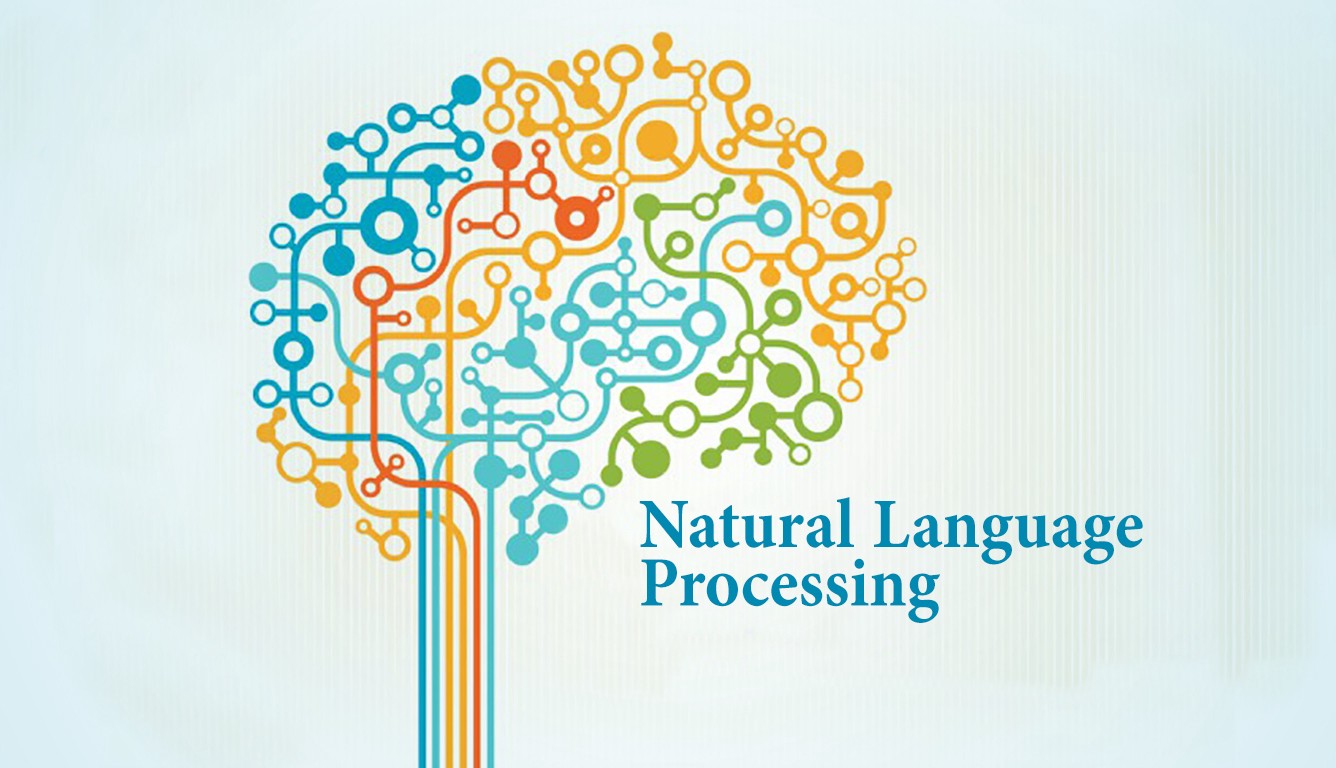
**گزارش پروژه پایانی پردازش زبان های طبیعی**



