**DOCUMENTO COMO PONER EL MODELO DEL EJERCICIO EN PRODUCCIÓN**

Para colocar el modelo del ejercicio 1 en producción se deben utilizar los siguientes servicios dentro de la GCP para poner a funcionar.

1. Front End que va a cargar el documento dentro de un storage.
2. Google Cloud Storage.
3. BigQuery.
4. Cloud Functions.
5. Vertex AI
6. App Engine,Cloud Build, Cloud Run (Docker Python 3.8)
7. Permisos de IAM para cuenta de servicio en cloud storage job editor, bigquery get jobs, bigquery editor.
8. Cloud Scheduler ( Programar re-entrenamiento del modelo)
9. Document AI para extraer textos de pdfs o imagenes.
10. Looker Studio.

Para desplegarlo en producción los pasos de funcionamiento serían los siguientes:

1. Primero el front end del modelo va a capturar el archivo o documento que va a ingresar a un cloud storage
2. El archivo se va a ingresar en un bucket del cloud storage , acto seguido se dispara una cloud function.
3. La cloud function es la encargada de disparar apenas se inserta el archivo dentro del bucket , va a llamar a la API de cloud Run que va a tener todo el contenedor , como parámetros le va a pasar el nombre del archivo.
4. El cloud run con el nombre docker en python y todo el modelo que esté entrenado previamente dentro de los scripts en máquinas virtuales de vertex AI para el procesamiento de grandes volúmenes de información. Este cloud Run disponibiliza un API que se crea en Flask de Python y responde al front con el tema al cual pertenece el documento , el front recibiría la respuesta .
5. Adicionalmente el docker en cloud run que utiliza librerías de bigquery para extraer información de los textos armaría un respaldo de los datos estructurados como una tabla para , donde se realizaría el almacenamiento del archivo el texto y la categoría a la cual pertenece.
6. Vertex AI se utilizan diferentes librerías como las utilizadas en este ejercicio y frameworks de modelos de topics clasification para estructurar los textos a un tema. Una vez esté el modelo construido se empaqueta para ser utilizado en el docker.
7. Se Crearia un Pipeline en Vertex AI para reentrenar el modelo con la información que vamos obteniendo almacenada en Bigquery, cada cierto tiempo y que funcione con un cloud scheduler para que cada cierto tiempo reentrene el modelo de manera automática.
8. Looker studio mostraría métricas de rendimiento del modelo datos descriptivos que indicarán la clasificación de manera adecuada.
9. En caso de requerirse mayor precisión dentro del modelo utilizaremos AutoML de GCP para generar un modelo con temas de clasificación muchos mas precisos por utilizar algoritmos de precisión.
10. StackDriver de la GCP o Logging para monitorear los errores que puedan existir dentro de la plataforma con todos los servicios que se disponen.