

1. Al evaluar un modelo de ML durante el entrenamiento, el objetivo es mejorar las métricas de nivel superior, como la precisión general. Esta información se usa para decidir si el modelo funciona bien o no, pero no muestra qué tan bien funciona en partes individuales de los datos. ¿Qué técnica es extremadamente útil para abordar esta deficiencia?

1 / 1 punto

- ☒ Segmentación de datos
- ☐ Métricas de transmisión
- ☐ Haz Apache
- ☐ Análisis de métricas de TensorFlow (TFMA)

✓ **Correcto**  
¡Así es! El corte se ocupa de comprender cómo se está desempeñando un modelo en cada subconjunto de datos.

2. Las métricas de transmisión son aproximaciones a las métricas de rendimiento de paso completo calculadas en \_\_\_\_\_.

1 / 1 punto

- ☐ el conjunto completo de datos de validación.
- ☒ mini lotes de datos
- ☐ rebanadas de datos
- ☐ el conjunto completo de datos

✓ **Correcto**  
¡Así es! Esta es una buena manera de aproximar las métricas de paso completo sin incurrir en un enorme costo de sobrecarga computacional.

3. Un reciente programa de fidelización de tarjetas de crédito ofrecido por una gran empresa de tecnología ha sido etiquetado como “**sexista**”, un claro ejemplo de discriminación social basada en algoritmos .

1 / 1 punto

Examinemos una queja de un usuario en Twitter: “Mi esposa y yo presentamos declaraciones de impuestos conjuntas, vivimos en un estado de propiedad comunitaria y hemos estado casados por mucho tiempo. Sin embargo, el algoritmo de la caja negra cree que merezco 20 veces el límite de crédito que ella merece. No funcionan las apelaciones. Estos y otros reclamos similares han desencadenado una investigación completa por parte del Departamento de Servicios Financieros del Estado de Nueva York. ¿Cuál de las técnicas revisadas en la conferencia podría haberse implementado para prevenir este embarazoso problema?

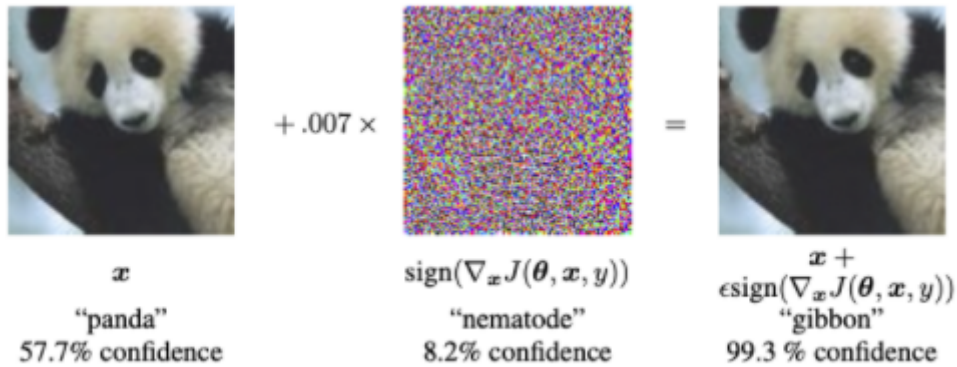
- ☐ Análisis residual
- ☒ Depuración de modelos
- ☐ Robustez del modelo
- ☐ Segmentación de datos

✓ **Correcto**

¡Así es! La depuración de modelos intenta mejorar la transparencia de los modelos al resaltar cómo fluyen los datos en el interior y, por lo tanto, puede prevenir la discriminación social dañina.

4. Se puede engañar a las redes neuronales convolucionales de última generación para que clasifiquen erróneamente imágenes corruptas de ruido con cambios que son completamente imperceptibles para el ojo humano, como se ilustra en la siguiente imagen:

1 / 1 punto



¿Qué tipo de análisis puede ayudarnos a detectar y prevenir este tipo de escenarios?

☒ Análisis de sensibilidad

☐ Ataque adversario

☐ Reducción de dimensionalidad

☐ Análisis de residuos

☒ **Correcto**

¡Así es! El análisis de sensibilidad ayuda a comprender un modelo al examinar el impacto que tiene cada característica en la predicción del modelo. En el análisis de sensibilidad, experimentamos cambiando el valor de una característica mientras mantenemos constantes las otras características y observamos los resultados del modelo.

5. Una brecha de métrica de rendimiento entre dos o más grupos podría ser una señal de que un modelo de ML puede tener sesgos injustos. Por lo tanto, ¿es el logro de la igualdad de desempeño (en los indicadores de equidad) entre los grupos una señal definitiva de que un modelo es justo?

1 / 1 punto

☐ Sí

☒ No

☒ **Correcto**

¡Así es! Los sistemas son muy complejos y lograr la igualdad en una, o incluso en todas las métricas proporcionadas, no puede garantizar la equidad. Las evaluaciones de imparcialidad deben realizarse durante todo el proceso de desarrollo y también después del lanzamiento.

6. Después de implementar un modelo, ¿es generalmente factible realizar un análisis residual?

1 / 1 punto

☐ Sí

☒ No

☒ **Correcto**

¡Así es! Una vez que se implementa su modelo, es posible que no tenga una buena cantidad de datos etiquetados y, en consecuencia, el análisis residual puede resultar un ejercicio costoso, ya que podría incluir la contratación de trabajadores para etiquetar los datos de prueba entrantes.