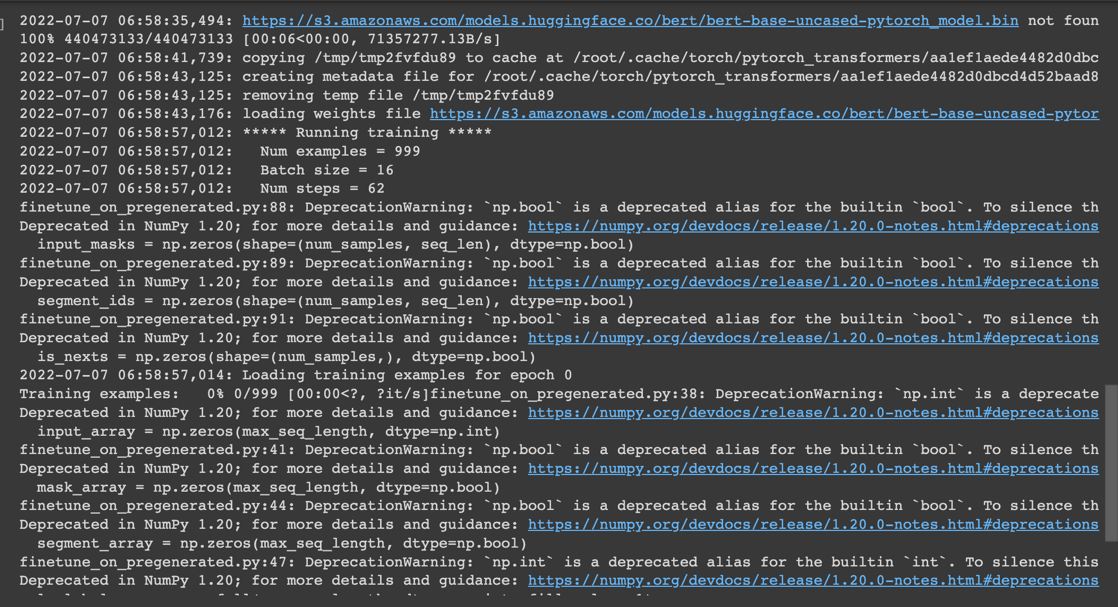
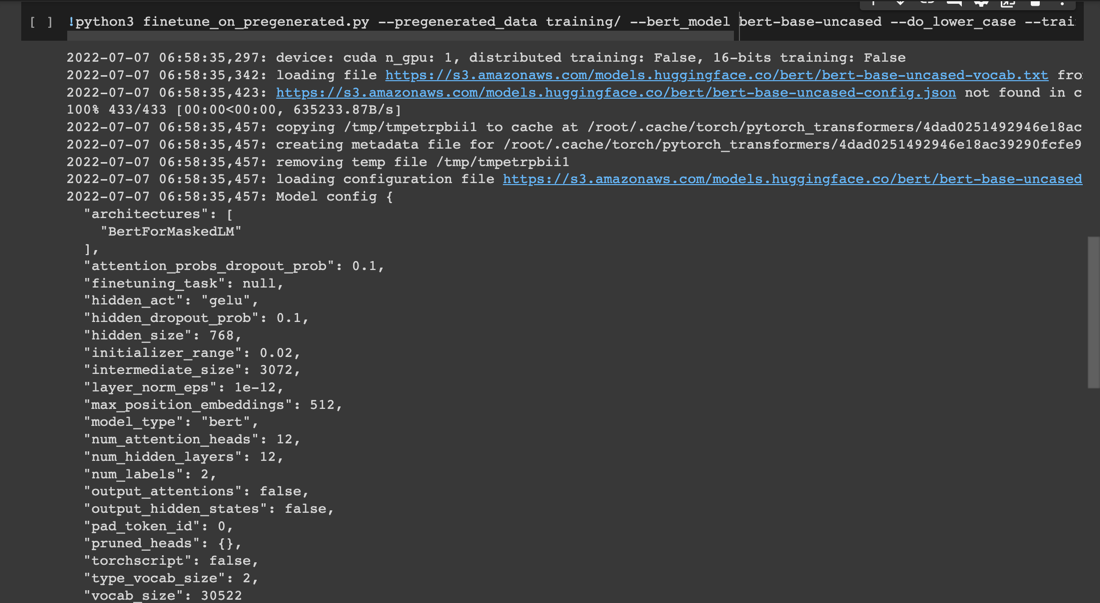
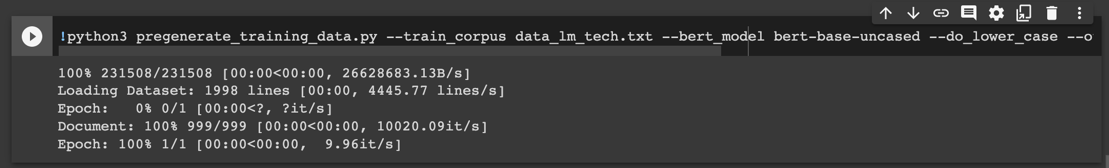
**نگین درخشان ، کیوان بوشهری**

