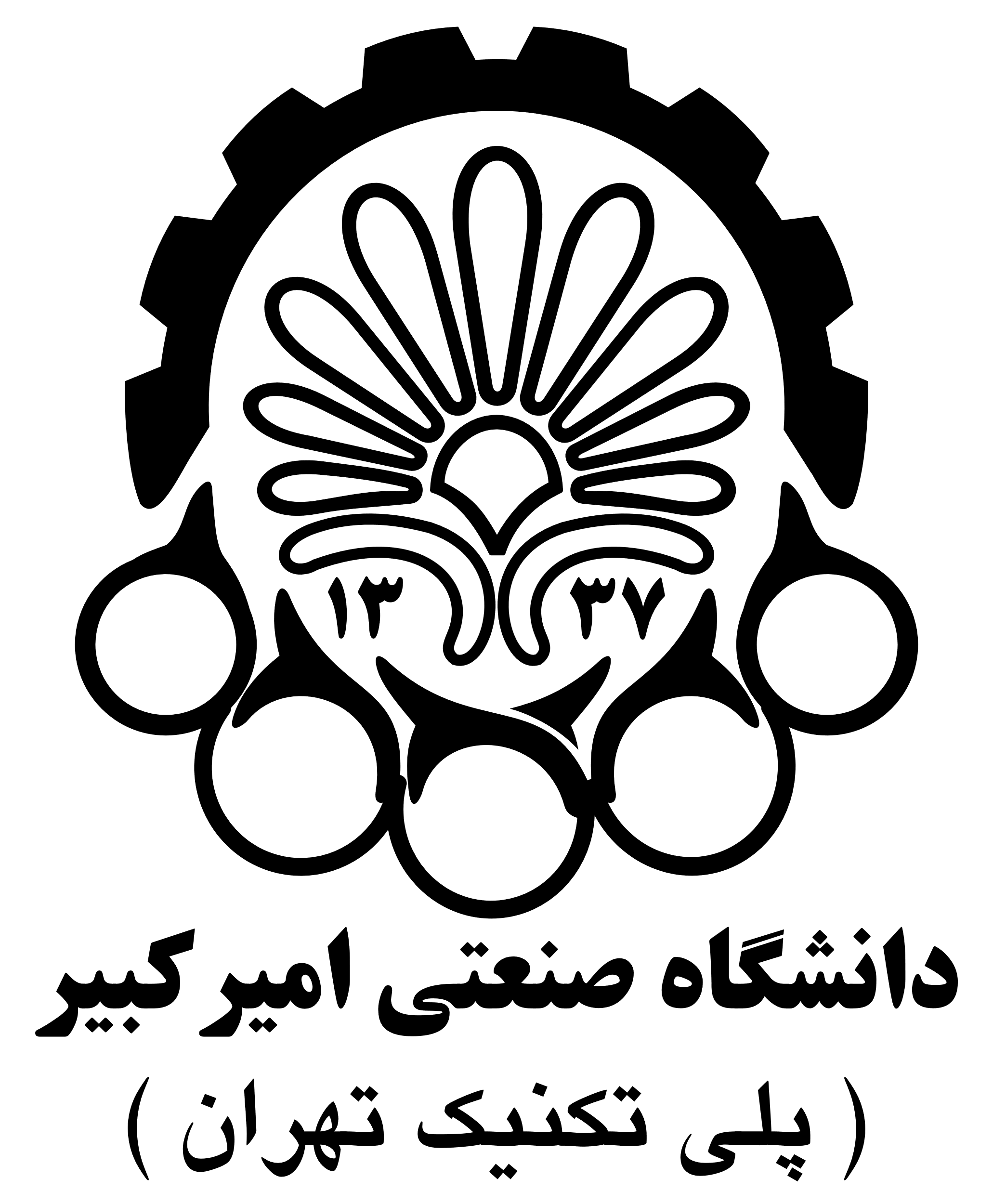
****

**دانشکده مهندسی کامپیوتر**

**پروژه پایانی درس فهم زبان**

**دکتر زینلی**

**غلامرضا دار 400131018**

**بهار 1401**

فهرست مطالب

[1) تولید شعر به کمک T5 3](#_Toc128241924)

[2) تولید شعر به کمک GPT-2 14](#_Toc128241925)

**نکته در مورد این تمرین:** با توجه به اینکه در کد مربوط به این تمرین از، دستوراتی برای متصل شدن به اکانت Huggingface و Google Drive و ... استفاده شده است، این کد قابلیت تبدیل به صورت ماژولار را نداشت و در نتیجه، کد این تمرین به صورت چند نوت بوک که در Google Colab قابل مشاهده و اجرا هستند ارائه می­شود. همچنین خروجی­های بیشتر در این نوت­بوک­ها قابل رویت است.

