

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

TECNOLOGÍAS DE LA INFORMACIÓN



Extracción de Conocimiento en Bases de Datos

DOCENTE:

Enrique Mascote

Diagnóstico Unidad 3: Análisis supervisado

PRESENTA:

Iván Eduardo Martínez Martínez

Grupo:

IDGS91N

Formato: 8 preguntas (5 opción múltiple, 3 respuesta corta)

1. ¿Cuál no es un algoritmo de regresión?
 - A) Regresión lineal
 - B) K-Means**
 - C) Lasso
 - D) Ridge
2. En clasificación, ¿qué métrica mide la proporción de verdaderos positivos correctamente identificados?
 - A) Precision
 - B) Recall**
 - C) Accuracy
 - D) MSE
3. ¿Qué técnica ayuda a evitar el sobreajuste ajustando penalizaciones?
 - A) Cross-validation
 - B) Regularización**
 - C) PCA
 - D) DBSCAN
4. ¿Cuál es el propósito de GridSearchCV?
 - A) Ajustar hiperparámetros**
 - B) Dibujar curvas ROC
 - C) Preprocesar datos
 - D) Reducir dimensionalidad
5. ¿Qué representa la curva ROC?
 - A) Relación entre precisión y recall
 - B) Tasa de verdaderos positivos vs. tasa de falsos positivos**
 - C) Error vs. tiempo
 - D) Correlación de variables

Respuesta Corta

6. Menciona un caso práctico donde usarías regresión.

Respuesta: Predecir el precio de una vivienda en función de sus características, como metros cuadrados, número de habitaciones, antigüedad y ubicación.

7. ¿Por qué es útil la validación cruzada?

Respuesta: La validación cruzada (cross-validation) es útil porque proporciona una estimación más robusta y confiable del rendimiento de un modelo.

8. Explica brevemente qué es una matriz de confusión.

Respuesta: Una matriz de confusión es una tabla que se utiliza para describir el rendimiento de un modelo de clasificación. Muestra las frecuencias de clasificación, comparando las etiquetas reales con las predichas.