**Prueba técnica: Ingeniero de Desarrollo**

Sección 1: Preguntas de conceptos básicos

1. ¿Cuál es la diferencia entre Aprendizaje Supervisado y Aprendizaje No Supervisado en Machine Learning?

La diferencia radica en el tipo de datos y objetivo del modelo, mientras que en el aprendizaje supervisado el conjunto de datos de entrenamiento contiene etiquetas y su objetivo es aprender la relación entre las variables de entrada y la salida para hacer predicciones en datos nuevos, en el aprendizaje no supervisado el conjunto de datos solo contiene las características de entrada sin ningún tipo de etiqueta o valores predefinidos, adicionalmente, el objetivo del aprendizaje no supervisado es encontrar patrones o agrupaciones dentro de los datos.

1. Explica brevemente qué es la regularización en el entrenamiento de modelos de IA y para qué sirve.‬

La regularización es una técnica utilizada para evitar el sobreajuste en los modelos, el objetivo es penalizar la complejidad del modelo, funciona introduciendo un término de penalización en la función de pérdida del modelo durante su entrenamiento. Tratando de evitar que el modelo aprenda patrones demasiado específicos o complejos lo que ayuda a mejorar la generalización del modelo sobre datos no vistos. Existen diferentes tipos de regularización como: regresión lasso, regresión rigde, elastic net, early stopping, entre otros.

1. Menciona por lo menos tres métricas de evaluación comúnmente utilizadas en‬ problemas de clasificación y explica en qué consiste cada una.‬

* Accuracy: Mide cuantitativamente cuántas predicciones fueron correctas.
* Precision: Proporción de verdaderos positivos sobre todas las predicciones positivas.
* Recall: Se define como la capacidad del clasificar de encontrar todas las muestras positivas.
* F1-Score: Es el promedio ponderado entre la precisión y el recall.
* AUC-ROC: Mide la capacidad del modelo para distinguir entre clases.

1. Explica en qué consiste la “Extracción de Información” (Information Extraction) a partir‬ de un texto y menciona al menos dos técnicas o enfoques utilizados en procesos de‬ Text Mining para identificar y extraer datos relevantes. Describa posibles aplicaciones o‬ casos de uso en la vida real de manera breve.

La Extracción de Información es un proceso de análisis de textos, que permite identificar, extraer y estructurar información de grandes cantidades de texto no estructurado. Dentro de las técnicas de Extracción de Información se encuentra: NER, Extracción de Relaciones, Extracción de Palabras Clave, Frecuencia de Palabras, Colocación, Concordancia, entre otras. Las posibles aplicaciones podrían estar orientas en diversas industrias como lo es el caso de análisis de currículums (Extracción de Nombres, Experiencia Laboral y otros), análisis de medios de comunicación (Extraer información de artículos por tanto facilita el monitoreo de las tendencias), implementación de modelo en sistemas de salud (Identificación de fórmulas médicas, tipo medicamentos prescritos).

1. ¿Cuál es la diferencia entre un entorno de desarrollo local y un pipeline de CI/CD‬ (Integración Continua/Despliegue Continuo) en un proyecto de IA?‬

En un entorno de desarrollo local, las tareas como obtención, carga, limpieza, procesamiento de datos y desarrollo del modelo se ejecutan en el computador del desarrollador. Esto implica una limitación de capacidad de cómputo (memoria RAM, CPU, GPU, disco, entre otras), por lo que si se trabaja con modelos complejos, la creación de un modelo podría tardar días o semanas. Además, solo estaría disponible de manera local.

Por otro lado, cuando se involucran procesos de CI/CD (Integración Continua/Despliegue Continuo), se habla de desplegar código en servidores, generalmente en ambientes distribuidos, los cuales cuentan con grandes capacidades de cómputo. Los beneficios de CI incluyen que muchas personas pueden trabajar en el código de un mismo proyecto colaborando en un código base compartido, mezclando código generalmente usando un sistema de control de versiones como Git, y adicionalmente automatizando la ejecución de pruebas y builds de la aplicación. CD ofrece otros beneficios como el despliegue automático del modelo en un ambiente de producción, de esta manera estará disponible para los usuarios

1. ¿Cómo se puede integrar un modelo de lenguaje a gran escala (LLM) de código abierto‬ en un servidor o contenedor? Explique brevemente qué es un LLM, describa un posible‬ proceso de despliegue y mencione algunos de sus casos de uso principales.‬
2. ¿Cómo se puede ejecutar un script o aplicación de manera automática en un entorno‬ controlado dentro de Azure, a partir de un evento desencadenante? Describa el tipo de‬ recurso de Azure que se debe utilizar, su proceso de despliegue y las consideraciones‬ clave a tener en cuenta. En caso de no dominar Azure, indique el proveedor de nube y‬ explique el proceso.
3. Explica en qué consiste la “Extracción de Información” (Information Extraction) a partir‬ de un texto y menciona al menos dos técnicas o enfoques utilizados en procesos de‬ Text Mining para identificar y extraer datos relevantes.‬

La Extracción de Información es un proceso de análisis de textos, que permite identificar, extraer y estructurar información de grandes cantidades de texto no estructurado. Dentro de las técnicas de Extracción de Información se encuentra: NER, Extracción de Relaciones, Extracción de Palabras Clave, Frecuencia de Palabras, Colocación, Concordancia, entre otras.