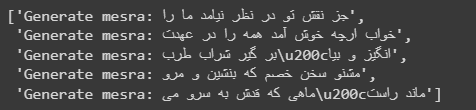
# 1) تولید شعر به کمک T5

در این بخش از پروژه می­خواهیم به کمک مدل T5، اشعاری مشابه اشعار حافظ تولید کنیم. برای این­کار به طور خاص از مدل MT5 که یک مدل چند زبانه از گوگل است استفاده خواهیم کرد. برای نوشتن این بخش از پروژه از یکی از نوت­بوک­های موجود در وبسایت Huggingface که در رابطه با گرفتن سورس کدی به زبان ruby و تولید چند خط توضیح برای آن بود استفاده کردیم. این نوت بوک از طریق لینک زیر قابل دسترسی است.

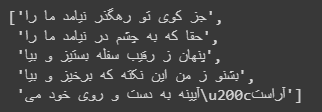


<https://colab.research.google.com/github/NielsRogge/Transformers-Tutorials/blob/master/T5/Fine_tune_CodeT5_for_generating_docstrings_from_Ruby_code.ipynb>

برای تولید مصراع بعدی به کمک مصراع قبل، داده را به شکل زیر در خواهیم آورد.



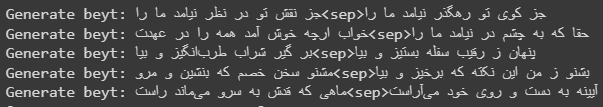
Inputs



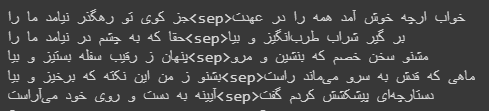
Labels

به ابتدای هر مصراع، عبارت Generate Mesra: را اضافه می­کنیم. ایده این­کار از نوت بوک اصلی آمد که در آنجا قبل از سورس کد، عبارت Summarize: را اضافه می­کردند. با این کار، مدل هر گاه Generate Mesra: و در ادامه یک مصراع را ببیند، یاد خواهد گرفت مصراع بعد را تولید کند.

به طور مشابه برای حالت تولید بیت جدید، از Generate beyt: استفاده کردیم و بین مصراع ها نیز <sep> گذاشتیم.

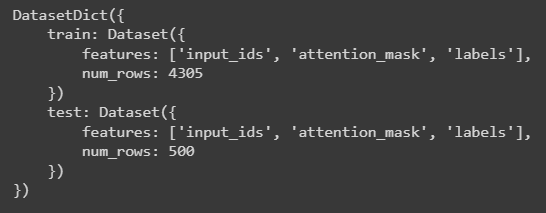


Inputs

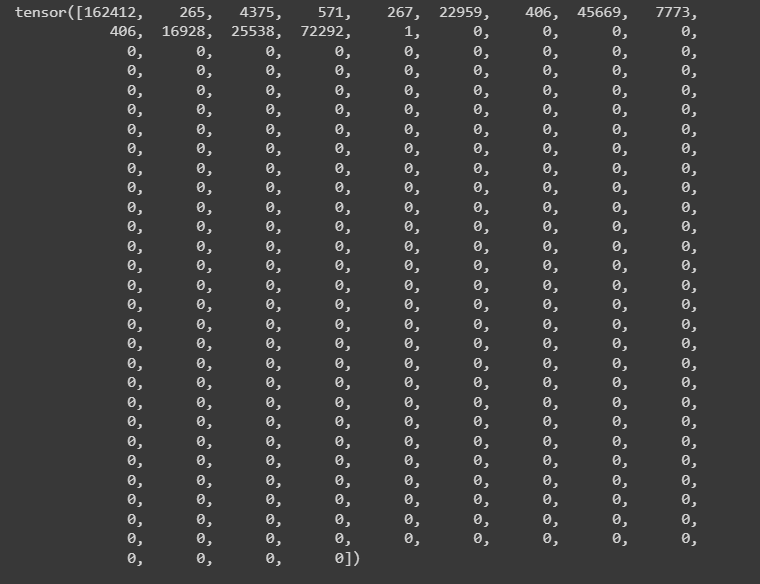


Labels

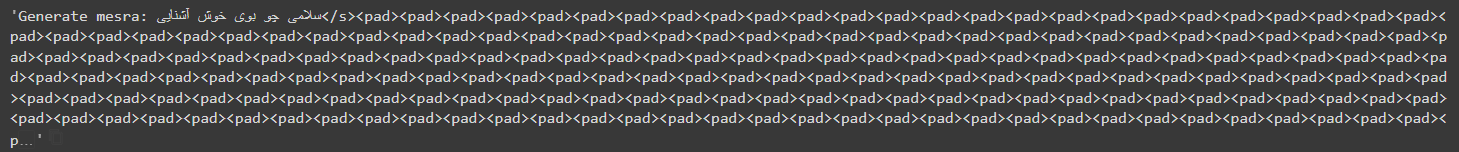
ساختار دیتاست پس از توکنایز شدن به شکل زیر در آمد: حدود 4300 مصراع آموزش و 500 مصراع تست. همچنین فیچر­های موجود در دیتاست، خروجی­های توکنایزر این مدل هستند.



