

Construcción de GPT from scratch con Keras - Fake News Maker

Por: Marcos Cortés Valadez, Cecilia Gómez Castañeda

LIGA A COLAB:

https://colab.research.google.com/drive/1nwQytUHFOyHwMsXt61J2pg3D7jMN_rDC

LIGA A CARPETA DE MODELOS:

https://drive.google.com/drive/folders/17_osohO1VPNtZbJ_iEQvDdWTkigwcd6y?usp=sharing

Introducción

Este trabajo está inspirado en el ejemplo “GPT text generation from scratch with KerasHub” de Keras, la idea es construir un Modelo Preentrenado Generativo que permita generar texto desde un prompt. La adecuación que se realizará al ejemplo es entrenar el modelo con textos de noticias en español. Por lo cual, esperamos que el modelo genere textos que simulen noticias.

Para entrenar el modelo se utilizaron fuentes de noticias, considerando que estas contienen bastante texto escrito en el lenguaje español. Las fuentes que usamos fueron los periódicos en línea de “La Jornada” y “El Economista”, así como el blog de noticias “Sopitas”.

Preparando los datos

La fuentes utilizadas para crear el conjunto de datos de entrenamiento y validación fueron los sitios web del periódico la jornada, el economista y el blog de noticias sopitas. Para cada uno de estos sitios web se utilizaron librerías en python para realizar técnicas de web scraping.

Una vez que se hizo la navegación del contenido de los sitios web se extrajeron los textos, en el caso de la Jornada se descargaron todas las noticias desde diciembre 2024 al 17 de abril de 2025; del periódico el economista se descargaron alrededor de 1000 noticias y en el caso del blog de noticias sopitas se descargaron noticias alrededor de 900 noticias.

En total, reuniendo las fuentes de información, logramos generar un texto en bruto con las siguientes características, incluyendo caracteres extraños, frases propias de la fuente, direcciones web y tags de referencia:

Número total de palabras: 5,856,825

Número de palabras únicas: 267,451

Número de líneas: 248,773

Una vez que aplicamos un proceso de limpieza de datos, generando todo el texto en minúsculas, quitando caracteres extraños que forman parte del formato de las fuentes, así como etiquetas a cuentas de twitter o referencias a páginas web y quitando oraciones con menos de 50 palabras, donde una oración se considera como el texto que se encuentra en puntos, obtuvimos un texto limpio de entrenamientos con la siguientes características:

Número total de palabras: 5,451,069

Número de palabras únicas: 223,296

Número de líneas: 185,639

El modelo

Se utilizó la configuración del modelo propuesto por el ejemplo de keras que a continuación se ilustra:

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, None)	0
token_and_position_embedding (TokenAndPositionEmbedding)	(None, None, 256)	1,312,768
transformer_decoder (TransformerDecoder)	(None, None, 256)	329,085
transformer_decoder_1 (TransformerDecoder)	(None, None, 256)	329,085
dense (Dense)	(None, None, 5000)	1,285,000

Total params: 3,255,938 (12.42 MB)

Trainable params: 3,255,938 (12.42 MB)

Non-trainable params: 0 (0.00 B)

La estructura del modelo se basa en un mecanismo de embedding y múltiples capas de decodificación tipo Transformer, finalizando en una capa densa para la predicción final.

Explicación de las capas del modelo original:

Capa de entrada: Esta capa recibe la secuencia de entrada.

Capa de tokenización y embedding posicional: esta capa realiza dos funciones fundamentales para el procesamiento de secuencias:

- **Embedding de Tokens:** Convierte cada token de la secuencia de entrada en un vector denso de dimensión 256. Esto permite al modelo capturar las relaciones semánticas entre los tokens. El número de parámetros asociados a esta parte dependerá del tamaño del vocabulario de entrada.
- **Embedding Posicional:** Agrega información sobre la posición de cada token dentro de la secuencia. Esto permite que el modelo pueda entender el orden de las palabras.

