

**دانشگاه صنعتي شريف**

**دانشکده مهندسي کامپيوتر**

**درس پردازش زبان های طبیعی**

**عنوان:**

**سیستم پرسش و پاسخ پزشکی**

**اعضای گروه:**

**سید محمدرضا حسینی**

**فرحان سراوند**

**وحیدالدین مقیمی**

**استاد: احسان الدین عسگری**

#### فهرست مطالب

1) بازیابی متون مرتبط

2) تولید پاسخ

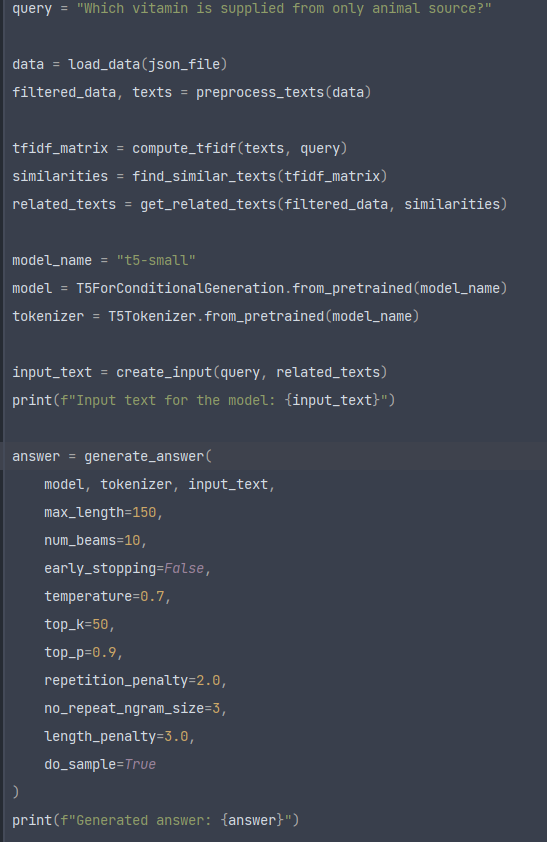
### بازیابی متون مرتبط

برای ساخت این سیستم در ابتدا نیاز به پیدا کردن دیتاست مناسب برای انجام این تمرین بود. Medmcqa دیتاست استفاده شده است که یک دیتاست برای پاسخ گویی به سوالات پزشکی میباشد. پس از دانلود دیتاست، برای اینکه بتوانیم جواب مورد نظرمان را بیابیم، میبایست بتوانیم از یک تکنیک مناسب برای بازیابی متون استفاده کنیم. تکنیک استفاده شده بدین صورت است که در ابتدا با استفاده از TF\_IDF ، کوئری و سوال فرد و جواب های موجود را تبدیل میکنیم سپس با استفاده از Cosine similiarity میزان شباهت کوئری و جواب ها را میسنجیم و بر اساس بالاترین شباهت جواب ها را مرتب میکنیم

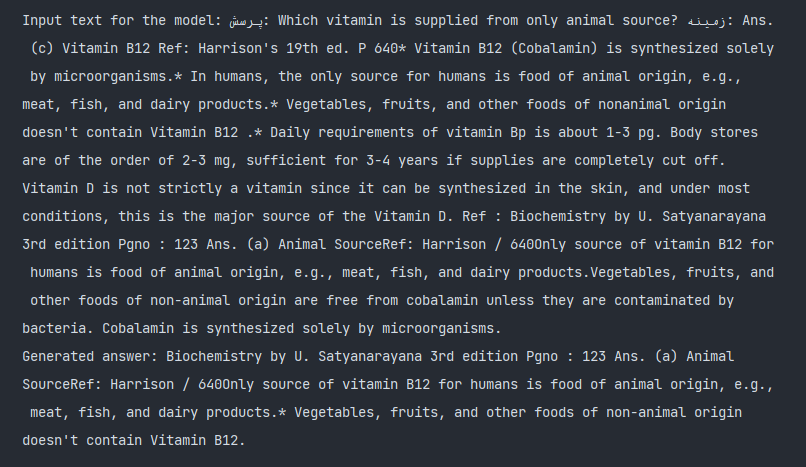
### 2. تولید پاسخ

هدف از این تسک پیاده سازی یک RAG(Retrieval Augment Generation) است که در آن با استفاده از داده های پزشکی به توان به سوالات مطرح شده پاسخ مناسبی داد.