یکی از مصراع­ها پس از توکنایز شدن:



و نسخه دیکد شده آن:



حال که مجموعه داده را به صورت توکنایز شده در اختیار داریم، می­توان مدل را آموزش داد.

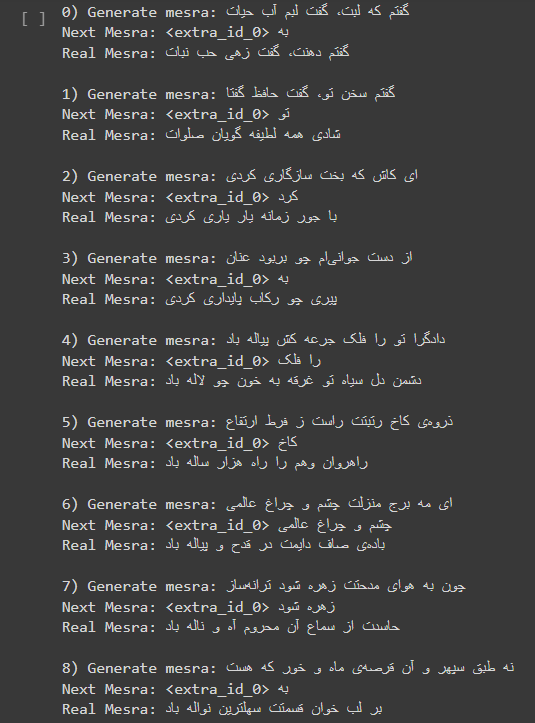
لازم به ذکر است که در این بخش به مشکل خوردیم و در نهایت نتوانستیم مدل مناسبی را آموزش دهیم. تعداد متغیر­های ممکن برای رسیدن به مدل خوب بسیار زیاد، مانند یافتن ماکسیمم اندازه هر ورودی و خروجی، به صورت دیفالت 256 و 128 توکن در نظر گرفته شده بود که ممکن است برای این مسئله زیاد باشد، هایپرپارامتر­های آموزش و تولید خروجی، بخش تولید خروجی پارامتر­های زیادی دارد که نیاز به آزمون و خطای زیادی دارد مانند اندازه beam و ... .

نکته دیگری که فهمیدیم این بود که این مدل به سرعت Overfit می­شود. داده­های حافظ برای همچین مدلی بسیار کم است و متاسفانه به دلیل کمبود وقت نتوانستیم این مدل را بر روی کل مجموعه داده شعرا نیز بررسی کنیم.

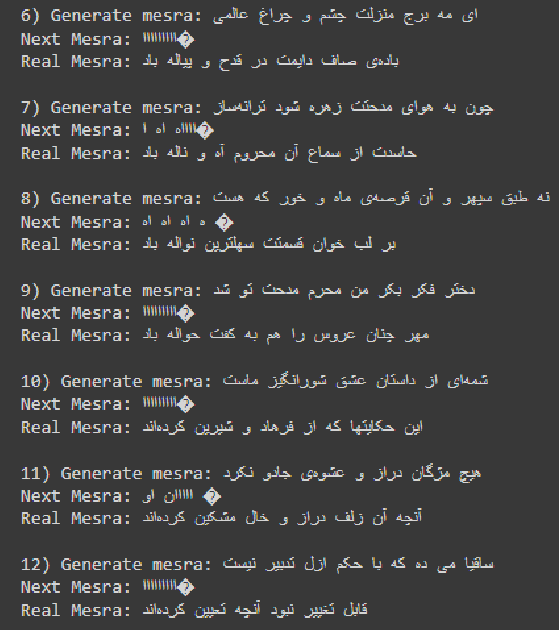
در ادامه چند نتیجه نسبتا نامطلوب از این مدل را خواهیم دید. این نتایج در اثر آموزش مدل به میزان epoch های مختلف بدست

آمده اند.

