

پروژه پایانی درس داده کاوی

ترم دوم سال تحصیلی ۱۴۰۱-۱۴۰۰

مهلت تحویل: ۳۱ تیر ۱۴۰۱

دادگان پروژه

دادگانی که در این پروژه در نظر گرفته شده است دادگان مربوط به شناسایی انواع برگ های درخت انگور است. علاوه بر میوه درخت انگور، برگ های درخت انگور هم قابل استفاده است و بر اساس نوع آن قیمت و کاربرد آن تعیین میشود. جداسازی این برگ ها در مقیاس بزرگ به صورت دستی امکان پذیر نیست. بنابراین سعی میکنیم با دادگانی که داریم عمل دسته بندی برگ ها را به صورت خودکار انجام دهیم. میتوانیم این دادگان را از این [لینک](#) دانلود کنید.

شرح پروژه

- همانند تمرین هایی که انجام دادید، داده ها را با نسبت ۸۰ به ۲۰ به داده های آموزشی و تست تقسیم کنید.
- با استفاده از data augmentation میتوانید تعداد تصاویر را بیشتر کنید. دقت داشته باشید که داده های تست نباید از داده های augment شده باشد. لذا ابتدا داده ها را به دو بخش تست و آموزش تبدیل کنید و قسمت آموزشی را augment کنید. بعد از اعمال augmentation از داده های آموزشی ۲۰ درصد داده ها را به عنوان داده های validation استفاده کنید.
- با استفاده از معماری های مختلف شبکه عصبی CNN (استفاده از keras یا pytorch) عمل دسته بندی را انجام دهید.
- از مدل های pretrained موجود در اینترنت و پکیج ها استفاده کنید.
- از شبکه های auto encoder برای کاربرد هایی نظیر denoising یا dimensionality reduction استفاده کنید و نتیجه را شرح دهید.
- با استفاده از ۱۰ بار اجرا به ازای seed های تصادفی و مختلف ، میانگین نتایج را ذکر کنید.
- گزارش خود از آزمایشات متنوع را در پایان گزارش به صورت جدول و confusion matrix و نمودار نشان دهید.
- از 10 fold cross validation استفاده کنید و در هر مورد دقت را گزارش کنید. داده های train و validation را باهم ادغام کنید و مجموعه داده های آموزشی جدید را برای این قسمت استفاده کنید. دقت داشته باشید از داده های تست نباید در فرآیند cross validation استفاده کنید.
- نتایج خود را با نتایج مقالات مربوطه مقایسه کنید.

بخش امتیازی*

- از تمامی روش های موجود در درس مثل روش های دسته بندی ، کلاسترینگ و ... به منظور بالا بردن دقت استفاده کنید. از روش های ترکیبی نیز میتوانید استفاده کنید.
- در این بخش گرفتن دقت خوب مورد نیاز است لذا مختار هستید از هر روش ابتکاری استفاده کنید.
- اگر بتوانید دقت بالاتری از دقت مقاله در این قسمت کسب کنید، نمره اضافه به شما تعلق میگیرد.
- به شرط کسب نتیجه بهتر می توانید فایل گزارش را به زبان انگلیسی بنویسید.

نحوه ارسال فایل های پروژه:

- کد های خود را به همراه فایل مربوط به توضیحات pdf به صورت یک فایل فشرده در سامانه ایلرن ثبت کنید.
- ترجیحا فایل کد ارسالی به فرمت ipynb باشد.
- نام فایلی که آپلود می کنید شامل نام خانوادگی و شماره دانشجویی باشد . مثال :

DM – FinalProject - your name - your student id

نکات و پیشنهادات مربوط به انجام پروژه:

- برای انجام قسمت مربوط به نوشتن کد سعی کنید از Google Colab استفاده کنید و لینک مربوطه را در کنار فایل پروژه ارسال کنید.
- استفاده از پکیج ها و کتابخانه های آماده بلامانع است.
- برای خواندن داده ها میتوانید از پکیج opencv استفاده نمایید.

لینک های مربوطه:

<https://www.sciencedirect.com/science/article/abs/pii/S0263224121013142?via%3Dihub>

موفق باشید.