

فاز دوم پروژه

درس مبانی داده کاوی

نیمسال دوم تحصیلی ۱۴۰۲

بخش اول - استخراج الگوهای مکرر (Extracting Frequent Pattern)

در این فاز دانشجویان می بایست حداقل ۷ مورد از الگوهای مکرر موجود در دیتاست `clean` حاصل از فاز قبلی را استخراج نمایند. استفاده از کتابخانه های مربوط به الگوهای مکرر مجاز می باشد. (می توانید از کتابخانه `mlxtend` استفاده کنید)

بخش دوم - Clustering & Classification

در این بخش می خواهیم با استفاده از `Clustering` و `Classification` به کشف الگوهای پنهان دیتاست بپردازیم. شامل:

خوشه بندی

۱. الگوریتم های خوشه بندی

- از این بخش به انتخاب خودتان ۲ الگوریتم خوشه بندی (`K-means`, `K-median`, الگوریتم های سلسله مراتبی و ...) را بر روی داده ها اجرا کنید تا بخش های مختلف دیتاست را خوشه بندی کنید.
- الگوهایی که در خوشه ها شناخته شوند را مورد بررسی قرار دهید.

۲. تصویرسازی الگوهای خوشه بندی

- از نمودار ها و تصویر سازی ها برای نمایش الگوهای خوشه بندی بهره ببرید.

دسته بندی

۱. انتخاب ویژگی ها: ویژگی های مربوطه را برای طبقه بندی انتخاب کنید. ویژگی هایی که بیشتر در پیش بینی **Rating** نقش دارند باید انتخاب شوند.
۲. انتخاب مدل: با حداقل ۳ مدل طبقه بندی مختلف (به عنوان مثال Naive Bayes، Random Forest، Decision Tree و SVM) آزمایش کنید تا مدلی را پیدا کنید که بهترین عملکرد را در مجموعه داده شما دارد.
۳. آموزش مدل: مدل انتخاب شده را با استفاده از مجموعه آموزشی از داده های از پیش پردازش شده آموزش دهید.
۴. ارزیابی مدل: عملکرد مدل را در مجموعه تست با استفاده از معیارهایی مانند Accuracy، Precision، Recall و F1-Score ارزیابی کنید.

در زمان انجام این فاز به نکات زیر توجه فرمایید:

- در بخش الگوهای مکرر، کاهش min_support برای رسیدن به الگوها حداقلی باشد.
- در بخش خوشه بندی با روش های مختلف مناسب بودن پارامترهای مختلف را بررسی کنید. (مانند متد Elbow برای تعداد K)
- برای آموزش مدل در بخش دسته بندی، امکان استفاده از ستون هایی که برآمده از ستون Rating می باشند (بطور مثال شکل نرمال شده مقادیر Rating) را ندارید.
- در بخش دسته بندی نتایج مدل ها را به طور کامل با یافتن بهترین پارامترها تحلیل کنید.
- به دلخواه خود می توانید بخش دسته بندی را به رگرسیون تغییر دهید.
- برای تمامی بخش های این فاز، نوشتن گزارش و تحلیل خروجی ها الزامی و بخشی از نمره این فاز خواهد بود.

موفق باشید