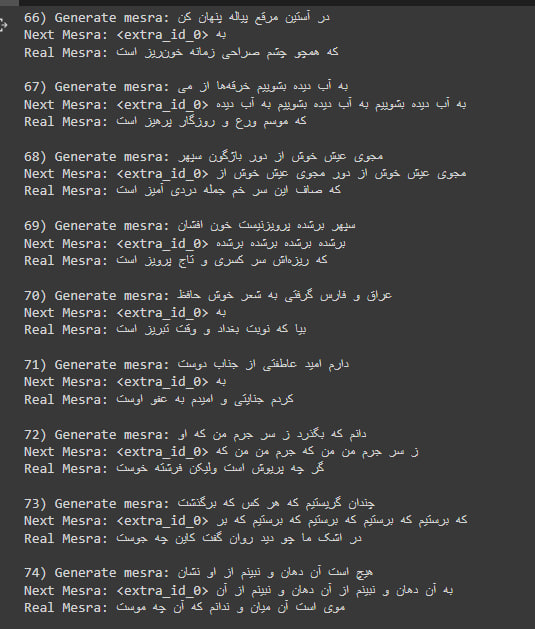
اینجا مدل یاد گرفته بود مصراع­های خیلی کوتاه تولید کند!



اینجا مدل یاد گرفت داد آه و ناله کند!😂



اینجا مدل یاد گرفته متن مصراع اول را کپی پیست کند!



در یکی دیگر از آموزش­ها که تصویرش موجود نیست، مدل یاد گرفته بود کلمه حافظ را در هر مصراع بیاورد که نشان دهنده Overfit شدن مدل بر روی این مجموعه داده است.

همچنین چون نتایج مطلوبی به دست نیامد، مدل را برای تولید **بیت** آموزش ندادیم اما کد نوشته شده، پس از یافتن مشکل آموزش، با استفاده از فلگ **mesra\_mode=True** قابلیت تولید بیت را نیز دارد.