Capas decoder: Se crean N capas de decodificadores Transformer, su función principal es procesar la secuencia de embeddings para generar una representación contextualizada de cada token. Se presume que apilar múltiples capas de Transformer permite al modelo aprender representaciones jerárquicas más complejas de la secuencia de entrada. Cada capa sucesiva refina la representación aprendida por la capa anterior, permitiendo capturar dependencias más abstractas y de largo alcance dentro de la secuencia.

Finalmente se aplica una capa densa con la cual se realiza una clasificación sobre el vocabulario generado para cada posición de la secuencia de salida.

Ejecución de pruebas

EJEMPLO CON 5 ÉPOCAS SIN PROMPT INICIAL

Top-K search generated text:

['[BOS] el reporte de las ong sentencia que las acciones de la ue para implementar la global gateway corren el riesgo de contradecir sus propios compromisos de mantener altos estándares de derechos humanos , sociales y laborales , transparencia , crear asociaciones entre iguales en lugar de dependencias y ofrecer una agenda de inversión democrática (como resultado de priorizar los grandes negocios , casi todos los países de la ue han realizado recortes a la ayuda al exterior 2025 - 28 para mejor finan']

EJEMPLO CON 200 ÉPOCAS SIN PROMPT INICIAL

Top-P search generated text:

['[BOS] , la inestabilidad que inició con el segundo mandato de donald trump en la presidencia de estados unidos y la guerra comercial que desató al imponer aranceles ya tuvo efectos en baja california , donde se reportan 17 mil empleos menos , entre los perdidos y los no creados por la espera en que se encuentran 10 empresas que decidieron suspender la ampliación de sus plantas , en una entidad donde 28 por ciento de la población económicamente activa (pea) vive de la manufactura , expuso el secretario de economía de']

Top-k search generated text:

```
['[BOS] a fin de atender los reclamos de la coordinadora nacional de
trabajadores de la educación ( cnte ) , el gobierno federal propone la
emisión del decreto y en las consideraciones advierte que la decisión se
apega a la recomendación 162 de la organización internacional del trabajo
, la cual plantea a los estados adoptar medidas para flexibilizar la edad
para las prestaciones de vejez , de acuerdo con las actividades
desempeñadas y la capacidad de las personas de edad avanzada , teniendo
también']
```

EJEMPLO 100 ÉPOCAS CON PROMPT INICIAL

Ejemplo de resultado con un modelo formado por 100 épocas y estableciendo una frase de inicio para la generación de texto, en este caso el prompt utilizado fue la frase: “El fenómeno”, a continuación se muestra el resultado utilizando dos métodos de búsqueda: Top - k Sampler y Top P sampler, ambos utilizados en el ejemplo original de Keras.

Top P Sampler

Se ilustran 10 pruebas aumentando el valor de P, en cada iteración se espera que la selección de token siguiente sea más rígida ya que se busca el token que tenga una mayor probabilidad.

```
p=0.1      ['[BOS] el fenómeno de la niña , que se acreditó de que el
gobierno de la ciudad de México , en el que se comprometió a que se le
acusa de arrestar a la compañera de la ciudad de México , en el que se
aprovechó a la compañera de los patrones de la candidatura de la ciudad de
México , y el periódico de la ciudad de México . [PAD] . [PAD] . [PAD] .
[PAD] de la ciudad de México . [PAD] de México ( 2 ) , en el primer lugar
de la']
```

```
p=0.2      ['[BOS] el fenómeno de la violencia que permaneció en el
estado , y la fiscalía general de la república ( fgr ) , y el gobernador
de Morelos , en el estado de México , en el que se permaneció en el centro
de la ciudad de México , en el estado de México . [PAD] , en el municipio
de San Juan de San Juan de la ciudad de México . [PAD] . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] . [PAD] .']
```

```
p=0.3      ['[BOS] el fenómeno de la niña en el que la fallecimiento de
la inseguridad , la filial de la candidatura de la corriente de los
peritos , que describió el Barcelona , la desconocida por la desconocida
de su filme de la Champions , y comunista . [PAD] de la atmósfera . [PAD]
```