پس از انتخاب دیتاست و پیاده سازی تکنیک بازیابی متون به سراغ تسک اصلی میرویم. در ابتدا از مدل پایه بدون هیچ فاین تیونیگ استفاده میکنیم.

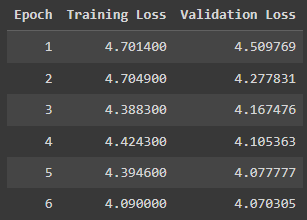


پرسش و پاسخ به صورت زیر میباشد:



سپس به سراغ فاین تیون کردن مدل T5 میرویم. برای اینکار تعداد داده های مختلفی آزمایش شد و در نهایت 2000 پرسش و پاسخ اول موجود در دیتاست برای فاین تیون کردن مدل استفاده شد تا بتوان اینکار را در یک زمان مناسب انجام داد. جواب ها در یک لیست و سوال ها در یک لیست دیگر ذخیره میشود. سوال ها به عنوان ورودی و جواب به عنوان خروجی مطلوب که مدل باید به آن دست پیدا کند فرض میشود. برای توکنایز کردن داده ها نیز از توکنایزر مدل T5 استفاده شده تا با آن همخوانی داشته باشد. این مدل 6 ایپاک فاین تیون شده و Loss های دریافت شده از آموزش و ولیدیشن نشان دهنده موفق بودن فاین تیونیگ میباشد.





پس از فاین تیون کردن مدل از این مدل برای تولید پاسخ با توجه به دیتا استفاده میکنیم. در ابتدا مطابق روش گفته شده پرسش به همراه جواب ها با استفاده از TF\_IDF تبدیل میشوند سپس با استفاده از cosine\_similiarity مرتبط ترین جواب ها پیدا میشوند. با دادن مرتبط ترین جواب ها و کوئری به مدل فاین تیون شده، جواب به آن سوال تولید میشود.

برای تولید پاسخ تعدادی پارامتر ست شده است که به صورت زیر میباشد:

max\_length = 150 که حداکثر میزان توکن تولیدی را نشان میدهد

num\_beams = 5 که تعداد پرتو های استفاده شده در beam search را نمایش میدهد. ست کردن درست این پارامتر سبب میشود که تعداد جواب های تولیدی بیشتر شود و بهترین آن ها برای خروجی استفاده شود.

early\_stopping : ست کردن این پارامتر سبب میشود که در صورتی که به اندازه تعداد پرتو جواب ها را تولید کردیم، عملیات متوقف شود

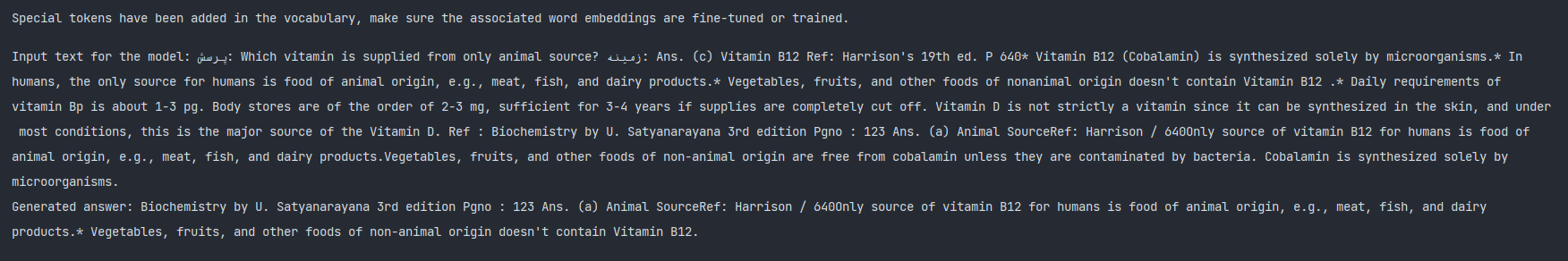
temprature=0.7 با ست کردن این پارامتر میزان قطعی بودن یا تخیلی بودن جواب را تعیین میکنیم. هر چه این عدد کم تر و نزدیک به صفر باشد جواب با توجه به منابع موجود در پایگاه دانش تولید میشود و هرجه بیشتر باشد مدل نیز سعی میکند جواب را طبق دانسته ها تغییر دهد.

top\_k=50 تعداد کاندیداهای برتر که باید از آن نمونه برداری شود را نشان میدهد.