**تعدادی از نتایج تولید مصراع برای یکی از مدل­ها**

0) Generate mesra: گفتم که لبت، گفت لبم آب حیات

Next Mesra: <extra\_id\_0> به

Real Mesra: گفتم دهنت، گفت زهی حب نبات

1) Generate mesra: گفتم سخن تو، گفت حافظ گفتا

Next Mesra: <extra\_id\_0> تو

Real Mesra: شادی همه لطیفه گویان صلوات

2) Generate mesra: ای کاش که بخت سازگاری کردی

Next Mesra: <extra\_id\_0> کرد

Real Mesra: با جور زمانه یار یاری کردی

3) Generate mesra: از دست جوانی‌ام چو بربود عنان

Next Mesra: <extra\_id\_0> به

Real Mesra: پیری چو رکاب پایداری کردی

4) Generate mesra: دادگرا تو را فلک جرعه کش پیاله باد

Next Mesra: <extra\_id\_0> را فلک

Real Mesra: دشمن دل سیاه تو غرقه به خون چو لاله باد

5) Generate mesra: ذروه‌ی کاخ رتبتت راست ز فرط ارتفاع

Next Mesra: <extra\_id\_0> کاخ

Real Mesra: راهروان وهم را راه هزار ساله باد

6) Generate mesra: ای مه برج منزلت چشم و چراغ عالمی

Next Mesra: <extra\_id\_0> چشم و چراغ عالمی

Real Mesra: باده‌ی صاف دایمت در قدح و پیاله باد

7) Generate mesra: چون به هوای مدحتت زهره شود ترانه‌ساز

Next Mesra: <extra\_id\_0> زهره شود

Real Mesra: حاسدت از سماع آن محروم آه و ناله باد

8) Generate mesra: نه طبق سپهر و آن قرصه‌ی ماه و خور که هست

Next Mesra: <extra\_id\_0> به

Real Mesra: بر لب خوان قسمتت سهلترین نواله باد

9) Generate mesra: دختر فکر بکر من محرم مدحت تو شد

Next Mesra: <extra\_id\_0> شدم

Real Mesra: مهر چنان عروس را هم به کفت حواله باد

10) Generate mesra: شمه‌ای از داستان عشق شورانگیز ماست

Next Mesra: <extra\_id\_0> عشق شورانگیز ماست عشق شورانگیز ماست عشق شورانگیز ماست

Real Mesra: این حکایتها که از فرهاد و شیرین کرده‌اند

11) Generate mesra: هیچ مژگان دراز و عشوه‌ی جادو نکرد

Next Mesra: <extra\_id\_0> نکرد

Real Mesra: آنچه آن زلف دراز و خال مشکین کرده‌اند

12) Generate mesra: ساقیا می ده که با حکم ازل تدبیر نیست

Next Mesra: <extra\_id\_0> می ده

Real Mesra: قابل تغییر نبود آنچه تعیین کرده‌اند

13) Generate mesra: در سفالین کاسه‌ی رندان به خواری منگرید

Next Mesra: <extra\_id\_0> به رندان

Real Mesra: کاین حریفان خدمت جام جهان‌بین کرده‌اند

14) Generate mesra: نکهت جانبخش دارد خاک کوی دلبران

Next Mesra: <extra\_id\_0> کوی کوی کوی کوی کوی کوی کوی کوی کوی

Real Mesra: عارفان آنجا مشام عقل مشکین کرده‌اند

15) Generate mesra: ساقیا دیوانه‌ای چون من کجا دربر کشد

Next Mesra: <extra\_id\_0> به

Real Mesra: دختر رز را که نقد عقل کابین کرده‌اند

16) Generate mesra: خاکیان بی‌بهره‌اند از جرعه‌ی کاس الکرام

Next Mesra: <extra\_id\_0> به

Real Mesra: این تطاول بین که با عشاق مسکین کرده‌اند

17) Generate mesra: شهپر زاغ و زغن زیبای صید و قید نیست

Next Mesra: <extra\_id\_0> و زغ و زغ و زغ و زغ و

Real Mesra: این کرامت همره شهباز و شاهین کرده‌اند

18) Generate mesra: آن میوه‌ی بهشتی کآمد به دستت ای جان

Next Mesra: <extra\_id\_0> ی بهشتی کآمد به دستت

Real Mesra: در دل چرا نکشتی از دست چون بهشتی

19) Generate mesra: تاریخ این حکایت گر از تو باز پرسند

Next Mesra: <extra\_id\_0> تو باز پرسند تو باز پرسند تو باز پرسند تو باز پرسند تو تو

Real Mesra: سرجمله‌اش فروخوان از میوه‌ی بهشتی

20) Generate mesra: می‌دمد صبح و کله بست سحاب

Next Mesra: <extra\_id\_0> به

Real Mesra: الصبوح الصبوح یا اصحاب

21) Generate mesra: می‌چکد ژاله بر رخ لاله

Next Mesra: <extra\_id\_0>

Real Mesra: المدام المدام یا احباب

22) Generate mesra: می‌وزد از چمن نسیم بهشت

Next Mesra: <extra\_id\_0> به

Real Mesra: هان بنوشید دم به دم می ناب

23) Generate mesra: تخت زمرد زده است گل به چمن

Next Mesra: <extra\_id\_0> است

Real Mesra: راح چون لعل آتشین دریاب

24) Generate mesra: در میخانه بسته‌اند دگر

Next Mesra: <extra\_id\_0> میخانه بسته اند

Real Mesra: افتتح یا مفتح الابواب

25) Generate mesra: لب و دندانت را حقوق نمک

Next Mesra: <extra\_id\_0> و دندانت را حقوق

Real Mesra: هست بر جان و سینه‌های کباب

26) Generate mesra: این چنین موسمی عجب باشد

Next Mesra: <extra\_id\_0> است

Real Mesra: که ببندند میکده به شتاب

27) Generate mesra: بر رخ ساقی پری پیکر

Next Mesra: <extra\_id\_0>

Real Mesra: همچو حافظ بنوش باده ناب

28) Generate mesra: خمی که ابروی شوخ تو در کمان انداخت

Next Mesra: <extra\_id\_0> تو در کمان اند ی ی ی ی ی ی

Real Mesra: به قصد جان من زار ناتوان انداخت

29) Generate mesra: نبود نقش دو عالم که رنگ الفت بود

Next Mesra: <extra\_id\_0> نبود نقش دو عالم که که نبود که نبود که

Real Mesra: زمانه طرح محبت نه این زمان انداخت

30) Generate mesra: به یک کرشمه که نرگس به خودفروشی کرد

Next Mesra: <extra\_id\_0> به خودفروشی

Real Mesra: فریب چشم تو صد فتنه در جهان انداخت

31) Generate mesra: شراب خورده و خوی کرده می‌روی به چمن

Next Mesra: <extra\_id\_0> می روی می روی می روی می روی می روی می روی می روی می روی می روی