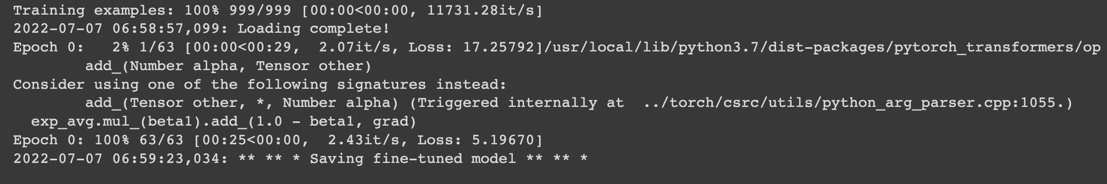
راهنما : رضا قهرمانی

استاد درس : دکتر مینایی

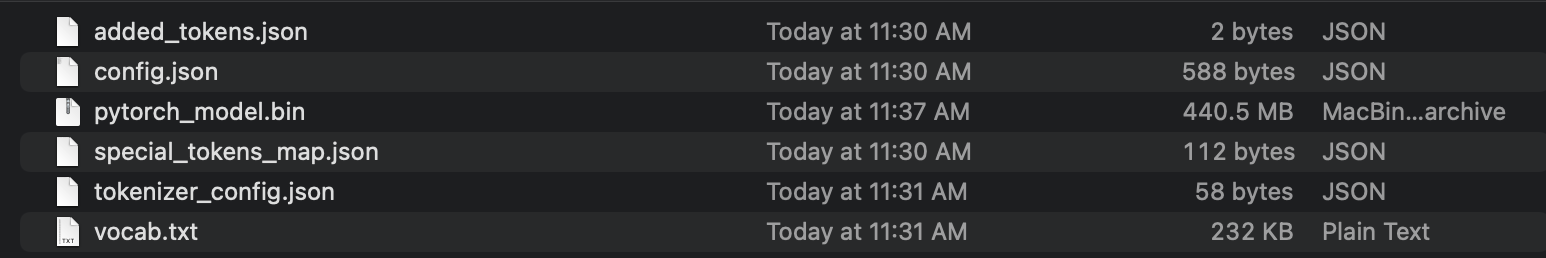
**بخش اول : تولید جملات**

در این بخش برای تولید جملات از مدل bert استفاده کرده ایم. و برای هر دسته از دیتایمان مدل bert را finetune کردیم:

