

Universidad Tecnológica de Chihuahua
Tecnologías de la Información



**Universidad Tecnológica
de Chihuahua**

Diagnóstico Unidad 3: Análisis
supervisado

Alumno:

Jatzel Israel Cruz Castruita

Grupo:

IDGS91N

Materia:

Extracción de Conocimiento en Bases de Datos

Docente:

Enrique Mascote

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

¿Cuál no es un algoritmo de regresión?

A) Regresión lineal

B) K-Means

C) Lasso

D) Ridge

En clasificación, ¿qué métrica mide la proporción de verdaderos positivos correctamente identificados?

A) Precision

B) Recall

C) Accuracy

D) MSE

¿Qué técnica ayuda a evitar el sobreajuste ajustando penalizaciones?

A) Cross-validation

B) Regularización

C) PCA

D) DBSCAN

¿Cuál es el propósito de GridSearchCV?

A) Ajustar hiperparámetros

B) Dibujar curvas ROC

C) Preprocesar datos

D) Reducir dimensionalidad

¿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

Menciona un caso práctico donde usarías regresión.

Para predecir el precio de una casa usando variables como tamaño, ubicación y número de habitaciones.

¿Por qué es útil la validación cruzada?

Porque permite evaluar el modelo con diferentes particiones de los datos, evitando sobreajuste y dando una estimación más confiable del desempeño real.

Explica brevemente qué es una matriz de confusión.

Es una tabla que muestra cómo se distribuyen las predicciones de un modelo de clasificación: verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.