Real Mesra: که آب روی تو آتش در ارغوان انداخت

32) Generate mesra: به بزمگاه چمن دوش مست بگذشتم

Next Mesra: <extra\_id\_0>م به

Real Mesra: چو از دهان توام غنچه در گمان انداخت

33) Generate mesra: بنفشه طره مفتول خود گره می‌زد

Next Mesra: <extra\_id\_0> می زده می زده می زده می زده می زده می زده

Real Mesra: صبا حکایت زلف تو در میان انداخت

34) Generate mesra: ز شرم آن که به روی تو نسبتش کردم

Next Mesra: <extra\_id\_0> شدم

Real Mesra: سمن به دست صبا خاک در دهان انداخت

35) Generate mesra: من از ورع می و مطرب ندیدمی زین پیش

Next Mesra: <extra\_id\_0> می زین من ورع می زین

Real Mesra: هوای مغبچگانم در این و آن انداخت

36) Generate mesra: کنون به آب می لعل خرقه می‌شویم

Next Mesra: <extra\_id\_0> می لعل خرقه می شویم

Real Mesra: نصیبه ازل از خود نمی‌توان انداخت

37) Generate mesra: مگر گشایش حافظ در این خرابی بود

Next Mesra: <extra\_id\_0> است

Real Mesra: که بخشش ازلش در می مغان انداخت

38) Generate mesra: جهان به کام من اکنون شود که دور زمان

Next Mesra: <extra\_id\_0> به من به من به من به من به من به من به من

Real Mesra: مرا به بندگی خواجه جهان انداخت

39) Generate mesra: به جان خواجه و حق قدیم و عهد درست

Next Mesra: <extra\_id\_0> به جان خواجه و حق خواجه و حق خواجه و حق خواجه و حق

Real Mesra: که مونس دم صبحم دعای دولت توست

40) Generate mesra: سرشک من که ز طوفان نوح دست برد

Next Mesra: <extra\_id\_0> من که

Real Mesra: ز لوح سینه نیارست نقش مهر تو شست

41) Generate mesra: بکن معامله‌ای وین دل شکسته بخر

Next Mesra: <extra\_id\_0> به

Real Mesra: که با شکستگی ارزد به صد هزار درست

42) Generate mesra: زبان مور به آصف دراز گشت و رواست

Next Mesra: <extra\_id\_0> گشت

Real Mesra: که خواجه خاتم جم یاوه کرد و بازنجست

43) Generate mesra: دلا طمع مبر از لطف بی‌نهایت دوست

Next Mesra: <extra\_id\_0> مبر از لطف مبر از لطف مبر از لطف بی نهایت

Real Mesra: چو لاف عشق زدی سر بباز چابک و چست

44) Generate mesra: به صدق کوش که خورشید زاید از نفست

Next Mesra: <extra\_id\_0> به صدق کوش که به صدق کوش که به صدق کوش که

Real Mesra: که از دروغ سیه روی گشت صبح نخست

45) Generate mesra: شدم ز دست تو شیدای کوه و دشت و هنوز

Next Mesra: <extra\_id\_0> کوه و دشت و دشت

Real Mesra: نمی‌کنی به ترحم نطاق سلسله سست

46) Generate mesra: مرنج حافظ و از دلبران حفاظ مجوی

Next Mesra: <extra\_id\_0> حافظ و از دلبران حافظ و از دلبران حافظ و از دلبران

Real Mesra: گناه باغ چه باشد چو این گیاه نرست

47) Generate mesra: تا سر زلف تو در دست نسیم افتادست

Next Mesra: <extra\_id\_0> تو در زلف تو در زلف تو تو در زلف تو در

Real Mesra: دل سودازده از غصه دو نیم افتادست

48) Generate mesra: چشم جادوی تو خود عین سواد سحر است

Next Mesra: <extra\_id\_0> است تو خود عین سواد سحر است است

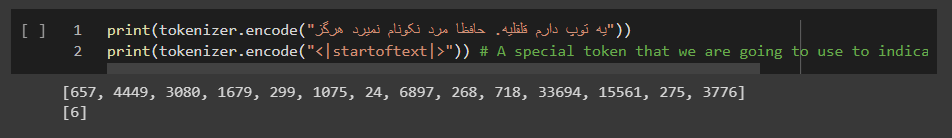
Real Mesra: لیکن این هست که این نسخه سقیم افتادست

**در مورد معیار BLEU به طور کلی در انتهای گزارش صحبت خواهیم کرد.**

# 2) تولید شعر به کمک GPT-2

در این بخش از پروژه می­خواهیم به کمک مدل GPT2-fa از هوشواره، مسئله بیان شده را حل کنیم. همچنین برای این بخش، از یکی از نوت­بوک هوشواره الهام­های زیادی گرفته شد که مربوط به فاین­تونینگ مدل GPT2-fa است.

در این بخش طبق پیشنهاد نوت­بوک اصلی، از توکن <|startoftext|> که یک توکن شناخته شده توسط توکنایزر این مدل است استفاده می­کنیم.



نحوه آماده سازی داده برای این سوال به این شکل است که مصراع اول به مدل داده می­شود، سپس توکن <|startoftext|> افزوده می­شود و انتظار می­رود مدل مصراع بعد را تولید کند. (درمورد تولید بیت بعد نیز به همین شکل عمل می­کنیم و بین مصراع­های هر بیت یک <sep> می­گذاریم.)

مثالی برای نمونه آورده شده است:

Sample input: ای دل گر از آن چاه زنخدان به درآیی<|startoftext|>

ای دل گر از آن چاه زنخدان به درآیی<|startoftext|> می‌آید. «در آن چاه زنخدان‌ها و در آن چاه زنی، همهٔ کسانی که از دل آن چاه زنخشان می‌کنند، در آن چاه

خروجی تولید شده که در مثال قبل میبینید برای مدل فاین تون نشده است. مدل GPT2-fa بر روی متن فارسی آموزش دیده است و سعی کرده است متن داده شده را کامل کند که در تکمیل متن نیز از کلماتی مانند چاه زنخدان که در ورودی آمده است استفاده کرده است.

