CORRECCIÓN QUIZ MACHINE LEARNING

NOMBRE: LORENA LUGO

1. Una correlación de Pearson entre dos variables es cero, pero sus valores aún pueden relacionarse entre si.

RTA: True, porque a pesar de que puede ser cero, pueden haber otro tipo de relaciones como lo son la lineal, cuadrática, cúbica, etc.

2. La gráfica de a continuación es correcta (AI=inteligencia artificial ML= Machine Learning, DL= Deep Learning)

RTA: True, Inteligencia artificial contiene ML y ML contiene a DL.

3. Al realizar una regresión o clasificación ¿Cuál es la forma correcta de preprocesar los datos?

RTA: Normalizar la data 🡪 Análisis de componenter principales (PCA) 🡪 Training

4. Cuál es una de las desventajas de los árboles de decisión?

RTA: Los árboles de decisión tienden a sobreajustarse, porque aprenden sólo los datos de entrenamiento y no tienden a predecir bien. Propensos a overfiting.

5. Un clasificador con accuracy 100% en train y el 70% en test ES MEJOR que uno con 70% de accuracy en train y un 75% en test.

RTA: False, porque el accuracy es mejor con el 75% de test que son los datos que no está viedo, mientras que en el de 100% hay un sobreentrenamiento.

6. Los algotirmos de **ML** permiten a los computadores aprender de la data y ser capaces de mejorar por si mismos.

7. Los tipos de problemas de ML más comunes son: Regresión, Clasificación y Clustering (Todos los anteriores), que hacen parte de aprendizaje supervisado y no supervisado.

8. Cual de las siguientes oraciones es falsa respecto a regresión?

RTA: Descubre relaciones causales, ya que es muy difícil de determinar.

9. La diferencia principal entre aprendizaje supervisado y no supervisado es tener una variable respuesta a modelar.

RTA: True, porque el supervisado siempre va a buscar predecir una variable dependiente de una o varias variables independientes, mientras que el no supervisado busca encontrar patrones y no modelar.

10. Machine Learning es una rama de la IA.

11. En qué casos el método de k-means NO dará buenos resultados?

RTA: Con outliers y grupos dominantes. Tener outlier afecta la clasificación porque se basa en distancias, entre más lejos esté un dato es más difícil de clasificar y cuándo hay grupos dominantes suele generar confusión al método. Es un buen método cuando se tienen grupos concentrados.

12. Es posible tener algoritmos que funcionen para resolver un problema de regresión y uno de clasificación? True

RTA: True, como el XGBOOST, Random forest, etc.

13. Cuál de los siguientes algoritmos no hace parte del aprendizaje supervisado?

RTA: PCA 🡪 Es una reducción de dimensionalidad, que hace parte del lenguaje no supeervisado.

14. Las técnicas de One Hot Encoding y Label Encoding se pueden usar indistintamente del tipo de variable

RTA: False, cada una debe usarse para un caso especial. Depende de las categorías que hayan en las variables, si hay pocas categorías se usa OHE y si hay muchas categorías se debe usar LE.

15. Cuál de las siguientes formas es una forma razonable para seleccionar el número de componentes principales K?

RTA: Elbow Method, es una gráfica que permite determinar el número de conglomerados en un conjunto de datos.