



درس داده کاوی

پروژه پایانی

01-02

پروژه شامل سه بخش می باشد: ابتدا باید روی داده های خود pre processing انجام دهید زیرا دیتا هایی موجود می باشند که مورد استفاده شما نیست. در قدم بعدی باید دیتا ها را طبقه بندی کنید به سه کلاس Normal, Benign, Malignant تقسیم کنید. در نهایت مدل خود را train کنید.

برای train کردن مدل حتما باید از فریم ورک Tensorflow استفاده کنید. دیتا های شما باید شامل train, validation, test باشد. که سهم تعداد سمپل های train باید بیشتر از بقیه باشند. در فرایند train کردن دیتاهای test نباید مورد استفاده قرار گیرند. Validation set شما نباید بیشتر از 20 درصد از سمپل های train شما باشد.

برای انجام پروژه حتما باید یک مدل pre trained شده انتخاب کنید و از این تاریخ به مدت یک هفته فرصت دارید اسم مدل خود را در گروه بفرستید مدل شما نباید تکراری باشد. البته می توانید روی چندین مدل اجرا بگیرید و در این صورت اگر یک مدل غیر تکراری داشته باشد کفایت می کند و بالاترین معیاری که به دست آمده برای شما در نظر گرفته می شود. (احتیاجی به نوشتن اسامی تمامی مدل ها در گروه نیست)

در فرایند train باید بهترین مدل خود را که روی validation set از نظر accuracy بهترین هست را در drive ذخیره کنید و بهترین مدل را روی سمپل های test ارزیابی می کنید.

برای ارزیابی مدل از معیار های زیر استفاده کنید.

Accuracy, AUC, Specifity, F1-score, Sensitivity

باید نمودار train accuracy , validation loss و train loss , validation accuracy را بکشید.

علاوه بر آن یک فایل csv ایجاد کنید و در هر epoch متریک هایی که به دست آمده را گزارش کنید.

پروژه دارای نمره تشویقی می باشد.

دقت داشته باشید هنگام ارائه پروژه باید تمرین ها را نیز ارائه دهید.

تاریخ تحویل پروژه 1402/04/07

<https://drive.google.com/file/d/14kfC3sUUvJX9deS8sJ3sqZQAoVkVnqpc/view?usp=sharing>

موفق باشید.