در ادامه این مدل را بر روی اشعار حافظ فاین تون خواهیم کرد.

در حین آموزش این مدل، تعدادی خروجی نیز از عملکرد مدل در تکمیل مصراع­ها گرفتیم که به ترتیب زمان تولید شدن به شرح زیر هستند.

در ره عشق از آن سوی فنا صد خطر است فیه، در

ز شعر دلکش حافظ کسی بود آگاه که می‌گفت آن که ز

گفتم ای شام غریبان طره شبرنگ تو گفت این دعا

هرگز که دیده باشد جسمی ز جان مرکب توبه آیینه‌ی د

مدامم مست می‌دارد نسیم جعد گیسویت و اندر

پیر گلرنگ من اندر حق ازرق پوشان رخصت خبث

معاشران گره از زلف یار باز کنید شبی خوش است بدین

گل بخندید که از راست نرنجیم ولی هیچ عاشق سخن سخت به

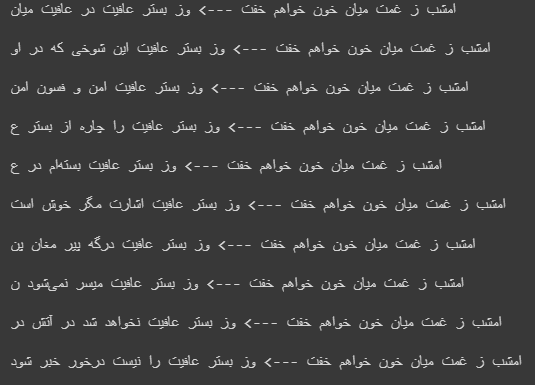
خط سبز و عارضت را نقش بندان خطا سایبان از عنبر تر

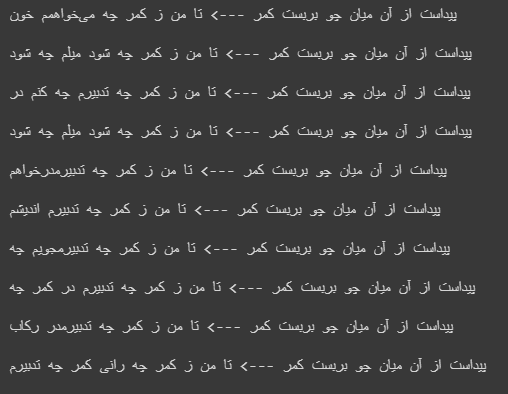
چنین قفس نه سزای چو من خوش الحانیست روم به گلشن رضوان

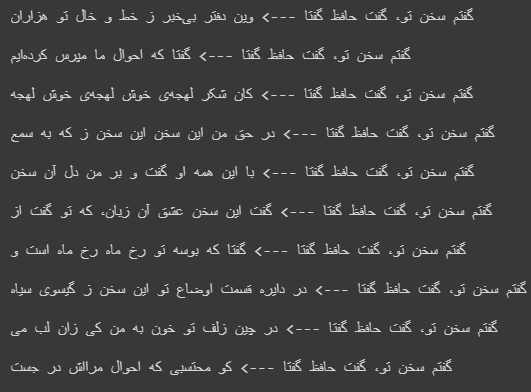
* ***حافظ جی پی تی***

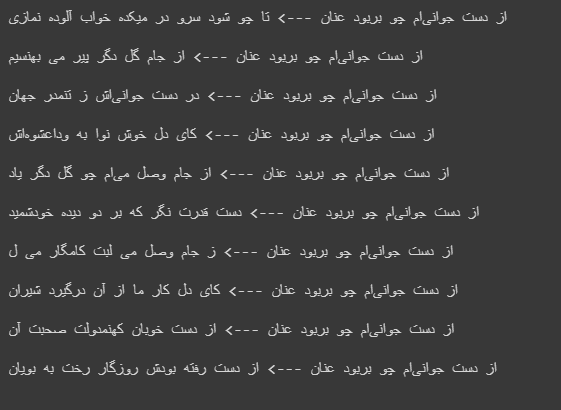
در ادامه تعدادی از نتایج تولید شده برای داده­های تست را می­بینید. در هر تصویر، تعدادی نمونه تولید شده برای یک مصراع را مشاهده می­کنید.

