

پروژه پایانی درس داده کاوی

برای انجام پروژه ابتدا باید پایگاه داده‌ای که ضمیمه‌ی این پیام شده است را دانلود کنید. این مجموعه دادگان دارای ۱۴ صفت (ستون) و ۴۸۸۴۲ سطر می‌باشد. شما باید بتوانید با اجرای الگوریتم‌های کلسیفیکیشن مختلفی که در این درس می‌آموزید بتوانید با بهترین دقت پیشبینی کنید که آیا یک شخص درآمدی بیش از ۵۰ هزار دلار دارد یا خیر. نتایج حاصل از این کلسیفایرها را در جداولی به شکل زیر گزارش کنید:

(Predicted) TRUE	(Predicted) FALSE	
TP	FN	(Actual) TRUE
FP	TN	(Actual) FALSE

همچنین برای هر روش، در چند خط تحلیلی از نتایج به دست آورده را در گزارش خود بیاورید. این تحلیل باید نشان دهد که به خوبی با مفاهیم این کلسیفایرها آشنا شده‌اید. معیار دقت را نیز گزارش کنید. اسم این فایل را شماره دانشجویی خود گذاشته و در قالب PDF به همراه موارد زیر ارسال کنید. مهلت تحویل پروژه تا روز سه‌شنبه ۳۰ آذر (حداکثر تا ساعت ۲۳:۵۹) می‌باشد. در صورت تاخیر تا دو روز، ۲۵٪ نمره به عنوان جریمه کسر می‌شود و پس از دو روز نمره‌ای در نظر گرفته نمی‌شود.

دانشجویان ریاضی و آمار :

برای هر روش با تنظیم پارامترها بهترین نتیجه را گزارش کنید و در گزارش خود تحلیلی از تاثیر پارامترها بر نتایج کلسیفایرهای خود بیاورید. پس از ساخت فرایند، انرا با نام دانشجویی خود ذخیره کرده و می‌فرستید. ترجیحا برای هر فرایند اسکرین شات بگیرید تا در صورت باز نشدن آن از عکس استفاده شود.

دانشجویان علوم کامپیوتر :

الگوریتم‌های مختلف را پیاده‌سازی کنید و سعی کنید تا با بهترین الگوریتم بهترین نتیجه را به دست آورید. در گزارش، الگوریتم خود را توضیح داده و تحلیلی از تاثیر پارامترها بر نتایج کلسیفایرهای خود بیاورید. برای کدینگ پایتون فقط و فقط امکان استفاده از پکیج‌های هسته‌ی SciPy را دارید (NumPy ، SciPy library ، IPython ، Matplotlib ، Sympy و pandas) و به هیچ عنوان نباید از Scikit و علی‌الخصوص Scikit learn استفاده کنید. کدهای خود را در پوشه‌ای با نام شماره دانشجویی خود ذخیره کنید و به همراه PDF ارسال کنید.

موفق باشید

ashkanalinejad@ut.ac.ir