

1. En ML, ¿los datos son ciudadanos de primera clase?

1 / 1 punto

☒ Sí

☐ No

✓ **Correcto**

¡Solo bien! Los datos son entidades que respaldan todas las operaciones generalmente disponibles para otras entidades.

2. Una canalización de datos es una serie de pasos de procesamiento de datos como:

1 / 1 punto

☒ Recopilación de datos

✓ **Correcto**

¡Camino a seguir! La recopilación de datos es el primer paso en la construcción de sistemas ML.

☒ Ingestión de datos

✓ **Correcto**

¡Justo en el camino! La ingesta de datos es el proceso de absorber datos de diferentes fuentes y transferirlos a un sitio de destino donde se pueden depositar y analizar.

☐ Análisis de los datos

☒ Preparación de datos

✓ **Correcto**

¡Tu lo tienes! La preparación de datos consiste en el formateo de datos, la ingeniería y la extracción de características.

3. ¿La canalización de datos es vital para el éxito del sistema de producción de ML?

1 / 1 punto

☐ No

☒ Sí

☒ **Correcto**

¡Exactamente! Consiste en los pasos increíblemente importantes para el éxito del sistema de producción ML.

4. ¿Qué aplica para maximizar las señales predictivas en sus datos?

1 / 1 punto

- ☐ Formateo de datos
- ☒ Ingeniería de características
- ☐ Cobertura de datos
- ☐ Selección de características

☒ **Correcto**

¡Lo has descubierto! La ingeniería de características es el proceso de utilizar el conocimiento del dominio para extraer características con altos niveles de señal predictiva a partir de datos sin procesar.

5. Sus datos de capacitación deben reflejar la diversidad y el contexto cultural de las personas que los utilizarán. ¿Qué se puede hacer para mitigar los sesgos inherentes en un conjunto de datos determinado?

1 / 1 punto

- ☒ Recopile datos de proporciones iguales de diferentes grupos de usuarios.
- ☐ Comprometerse con la equidad.
- ☐ Adáptese a los datos que cambian continuamente
- ☐ Diseñar mejores características

☒ **Correcto**

¡Excelente! El muestreo equilibrado de diferentes grupos de usuarios ayuda a evitar sesgos inherentes.

6. La mayoría de las veces, los sistemas de ML pueden fallar a los usuarios a los que sirven. En este contexto, ¿qué es el **daño representativo**?

1 / 1 punto

- ☒ La amplificación o reflejo negativo de ciertos estereotipos colectivos.
- ☐ Hacer predicciones y decisiones que impiden que ciertos grupos accedan a recursos u oportunidades.
- ☐ Dar salidas sesgadas con más frecuencia para ciertos grupos de usuarios
- ☐ Inferir vínculos perjudiciales entre ciertos rasgos demográficos y comportamientos de los usuarios.

☒ **Correcto**

¡Buena llamada! Esta es una forma prototípica en que un sistema ML puede fallar a los usuarios a los que sirve.

7. Las etiquetas precisas son necesarias para entrenar adecuadamente a los modelos supervisados. Muchas veces, sujetos humanos conocidos como evaluadores realizan este esfuerzo de etiquetado. ¿Cuáles son las principales categorías de evaluadores humanos? (marque todo lo que corresponda).

1 / 1 punto

☒ generalistas

☒ **Correcto**

¡Buena elección! Los generalistas generalmente provienen de sitios de crowdsourcing.

☒ Expertos en la materia

☒ **Correcto**

¡Bien hecho!. Un ejemplo clásico son los radiólogos que etiquetan imágenes médicas para herramientas de diagnóstico automatizadas.

☒ tus usuarios

☒ **Correcto**

¡Buena elección! Los usuarios pueden proporcionar etiquetas dentro de su aplicación. Un ejemplo clásico es el etiquetado de fotografías.

☐ Registradores

☐ Agregadores

☐ Clasificadores

