.ir)

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده علوم و فنون

ترم اول سال تحصیلی - پاییز ۱۳۹۹

نکات تکمیلی

- از زبان برنامه سازی پایتون کنید.
- کدها با نامگذاری مناسب در پوشه SOURCES قرار گیرد.
- گزارش کامل تمرین را با ساختار مناسب، بدون ابهام و به زبان ساده نوشته و با فرمت PDF در پوشه دیگری قرار دهید. برای نگارش گزارش‌ها نیز امتیاز مثبت و منفی در نظر گرفته شده است، پس برای نگارش گزارش خود وقت مناسبی اختصاص دهید.
- پاسخ تمرین خود را به ایمیل barazandeh.iman@ut.ac.ir ارسال نمایید. هر دو پوشه را در یک فایل ZIP با نام MMDS_STID_FullName_HW# مثلا MMDS_830123451_ImanBarazandeh_HW1 قرار داده و در ایمیلی با همین عنوان نیز ارسال کنید.
- در صورت مشخص شدن هر نوع تقلب انفرادی یا مشترک نمره تمرین صفر در نظر گرفته می‌شود.
- به ازای هر روز تاخیر ۲۵٪ از امتیاز تمرین از دست خواهید داد.
- هر گونه ابهام را با دستیاران آموزشی مطرح کنید.