**Prueba técnica**

**Teoría:**

1. En la empresa GA, en el área de compras necesitan CLASIFICAR y organizar los correos que llegan a la bandeja de entrada entre 4 tipos de correos (Compras cementos, Compras energía, Compras concretos y correos generales o de otra índole). Esta tarea se le encomienda a usted, gracias a su rol puede solicitar al área interesada los recursos humanos que necesite para llevar a cabo este proyecto, también puede solicitar en tecnología todo lo que necesite, además tiene las bandejas de entrada de correos históricos de los analistas que reciben estas solicitudes con aproximadamente: 5500 correos de compras cementos, 2700 correos de compras de energía, 1100 correos de compras concretos y 12876 correos generales o de otra índole.

**R/**

**Entendimiento del Problema**

El objetivo es crear sistema eficiente que permita clasificar automáticamente los correos electrónicos entrantes de acuerdo a las cuatro categorías predefinidas con un sistema que debe ser capaz de procesar un gran volumen de correos y asegurar una alta precisión a la hora de clasificación.

**Metodología**

1. **Tipo de trabajo:** Se estructuraría el proyecto operando bajo una metodología Agíl, dado que, nos permite iterar sobre los inconvenientes que vayan surgiendo y mejorar en pequeños periodos de tiempo el proyecto.
2. **Recolección de datos y almacenamiento:** Se recopilarán todos los correos electrónicos históricos de los analistas y se manejaría como base de datos para el almacenamiento una NoSQL para nuestro caso recomendamos MongoDB para mejorar el procesamiento de la consulta.

* Otra alternativa puede ser almacenarlas en Azure y generar como trigger power automate.

1. **Análisis exploratorio y limpieza de datos:** Revisar las estructuras de datos de los analistas y realizar una limpieza de los datos para eliminar duplicados, correos dañados, quitar caracteres inválidos y normalizar la información.
2. **Definición de regla de negocio:** Se organizan grupos de trabajo con los expertos del área de compras para poder definir el conjunto de reglas de negocio que se deben aplicar a cada uno de los correos cuando estos lleguen, basándose en el conocimiento y la experticia de cada uno de los expertos del área.
3. **Definición e implementación del modelo de aprendizaje automático:** Se utilizaría un modelo supervisado, dado que tenemos las reglas de negocios predefinidas y es fácil de realizar evaluaciones e intervenciones al mismo.
4. **Evaluación y optimización del proyecto:** Se aplicarán métricas como precisión, recall y F1-score: Estas métricas nos permiten ver la exactitud y.

* Ajuste de hyperparametros, nos permitirán mejorar como opera el algoritmo.

1. **Despliegue :** Una vez que el modelo esté entrenado y validado, lo lanzaremos al entorno productivo y revisaremos su comportamiento.
2. **Mantenimiento y mejora continua:** Se realizará una evaluación periódica del sistema para medir su desempeño y detectar posibles problemas.

Se implementará un mecanismo de retroalimentación con periodicidad mensual para que los usuarios puedan corregir las clasificaciones incorrectas y mejorar así el modelo.

2. Seis meses después de haber desplegado un modelo de regresión en producción, los usuarios se dan cuenta que las predicciones que este está dando no son tan acertadas, se le encarga a usted que revise que puede estar sucediendo.

**¿Cree que el modelo esté sufriendo Drift?**

**¿Cómo puede validarlo?**

**¿De ser así, que haría usted para corregir esto?**

**Explique sus respuestas.**

3.Su equipo de trabajo está trabajando en un chatbot con generación de texto utilizando el modelo GPT-3.5, según cómo funciona este modelo, ¿cómo haría usted para hacer que las respuestas del chatbot estén siempre relacionadas a conseguir cierta información particular del usuario y no empiece a generar texto aleatorio sobre cualquier tema?

**Explique su respuesta.**