

فاز دوم پروژه

درس مبانی داده کاوی

نیمسال دوم تحصیلی 1402

بخش اول - استخراج الگو های مکرر (Extracting Frequent Pattern)

در این فاز دانشجویان می بایست حداقل 6 مورد از الگو های مکرر موجود در دیتاست clean حاصل از فاز قبلی را استخراج نمایند. استفاده از کتابخانه های مربوط به الگو های مکرر مجاز می باشد. (می توانید از کتابخانه mlxtend استفاده کنید)

بخش دوم - Clustering & Classification

در این فاز می خواهیم با استفاده از Clustering و Classification به کشف الگو های پنهان دیتاست بپردازیم. در ادامه باید از خروجی مرحله قبل، بخش های زیر انجام شود.

خوشه بندی

1. الگوریتم های خوشه بندی

- از بخش به انتخاب خودتان 2 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.

2. تصویرسازی الگوهای خوشه بندی

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

دسته بندی

1. انتخاب ویژگی ها: ویژگی های مربوطه را برای طبقه بندی انتخاب کنید. ویژگی هایی که بیشتر در پیش بینی خطر حمله قلبی نقش دارند باید انتخاب شوند.
2. انتخاب مدل: با مدل های طبقه بندی مختلف (به عنوان مثال، درختان تصمیم، جنگل تصادفی، رگرسیون لجستیک، SVM) آزمایش کنید تا مدلی را پیدا کنید که بهترین عملکرد را در مجموعه داده شما دارد.
3. آموزش مدل: مدل انتخاب شده را با استفاده از مجموعه آموزشی از داده های از پیش پردازش شده آموزش دهید.
4. ارزیابی مدل: عملکرد مدل را در مجموعه تست با استفاده از معیارهایی مانند accuracy، precision، recall و F1 measure ارزیابی کنید.

موفق باشید