

Enunciado Prueba - Científico de Datos IA

El proceso de selección consta de dos componentes principales:

1. Componente Teórico:

Evaluaremos conocimientos fundamentales asociados a metodologías de programación, estadística aplicada, y principios de integración de sistemas. Este apartado busca validar la comprensión conceptual y teórica sobre las bases que sustentan el desarrollo e implementación de soluciones de Inteligencia Artificial y análisis de datos.

2. Componente Teórico-Práctico:

En esta etapa se evaluarán habilidades técnicas específicas mediante la resolución de problemas prácticos. Se abordarán temas avanzados de programación en Python, incluyendo manipulación y análisis de datos, protocolos de comunicación mediante APIs, técnicas de procesamiento de lenguaje natural (NLP), y consultas intermedias a avanzadas a bases de datos. Esta fase está diseñada para medir la capacidad de aplicar los conocimientos teóricos en escenarios reales y complejos.

Preguntas de evaluación

A. Preguntas Teóricas (10)

- Explique el concepto de programación orientada a objetos y cómo se aplica en Python.
- ¿Qué es el overfitting en un modelo estadístico o de machine learning? ¿Cómo puede evitarse?
- Defina qué es una API REST y mencione sus principales características.
- Explique la diferencia entre series temporales estacionarias y no estacionarias.
- ¿Qué son las distribuciones de probabilidad y cuál es su importancia en estadística?
- Describa el principio de inyección de dependencias en desarrollo de software.
- ¿Qué es el preprocesamiento de texto en NLP y por qué es importante?
- Explique la diferencia entre un modelo supervisado y uno no supervisado en aprendizaje automático.
- ¿Qué es una consulta SQL JOIN y cuáles son sus tipos principales?
- Explique la diferencia entre métodos síncronos y asíncronos en programación.

B. Preguntas Prácticas

1. Programación en Python (nivel avanzado)

- Escriba una función en Python que reciba una lista de cadenas y devuelva un diccionario con la frecuencia de cada palabra ignorando mayúsculas y signos de puntuación.

- Implemente un decorador en Python que mida y muestre el tiempo de ejecución de una función.
- Dado un dataframe de pandas con columnas 'fecha' y 'ventas', escriba código para calcular el promedio móvil de ventas en una ventana de 7 días.

2. Protocolos de Comunicación (APIs)

- Explique cómo funciona el método HTTP POST y cuándo se utiliza en una API REST.
- En Python, ¿cómo se puede realizar una solicitud GET a una API y manejar una respuesta en formato JSON? Escriba un ejemplo básico.
- Mencione y explique tres buenas prácticas de seguridad que se deben considerar al desarrollar o consumir APIs.

3. Manipulación de Texto (NLP)

- Describa cómo funciona la tokenización en el procesamiento de lenguaje natural y mencione una librería en Python que la implemente.
- Explique qué es el "stemming" y cómo se diferencia del "lemmatization".
- Escriba un fragmento de código en Python que elimine stopwords de un texto dado.
- Explique cómo los modelos generativos (GenAI) como GPT pueden ser utilizados para tareas de NLP, y mencione uno o dos casos de uso prácticos en la industria.

4. Manipulación de Datos y Consultas SQL (nivel intermedio-avanzado)

- Escriba una consulta SQL que utilice una subconsulta correlacionada para obtener, de una tabla 'ventas', los productos cuyo monto de ventas sea superior al promedio mensual de ventas de todos los productos.
- Explique cómo funcionan los índices compuestos en bases de datos y cómo afectan el rendimiento de consultas con múltiples condiciones.
- Escriba una consulta SQL que utilice funciones ventana (window functions) para calcular el ranking mensual de ventas por producto en una tabla 'ventas'.