de brav , que habilita la historia de la magia del tokio . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] . [PAD] . [PAD]']

p=0.4 ['[BOS] el fenómeno de la niña , que es una distinta que se
contempla la inscripción de la prestación de servicios de salud y
servicios de salud , que en el caso de las personas con discapacidad , a
los trabajadores y sus corridos , por lo que la persona que se perciben un
mayor trabajo . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] debut .
[PAD] , chara , de kankistán , chuapan , en la plal spiniko . [PAD] .
[PAD] , de ken kuenz']

p=0.5 ['[BOS] el fenómeno de la violencia en méxico es una muestra
de indagatoria , que se describieron en el motor de la violencia , la de
las mujeres y la niña , así como los hombres que ha hecho un compañero ,
un vínculo de él . [PAD] . [PAD] que tuvo un robar a la voz de los
habitantes de la tierra . [PAD] , que le deseaba . [PAD] . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] , en un momento fue un distin (meday , may , alus
tor (torona) de tor']

p=0.6 ['[BOS] el fenómeno de la violencia , en la región de los
ángeles , en el municipio de nezahuacán , dijo que en el evento , por la
mañana se detuvo a dos personas , las cuales fue desechos de la zona de
australia , se iniciaron en el aeropuerto internacional de la zona norte
de chiapas , con el fin de semana pasado . [PAD] . [PAD] . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] de la zona de la cruz azul , la caltaña del moda (
world . [PAD] (ki) , carbaum (2 - 2) .']

p=0.7 ['[BOS] el fenómeno de la pérdida de agua (unex) de las
bandas y se incrementó un espacio de la acumulación de los recursos del
invierno , pero no sólo a la batería de servicios básicos del agua , así
como la progenidad de las limnas , como un existo del vínculo y la tesis
delictiva en el museo . [PAD] . [PAD] de la cultura perpetrógeno en el
cedral en un contexto de automóvil , así como el gran premio de las
naciones , en méxico y no se sofis']

p=0.8 ['[BOS] el fenómeno de taquiler de extrajenos erróneamente
como una zona de la tierra , la aprecia desacaa , es un madero más , que
los agricultores de santander , tringue , sin una prestigio de una cirugía
como una mujer individual y les pague la instalada en la ciudad de la
dictadurante en un campo de a la piel . [PAD] . [PAD] abra de burdal , de
la cantata , chelyer']

p=0.9 ['[BOS] el fenómeno de tecrimo construido comprobado : no sólo
imaginarse con espectro y viajar mal que , hace unos días también es como
chor , también conocidos y traen virahu , se dice la obtiene . [PAD] baus
sensky que le prepore la vida y el gran amor arro de josé mochindlo , la
fe , y él es imposible que es una mandó y de diana martínez , montez de su
amigo liverpo del']

p=1.0 '['[BOS] el fenómeno de dificultades para personas que les da cuenta con un aumento de 70 por ciento , por medio en algún nivel de interés , aunque el árbol de mil mensuales de otros) , y por lo que pilaron de las bajas de la seguridad hacia las empresas que se ponga en el sector . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] y del norte (fy kon) . [PAD] de china (kamie . [PAD]) . [PAD] refirió que conversa escalaían los peor daños porque carter gavuelga hizo un cuarto y fuese del cimo'

Top K Sampler

Se ilustran 10 pruebas aumentando el valor de K, donde se espera tener más variedad de selección del toque siguiente muestra la K incrementa.

