



تمرین ششم (مینی پروژه) پردازش زبان طبیعی

توجه:

- کد باید فقط در زبان پایتون باشد.
- استفاده از کتابخانه‌های از پیش آماده مجاز است. در گزارش کتابخانه‌ها ذکر شوند.
- لطفاً علاوه بر ضمیمه کردن کد، نتایج را تحلیل و در فایل گزارش خود ضمیمه کنید.
- فایل گزارش به فرمت pdf و به زبان فارسی باشد.
- تمرین در سایت کوئرا باید آپلود گردد. لینک ورود، رمز ورود: NLP-SP-03

در این تمرین با استفاده از تنها یک شبکه عصبی مسئله POS و NER را به صورت همزمان حل می‌کنید. دادگان به سه دسته آموزشی، اعتبارسنجی و تست تقسیم شده و در این لینک قرار دارند و با استفاده از پایتون pickle شده‌اند. هر بخش داده دارای ۳ فیلد `tokens`، `pos_tags` و `ner_tags` است، اما دادگان تست برچسب ندارند.

۱. مدلی بر پایه شبکه پیش آموزش دیده `bert-base-cased` برای پیش بینی POS و NER طراحی کنید (این مدل پیش آموزش دیده فقط نقش `backbone` را دارد، در نتیجه شما آزادی عمل در طراحی مدل کلی را دارید). در نظر داشته باشید که مدل طراحی شده باید به صورت پایان-به-پایان^۱ دو برچسب را پیش بینی کند (Multi-Tasking Learning).

۲. مدل و روش‌های تان با استفاده از معیارهای دقت، `precision`، `recall` و `f1-score` مورد ارزیابی قرار دهید و گزارش کنید.

۳. مدل بهینه تان را انتخاب کنید و بر روی دادگان تست پیش‌بینی انجام دهید. پیش‌بینی را با فرمت مشابه برچسب‌های داده آموزشی `test_pred.pickle` کرده و با نام `test_pred.pickle` ضمیمه کنید.

۴. گزارش کامل از پیش پردازش، طراحی مدل، نحوه آموزش و ... به همراه کد و پیش بینی‌های تان ضمیمه کنید.

۵. در نظر داشته باشید که میزان دقت بر روی دادگان تست و گزارش شما در نمره نهایی پروژه تاثیر بالایی خواهد داشت.