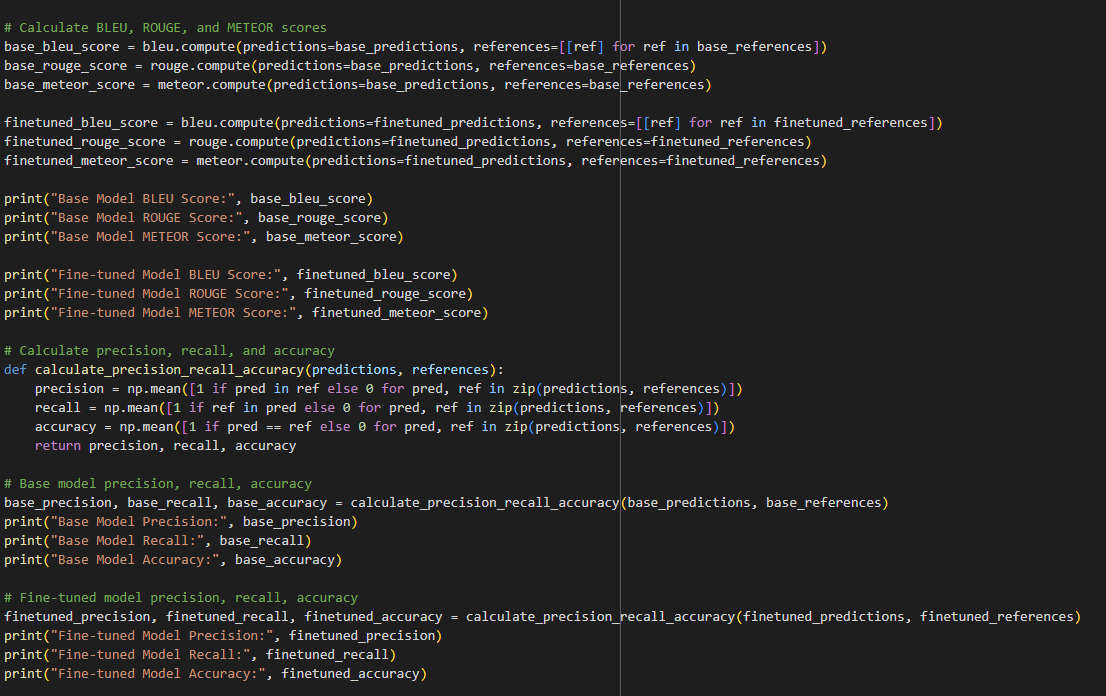
top\_p = 0.9 نمونه گیری هسته ای که مجموع احتمالات را تا یک آستانه خاص شامل میشود.

repetition\_penalty=0.2 این پارامتر مقدار جریمه ای که به ازای تکرار کلمات اتفاق میفتد را نشان میدهد. این جریمه از تکرار کردن کلمات توسط مدل جلوگیری میکند.