k=1 '['[BOS] el fenómeno de la niña , que se desconoce , en el que se comprometieron a la comunicación de la comunitaria , la comunitaria de la comunitaria , la descripción de la comunitaria de la cultura , la ciencia y la cultura (unesco) . [PAD] (universo) . [PAD] de la ciudad de méxico (198) . [PAD] (198) , el 19 de abril de 2025 . [PAD] de 2026 . [PAD] (197) . [PAD] (198) . [PAD] (1994']

k=2 '['[BOS] el fenómeno de las desapariciones de la región , que se ha desplegado a la población de la ciudad de méxico , que se apreció a los trabajadores de la educación , la ciencia y el desarrollo de la ciencia , la cultura , la ciencia y la cultura (unesco) , en el desarrollo de la ciencia y el desarrollo tecnológico de méxico (unigh) . [PAD] (lactur) , la unam , en la unam , de la universidad nacional autónoma de méxico (unam) , la unam , la unam , el colegio de méxico (unam) , de la universidad autónoma de méxico ,']

k=3 '['[BOS] el fenómeno de la violencia , que se ha descubren en el país , en la que se ha convertido en una de las últimas horas de la casa blanca , en la que se ha demostrado en un país que se prepara el control de la frontera norte , que ha estado a los cárteles de la droga , la droga , y la droga . [PAD] de la droga . [PAD] . [PAD] . [PAD] . [PAD] , el caphuapan . [PAD] , la bob , el bonetror , y la brita de la brazo , lanzar un gran']

k=4 '['[BOS] el fenómeno de la humanidad es un desafío de la humanidad , que ha sido un país , que ha hecho de que el mundo no tiene un país de origen , sino también una de sus principales índices de la violencia de género , y la de las mujeres , y los hombres , de todo el país . [PAD] . [PAD] . [PAD] es una de las más importantes del mundo . [PAD] . [PAD] . [PAD] (en méxico) , en la ciudad de méxico y el caribeño de la ciudad de méxico , en la que se compara con los mexicanos , en el país . [PAD] (seguina) . [PAD] de la']

k=5 ['[BOS] el fenómeno de los incendios , en el centro de la ciudad de méxico , que en el municipio de villahermosa , se ha calzado a una altura de la calle de chicago . [PAD] , en la zona centro de la ciudad de méxico , informó el secretario del tesoro , juan ramón de la fuente y el titular de la unidad nacional de la unidad de protección ciudadana (ss) , de las mujeres en la ciudad de méxico . [PAD] . [PAD] , en el sur , la capital del estado de méxico (méxico) . [PAD] de guadalajara , en el que se compece la tra']

k=6 ['[BOS] el fenómeno de la violencia que no es de una manera de los pueblos , sino a los pueblos indígenas y afromexicanos , que aprehenquemos , descartaron , chips en su mayoría de edad , a los acomocratificantes de las víctimas . [PAD] . [PAD] . [PAD] . [PAD] debilic . [PAD] y de la preca . [PAD] , el manejaba , y la caminata . [PAD] maz bavi ,orros . [PAD] de la corona . [PAD] . [PAD] (ch']

k=7 ['[BOS] el fenómeno de la niña , es una de las más que en el traductor de que el tienzo , el matrimonio de la desconoce a los pabels de bars y el aglomer , y que no cambió como el bracking en el sprendente de los mannes , adjuntan , en un audito a su familia alternativo en las que los agremiados en un rodurte en el aglomerado de la música en el centro del']

k=8 ['[BOS] el fenómeno de que se cometa a las preocupaciones que están cerrando la responsabilidad del pueblo y la justicia en la nación , de la mandataria claudia sheinbaum , por la tarde , la mandataria federal ha sido absulado . [PAD] a la embajada de estados unidos en la entidad . [PAD] . [PAD] despojo , el salvador aguiz aguilera y supo . [PAD] . [PAD] . [PAD] . [PAD] de la balla , de chilapa y en méxico . [PAD] . [PAD] . [PAD] se ap tra . [PAD] (eln) y descarril de la']

