

Universidad Peruana de Ciencias Aplicadas.

Curso: Data Mining Tools.

Semestre: 2024-2.

Docente: Carlos Fernando Montoya Cubas.



Examen Parcial.

Instrucciones:

Cada pregunta vale cuatro puntos, dos puntos por codificación y dos por interpretación y documentación.

Entregar un documento y/o notebook conteniendo los nombres de los integrantes del grupo.

Dado el dataset de **Hotel Reservations Dataset** disponible en:

<https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset>

1. Realizar el EDA, cada paso debe contener la interpretación de los resultados.
2. Definir por lo menos un corte en el dataset para realizar la interpretación de los resultados, justificar la elección de la variable de corte y explicar detalladamente los resultados encontrados.
3. Realizar la extracción de características, extraer por lo menos una característica manual, y dos características automáticas, justificar y documentar.
4. Usar pipelines de sklearn para realizar la conversión de variables categóricas (nominales y ordinales) a numéricas y realizar el escalado correspondiente, justificar el tipo de escalado escogido.
5. Realizar un proceso de selección de características y reducción de características de los datos. Explicar los resultados.