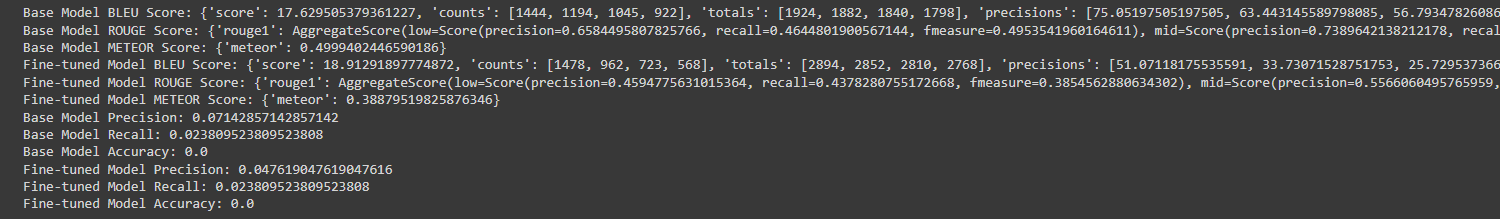
سپس پس از فاین تیون با استفاده از روش گفته شده و استفاده از TF\_IDF و Cosine Similiarity جواب های پیدا شده به مدل داده میشود و مدل جواب نهایی را تولید میکند. جواب RAG پیاده سازی شده با مدل فاین تیون شده به صورت زیر است:



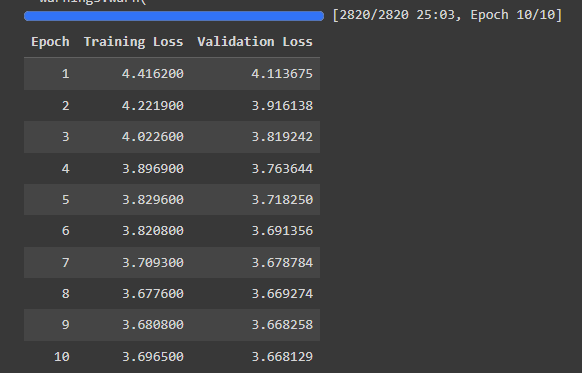
برای ارزیابی مدل از بخشی از دیتاست استفاده شده که برای فاین تیون مدل استفاده نشده است تا مدل آن را ندیده باشد و بر اساس آن تست ها انجام شود.



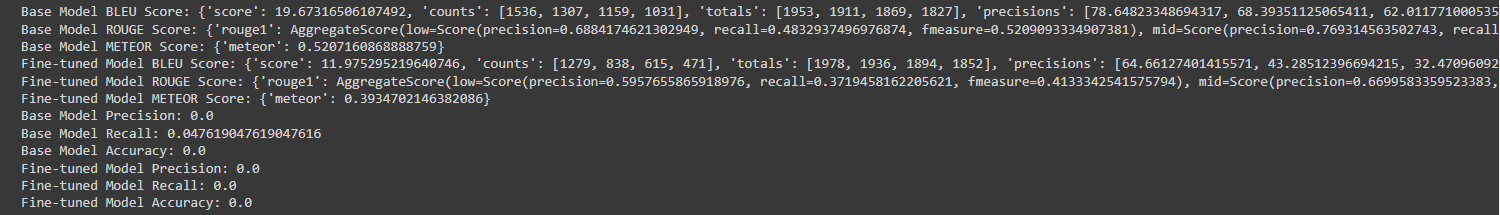
در هنگام فاین تیونیگ پارامتر های مختلفی امتحان شد تا عملکرد مدل بهبود یابد ولی تنها حالتی که بهترین جواب را از فاین تیون کردن مدل دریافت کردیم در حالت فاین تیون توضیح داده شده است. که همانطور که مشخص است در معیار BLUE نسبت به حالت پایه بهبود داشته ایم. دلیل اینکه در سایر معیار ها دچار افت شده ایم به دلیل وسیع بودن داده استفاده شده و حجم اطلاعات زیادی است که سعی در آموزش مدل داریم.



به عنوان مثال حالت زیر حالتی بوده که مقدار Training loss , evaluation Loss بیشتر کاهش پیدا کرده و به نظر در فاین تیون کردن مدل موفق تر بوده ایم.



در معیار BLUE این مدل به شدت نسبت به مدل گفته شده در بالا عمل میکند ولی توانسته METEOR و ROUGE بالاتری نسبت به مدل بالا داشته باشد ولی عملکرد آن نسبت به مدل پایه ضعیف تر است.س



سعی کردیم تعداد داده ها و ایپاک ها را نیز تغییر داده و پس از لرنینگ ریت و وارمینگ استپ را نیز تغییر دهیم ولی باز نتایج از مدل پایه ضعیف تر عمل کرده اند :