k=9 ['[BOS] el fenómeno de la niña , la franquicia de los jóvenes , el antimigrante , que descubría y traficaz , la genética y de la plástico transportaban que son los resurridos en el que los desperdicios y la padrón , que se desatengamos , a la que los traficantes son de su país , pero a las personas que no se complican su trabajo . [PAD] , la ama a los cortaús de sufrimiento de']

k=10 ['[BOS] el fenómeno de que la magdalena de fanster , que es el piloto combinado de la montaña , de acuerdo con los reportes de los chelinos , que en los preciban la inspección de la trama y la corporación del agua de la pluma de la presa , en un contexto más alto desde que se influencia del cocineros y en la capital de los municipios . [PAD] con chenúl . [PAD] de las bulinas en la zona del centro de t']

EJEMPLO 200 ÉPOCAS CON PROMPT INICIAL

Ejemplo de resultado, corriendo 200 épocas y estableciendo una frase de inicio para la generación de texto, en este caso el prompt utilizado fue la frase: “El fenómeno”, a continuación se muestra el resultado utilizando dos métodos de búsqueda: Top - k Sampler y Top P sampler, ambos utilizado en el ejemplo original de Keras.

Top P Sampler

Se ilustran 10 pruebas aumentando el valor de P, en cada iteración se espera que la selección de token siguiente sea más rígida ya que se busca el token que tenga una mayor probabilidad.

p=0.1 ['[BOS] el fenómeno de la niña de 13 años de edad , que se encuentra en el hospital general regional número 40 , que se han registrado en el torneo local , donde se disputará con 3 mil 946 unidades de medida y actualización (uma) , que se han registrado en el torneo local . [PAD]ices en el estadio cuauhtémoc . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD]dos afectados . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD] . [PAD]s . [PAD]s . [PAD]s .']

p=0.2 ['[BOS] el fenómeno de la niña de 13 años de edad , que se encuentra en la ciudad de méxico , donde se comprometió a respetar el número de personas en situación de pobreza , pues no se han convertido en un problema , sino que también acreditan la cotización de la población . [PAD] serias . [PAD]ices yorksrrarados presentadas . [PAD] : . [PAD] : meses . [PAD]rarrarrarrarrarrarrarrarrarrarja perez :gers : algún viajes :rarm : algún periodo dem . [PAD] :rar :rar años de sugersduras']

p=0.3 ['[BOS] el fenómeno de la niña de 13 años de edad , se encontró con un poco más de dos minutos , en el capitolio , en el que se incorporó a la mitad de las canchas de la especie en el escenario . [PAD] adeduras , como presidente de estados como en el zócalo capitalinos . [PAD] . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s : perez perez perez perez perez perez perez perez perez perez . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD] . [PAD]']

p=0.4 ['[BOS] el fenómeno de desaparición forzada (ced) , la
secretaría de relaciones exteriores de la defensa nacional de la
comisionada por la desaparición de un ciudadano colombiano , quien fue
detenido acusamente sexual y un homicidio con el fin de semana . [PAD]s de
su puesta en escenas . [PAD] que acudieron personajes . [PAD] : . [PAD] :
algún periodo que perez que acudieron que se encuentran en fiestas .
[PAD]dos buscannciones : el legados . [PAD]méricas : puedan acceder ,
perez perez perez perez perez león . [PAD] :gers . [PAD] : algunos dem .
[PAD] : el legado : fiestas :']

p=0.5 ['[BOS] el fenómeno de análisis detallado de la reunión con el
secretario de comercio , hograma , mario vargas llosa colga , informó que
se harán más de 300 asistentes y el intercambio de permanecer en los
hechos históricos de la migración y el flujo de migrantes . [PAD] que
acudieron diego en el parque de xaso necesario . [PAD]rarlos . [PAD]rar
acompañadas de x infaméticosrardo :rarias que acudieron que rubios
regresarles puedan regresarles . [PAD] de visitantes . [PAD] . [PAD]
ubicados de plaza : ubicados . [PAD] :rarger suger rubiosgers']