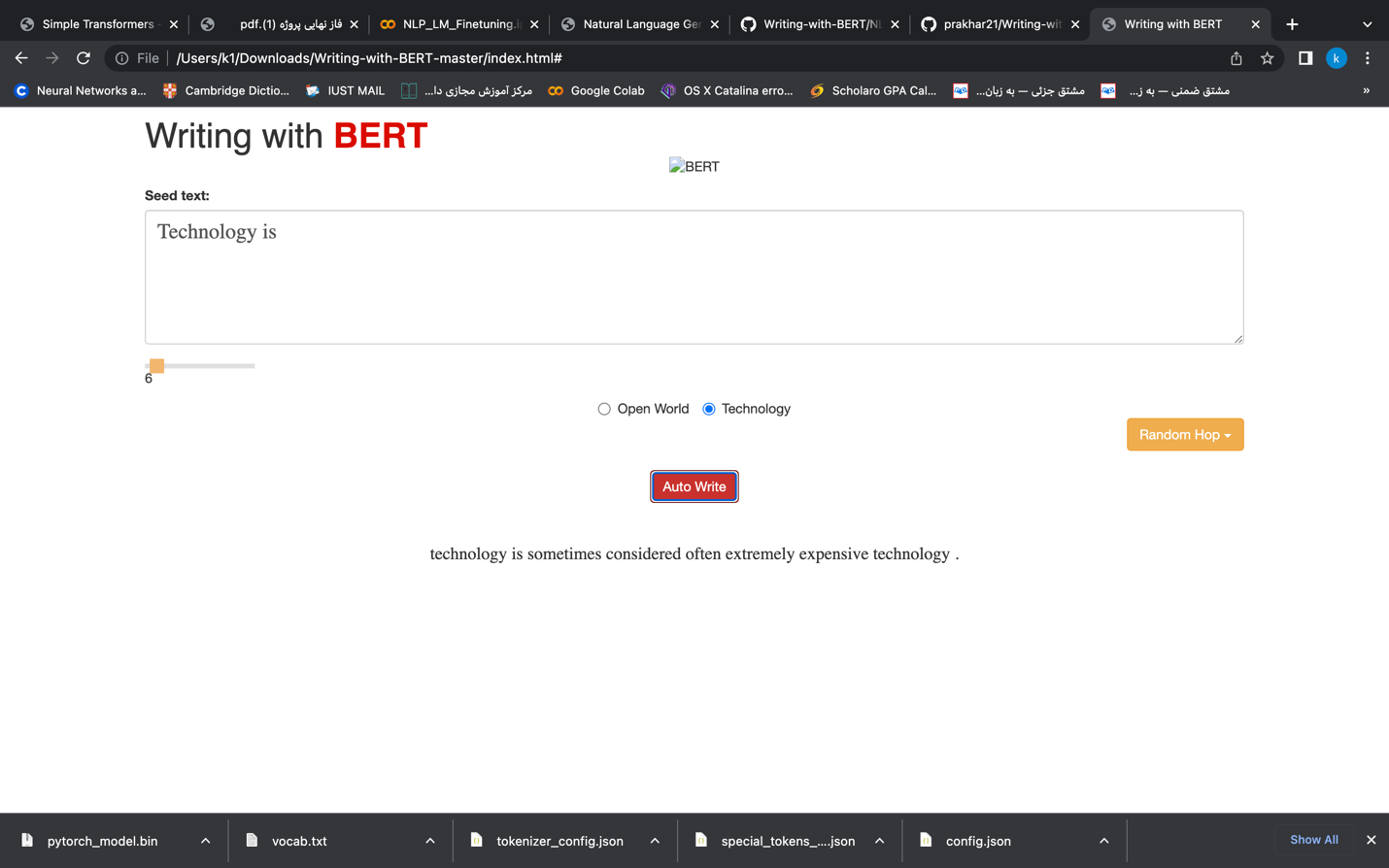




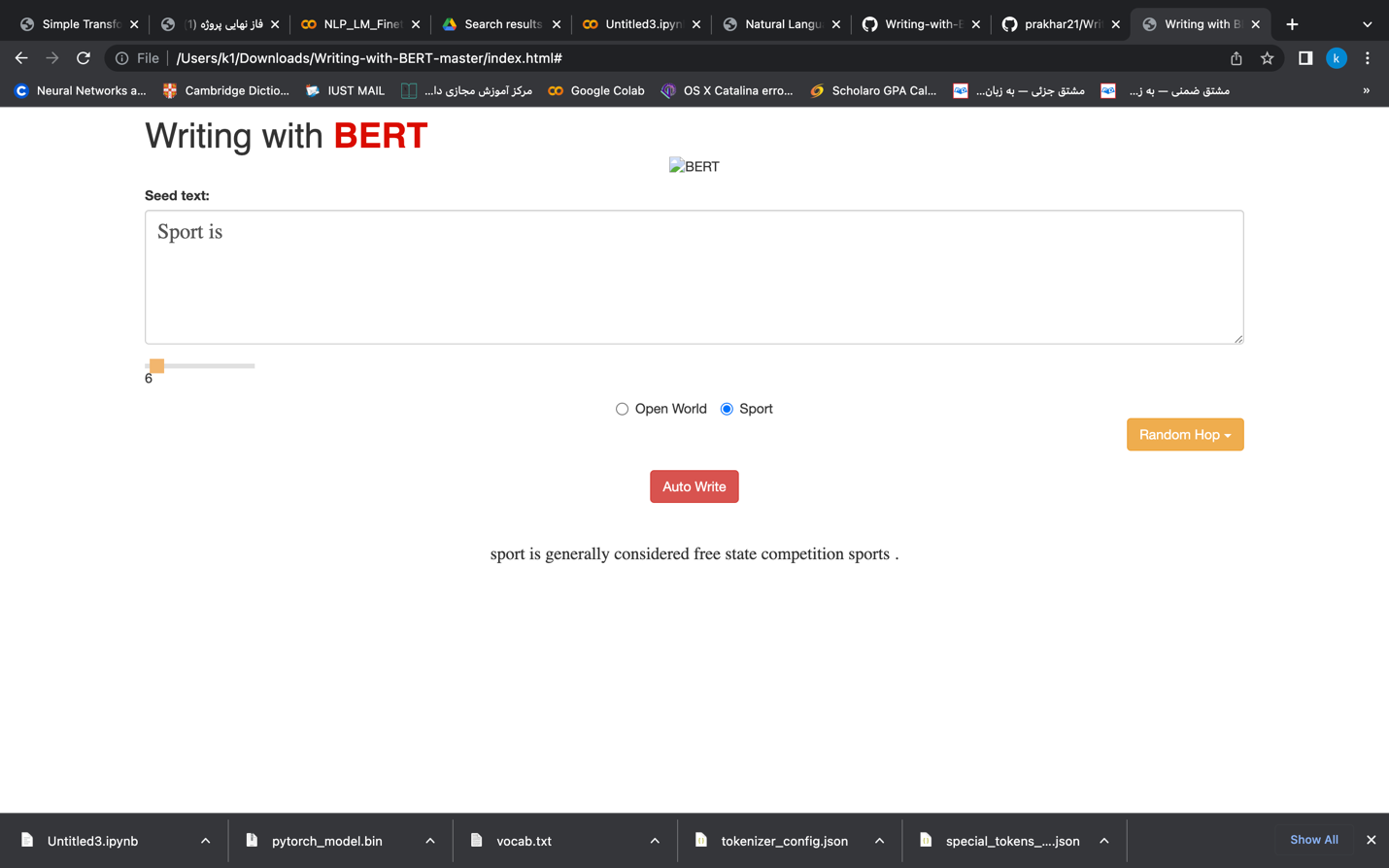
نتیجه نهایی برای مدل تکنولوژی در فولدر finetuned\_lm:



حال برای تولید چند جمله در زمینه تکنولوژی از مدلمان استفاده می کنیم:



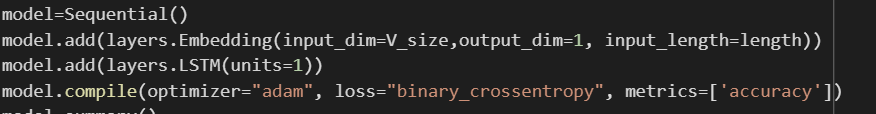
همین کار را برای دسته ورزشی تکرار کردیم:



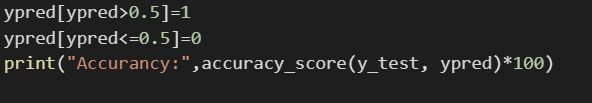
بخش دوم:شبکه با معماری ساده

در این قسمت ابتدا جملات تولید شده در قسمت های قبلی را که در دو دستهtest , train هستند را خوانده و آرایه های xtrain, ytrain, xtest, ytest را تشکیل دادیم. در قدم بعدی با استفاده از tokenizer جملات آرایه x را token کرده و تبدیل به دنباله(sequences) کردیم و دنباله را به آرایه ورودی شبکه عصبی تبدیل کردیم.

پس از تولید ورودی مناسب، شبکه عصبی را ساختیم. ابتدا مدل را به صورت Sequential تعریف کرده و یک شبکه lstm به آن اضافه کردیم. در ادامه مدل را کامپایل کرده و روی داده های train آموزش دادیم.



در ادامه با استفاده از predict و داده تست، داده پیشبینی شده را بدست آوردیم. این داده مقدار بین 0 و 1 دارد که با توجه به 0 یا 1 بودن کلاس ها با تابع sigmoid کلاس پیش بینی شده را بدست اوردیم. در نهایت با استفاده از مقدار کلاس واقعی مقدار دقت را بدست آوردیم.



شبکه ساخته شده و مقدار دقت این شبکه روی داده تست به صورت زیر است:

