

## Prueba técnica Gestor SR con énfasis en IA.

### 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, como es el Gestor SR en temas de analítica e IA 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.

**Explique como resolvería este problema, metodología, algoritmos, modelos, arquitectura del proyecto etc.**

Para empezar, se debe estructurar esto como un proyecto, con su respectiva metodología, repositorios, ambientes y miembros.

Procedemos al entendimiento del negocio, objetivos y posibles soluciones. así como de los datos, para poder identificar que datos sirven, y como se deben obtener.

Una vez identificados las fuentes de los datos y los datos, se comienza con un proceso de guardado de la información de tal suerte que tenga una estructura adecuada para empezar. Se puede hacer uso de APIs de lecturas de correos, o si esta en un correo de microsoft hacer uso de Azure, usando por ejemplo Graph, o en su defecto un RPA.

Junto a esto se define el espacio donde se van a almacenar y si los datos serán estructurados o no estructurados.

Finalmente podemos proceder con un modelo de clasificación de texto empleando modelos de clasificación como KNN, Naive Bayesiano Supervisado o vectores de soporte SVM y random forest, arboles de decisión, redes neuronales e incluso PLN.

Otro recurso a usarse pueden ser los PLN, LLM y se puede auxiliar de librerías como Transformers de HuggingFace con datos ya entrenados, o una elaboración propia de un diccionario con palabras que tienen posibilidad de aparecer en determinado correo y se categorizan.

Obtenido el resultado, se disponibiliza el modelo ya sea con una API, docker o ya en Azure con un procesamiento orquestado desde Data Factory o Databricks, o que se habilite azure function o logic app.

Por último, si es necesario, se construye un front ya sea cómo plataforma web o en tableros de PBI.

1. Caso de negocio
2. Entendimiento del negocio y data
3. Obtener información, limpiarla y almacenarla.
4. Modelación
5. Orchestation, llevar a productivo, disponibilizar el modelo.
6. Exposición o visualización (Si aplica)

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 como Gestor SR en temas de IA 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.**

Si, el modelo puede estar perdiendo predicción dado que esta dejando de lado los eventos y situaciones que vienen ocurriendo en los últimos 6 meses que afectan el comportamiento de la predicción, por tanto, es importante tener un tiempo de reentrenamiento más corto para que estos efectos no sigan ocurriendo. La forma de evidenciar si está o no perdiendo efectividad es calculando las predicciones versus la realidad y calcular el valor de algunas métricas de acierto de predicción y compararlas con las que dieron en la última puesta en producción en su entrenamiento, si es evidente su caída, podemos asumir que el modelo perdió asertividad.

Como lo mencioné anteriormente, una de las formas de asegurar que esto no ocurra, es reentrenando los modelos con menor tiempo, para así posibilitar que los modelos aprendan lo que sucedió recientemente y calculen un pronóstico mucho más semejante a la realidad, puesto que es probable que ocurriesen eventos de datos atípicos, tendencias, estaciones o incluso situaciones geopolíticas.

Si es muy difícil poder lograr que se reentrene por los altos costos, se debería hacer una especie de alerta que muestre el nivel de acierto que va teniendo semana a semana y poder haciendo ajustes que no requieran modelación (por ejemplo, la suma de datos atípicos, etc).

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.**

Según se indica, el chatbot es propio del desarrollo, por ende, se puede personalizar, ajustar los parámetros, mantener prompts específicos y dando contexto persistentemente.

Por último, se puede hacer fine tuning o Few-shot Learning