p=0.6 ['[BOS] el fenómeno de meteorológico que debilitará un incendio
en el basurero de santa maría coixavidad , debido a que la mujer está
concurrido de violencia en el estado , y cuando se le cale , a los niños
de la delincuencia organizada . [PAD] que acudieron x a tijuana en la
plaza acudieron algunas . [PAD] correspondientes : . [PAD] de funcionarios
de su parque de x horas en la plaza regresar con subjeción . [PAD] . [PAD]
:rartar suger oramos algún incofaméticosrardo que perez :gers , acudieron
que encogers']

p=0.7 ['[BOS] el fenómeno de la niña se encontraba de un hombre , que
en los jóvenes , en el camión de las montañas , en aceras , con un hombre
y los reyes cabrera y abuela , barrativos , emocionales y tienda de alta
depítrico , y tos se encuentre la sierra de berragán . [PAD]randroles al
legadosburias que acudieron x su parque en basuras . [PAD] encaciones
puedancacioness puedan regresarles puedan regresar su importancia de
visitantes . [PAD] alguna']

p=0.8 ['[BOS] el fenómeno migratorio que será la nueva administración
de trump , y al repetir el acuerdo comercial bilateral , en mocte , en el
mismo periodo del gobierno federal , el que le acusaron la decisión de la
comisión de relaciones exteriores , la cual tiene fecha produjo opciones a
las personas rusas . [PAD] reuniones . [PAD]das : fiestales puedan ,
asistir a la partura de x : cambiar acá : salir hoteles y no están
llegandos . [PAD]dos . [PAD]s acudieron a deportados . [PAD] simmbrados :
fiestales , ubicados ? : expresados . [PAD] perezs']

p=0.9 ['[BOS] el fenómeno tumbado y , porque hacemos una temporada de lluvias y conflictivas las operaciones de estabilizar desapariciones , o amplíruen desaparecer en todos los niveles que les dejaron poco muertos y el resto son dolor . [PAD] . [PAD]dos afectados . [PAD] ques como aseguradorass . [PAD] . [PAD]rse . [PAD]gers , carmen castillos . [PAD] : algunos que se puedan acceder si monreal : algún vigs dentro , aplicarles . [PAD] : gentes . [PAD] ? pérezs . [PAD]ger rechadorasm pérez pérez pérez pérez león . [PAD]gers :']

p=1.0 ['[BOS] el fenómeno meteorológico conocido gradualidaddes predicción de un espíritu ágiles y que se vea la cantante , inundaciones técnicas , como la carranza poquito sigue emergencias desarrollón con un entorno indistivo y personal de hospitales para garantizar impunidad en los casos de cáncer y discapacidad , la próspero en el campo . [PAD]r vaticano rubitivas . [PAD]bles que acompañadasción ? . [PAD] pérez serias ventales : infasción de algún personal de argentina , figura trancaciones mesesbeciones endeles']

Top K Sampler

Se ilustran 10 pruebas aumentando el valor de K, donde se espera tener más variedad de selección del toque siguiente muestra la K incrementa.

=1 ['[BOS] el fenómeno de la niña de 13 años de edad , que se encuentra en el hospital general regional número 40 , que se les ha convertido en un centenar de generaciones en la población de la población en situación de pobreza , y en particular en el sector de la población . [PAD]lo . [PAD] alias . [PAD] : . [PAD] pérez : . [PAD] pérez pérez pérez pérez pérez pérez pérez pérez pérez que acudieron afectados . [PAD] afectados . [PAD] afectados . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s . [PAD]s : algún xilom pérez pérez']

k=2 ['[BOS] el fenómeno que se describió como un año , la compleja de los años que desacelerará el crecimiento mundial y el crecimiento económico mundial seguirá siendo la diferencia de los países de la ue , señaló el ex campeón mundial de estados unidos y descenso en el año pasado . [PAD] adm torno : miles de turistas que competirán en países de españa . [PAD]rar . [PAD] gente . [PAD] :rar suger importancia enco vaticanos . [PAD]s que puedan regresar futurosdurasdurasduras deculación :rarrarrsen x mismo : fiestas . [PAD]dos en']

