Logotipo, nombre de la empresa

Descripción generada automáticamente

Universidad de Montevideo – Introducción a la Ciencia de Datos

Examen 2020 - S2

22 de Febrero, 2021

1. Explique qué es el operador JOIN y cite un ejemplo práctico de uso.
2. Explique qué es One Hot Encoding y ponga un ejemplo.
3. ¿Qué es un confounding factor? Cite un ejemplo.
4. ¿Cómo se evalúa la significancia estadística de un insight? Detalle el proceso.
5. ¿Qué es una regresión lineal y qué problemas hay asociados a la misma?
6. Usted ha entrenado un modelo predictivo para predecir los costos de construcción de edificios a partir de datos de empresas constructoras en Uruguay. Tras haber tenido éxito, una empresa constructora en Inglaterra lo contrata para que le realice un trabajo similar. ¿Qué es esperable que suceda con la precisión de su modelo con los datos ingleses y por qué? Justifique su respuesta.
7. Explique los pasos para construir un árbol de decisión. ¿Qué métodos usaría para decidir qué atributo colocar en la raíz o en los diferentes nodos internos? ¿Es correcto hacerlo de manera aleatoria? Justifique.
8. Explique el algoritmo K-means, mencionando problemas asociados y cómo determinar el número óptimo de vecinos.
9. En base a un dataset de detección de cáncer usted entrena un algoritmo de clasificación que alcanza un *accuracy* de 95%. ¿Qué podría preocuparle de su modelo?
10. Si usted quiere predecir la probabilidad de que un potencial cliente compre su producto, en base a su edad, sexo y nivel de ingresos, ¿qué algoritmo le parece más apropiado? Justifique su respuesta.
    1. Regresión Logística
    2. Regresión Lineal Simple
    3. Regresión Lineal Múltiple
    4. K-Means
    5. Ninguna de las anteriores
    6. Todas las anteriores