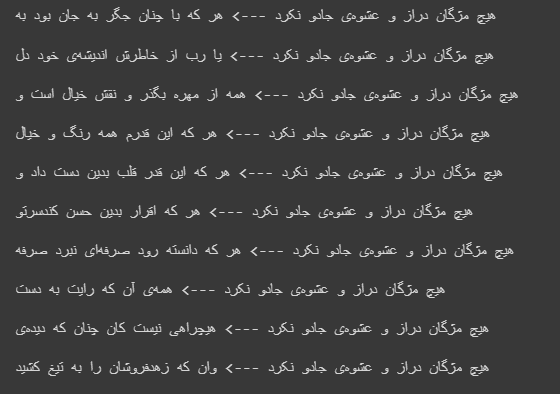


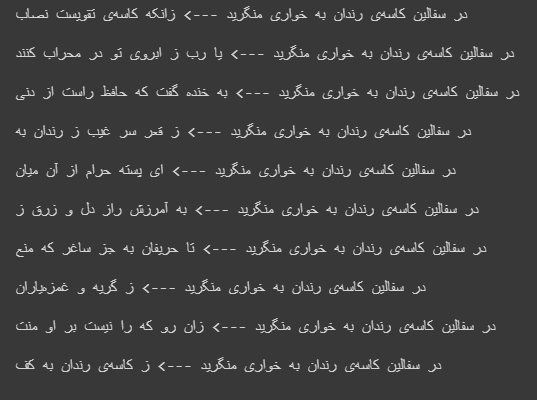












همین­طور که از نتایج مشاهده می­شود، این مدل بهتر از مدل mt5 آموزش دیده و نتایج معقول تری(هر چند باز هم نه چندان مناسب) تولید می­کند. قطعا نمی­توان از این نتایج برداشت کرد که مدل gtp2-fa از mt5 بهتر بوده، و اتفاقا از نظر بنده در صورت بررسی و دیباگ و آموزش درست، هر دو مدل بتوانند به نتایج مناسب و نزدیک هم نیز برسند.

یکی از تکنیک­هایی که در صورت سوال نیز پیشنهاد شده بود و در این تمرین مورد استفاده قرار نگرفت استفاده از اشعار سایر شعرا بود که همان­طور که دیدیم مشکل Overfit شدن مدل را باید تا حدی رفع کند.

بحث آخر در مورد معیار BLEU است. درست است که این مدل­ها نتایج نامطلوبی دارند، اما اگر یک مدل بسیار خوب که اشعار باورپذیری را تولید می­کند نیز داشته باشیم، بعید است معیار BLEU برای ارزیابی آن مناسب باشد و امتیاز بالایی به آن دهد. این معیار paraphrasing و نزدیکی معنایی را اصلا در نظر نمی­گیرد و از نظر بنده معیار مناسبی برای ارزیابی این گونه مدل­ها نیست.

به طور مثال معیار BLEU برای مدل GPT2-fa برابر با 0.05 شد که اصلا مناسب نیست در صورتی که برخی مصراع­های تولید شده شباهت اندکی به اشعار حافظ داشتند.

در انتها باید یک عذرخواهی از حافظ عزیز نیز بکنیم. این تمارین جنبه آموزشی دارد و شما خیلی خودتان را ناراحت نکنید!

**منابع و نوت­بوک­ها:**

Hooshvare GPT-2 finetuning notebook: <https://colab.research.google.com/github/hooshvare/parsgpt/blob/master/notebooks/Persian_Poetry_FineTuning.ipynb#scrollTo=Osc729oi5U1w>

Huggingface Ruby code summarization T5 notebook: <https://colab.research.google.com/github/NielsRogge/Transformers-Tutorials/blob/master/T5/Fine_tune_CodeT5_for_generating_docstrings_from_Ruby_code.ipynb>

How To Make Custom AI-Generated Text With GPT-2 : <https://minimaxir.com/2019/09/howto-gpt2/>

Train a GPT-2 Text-Generating Model w/ GPU (gpt-2-simple lib) : <https://colab.research.google.com/drive/1VLG8e7YSEwypxU-noRNhsv5dW4NfTGce>

GPT-2 Neural Network Poetry : <https://gwern.net/gpt-2>

Generating an Edgar Allan Poe-Styled Poem Using GPT-2 : <https://scottmduda.medium.com/generating-an-edgar-allen-poe-styled-poem-using-gpt-2-289801ded82c>

<https://reyfarhan.com/posts/easy-gpt2-finetuning-huggingface/>

<https://towardsdatascience.com/fine-tune-a-non-english-gpt-2-model-with-huggingface-9acc2dc7635b>

**نوت­بوک­های این تمرین:**

MT5 GHD: <https://colab.research.google.com/drive/1dWyJhJIuAIsKY0c9P904Z3Ag-aDZi9JG?usp=sharing>

GPT2-fa GHD: <https://colab.research.google.com/drive/1cjHUpLQFiQTcTr1I2ZLjK0DWVC5bUQlI?usp=sharing>