k=3 ['[BOS] el fenómeno de desaparición forzada de personas unidas advirtió que la situación de desapariciones forzadas , responsable de desaparición y de manera inmediata la detención de quienes se han realizado en un domicilio en la unidad de política migratoria de la policía federal pasada investigación , estatal y detenciones de personas que se despedó . [PAD] adm torno a la puesta en elizar el desempeño . [PAD] su legado :rar que acudieron funcionarios de su secretario de x .

turistas de acceso . [PAD] puedan acceder acorarias . [PAD] . [PAD] :
algunos afectados . [PAD] : castillos . [PAD] pérezs . [PAD] xgers']

k=9 ['[BOS] el fenómeno de que las personas que son defendieron los
ministros de la suprema corte de justicia de la nación (scjn) , xóchitl
impulsan la innovación de la reforma al poder judicial y de la voluntad
política del gobierno de estados unidos y de los pueblos indígenas , así
como afromexica jurídicati el derecho a la justicia . [PAD]r sudro de su
situación futuro : . [PAD] agud ar : su presencia rubios . [PAD] rbergles
: algún recintos de los funcionarios de las que puedan regresar agrarias
engens , periodos . [PAD] : acompañada']

k=10 ['[BOS] el fenómeno de abuso de drogas de avenidas y desafío
retomarlas y a través del sistema premisaitario , la descentuación de la
carretera que se deje de la zona . [PAD] el parque juárez de san trinidad
tas , vacaciones . [PAD]das . [PAD]gers : algún periodo que pérez " ,
pérez león " . [PAD] pérez pérez "dura lópez . [PAD] . [PAD] . [PAD] .
[PAD] . [PAD] . [PAD] . [PAD] congersduras : algún periodos :gers :
visitas : xgers : x york : : miles de años ? :ger reyes de alguna gente']

Conclusiones

Este proyecto representó una serie de desafíos, iniciando por decidir cual sería un buen conjunto de información que nos sirviera para entrenar un modelo con propósito de generación de texto, decidimos entrenar con noticias porque representan un vocabulario con un discurso tipo reporte y pensamos que eso puede ayudar a focalizar el contexto del modelo, además que gracias a las fuentes disponibles por internet podemos descargar un volumen considerable de datos. El siguiente reto fue representado por realizar procesamiento de limpieza sobre las diferentes fuentes.

Entender la aplicación del modelo y hacer pruebas buscando tener mejor resultados, nos ayudó a tener mayor sensibilidad sobre el efecto que realiza cada una de las capas del modelo, así como decidir una mejor selección de hiper parámetros.

Sobre los resultado obtenidos, aunque nos sorprende la mejora en el resultado bajo la ejecución de cada época, pensamos que también se puede generar un problema de sobreentrenamientos ya que nos dimos cuenta que el algunos escenario el modelo hizo copia exactos del texto de noticias.

Finalmente, podemos decir que algunos textos obtenidos en pruebas contiene una narrativa conforme a la redacción de noticias, sin embargo, considerando los siguientes pasos para mejorar el modelo:

1. Crear un conjunto de datos para entrenamiento del modelo más grande, porque el conjunto de datos utilizado en el ejemplo original cuenta con 75 millones de palabras

repetidas, consideramos dadas las pruebas que realizamos con un conjunto de datos menor al final , que mejora significativamente la generación de texto aumentando el modelo de entrenamiento.

2. Sumar una capa del modelo para que mejore las conclusiones de una noticia, porque la mayoría de noticias generadas carecían de conclusiones, por lo que pensamos que se puede entrenar un modelos que considere el inicio y fin de una conclusión,
3. Realizar un flujo de trabajo en el cual pueda ser posible ingresar trompo por voz, hacer la interpretación a texto, generar una noticia y finalmente regresar la noticia con voz.