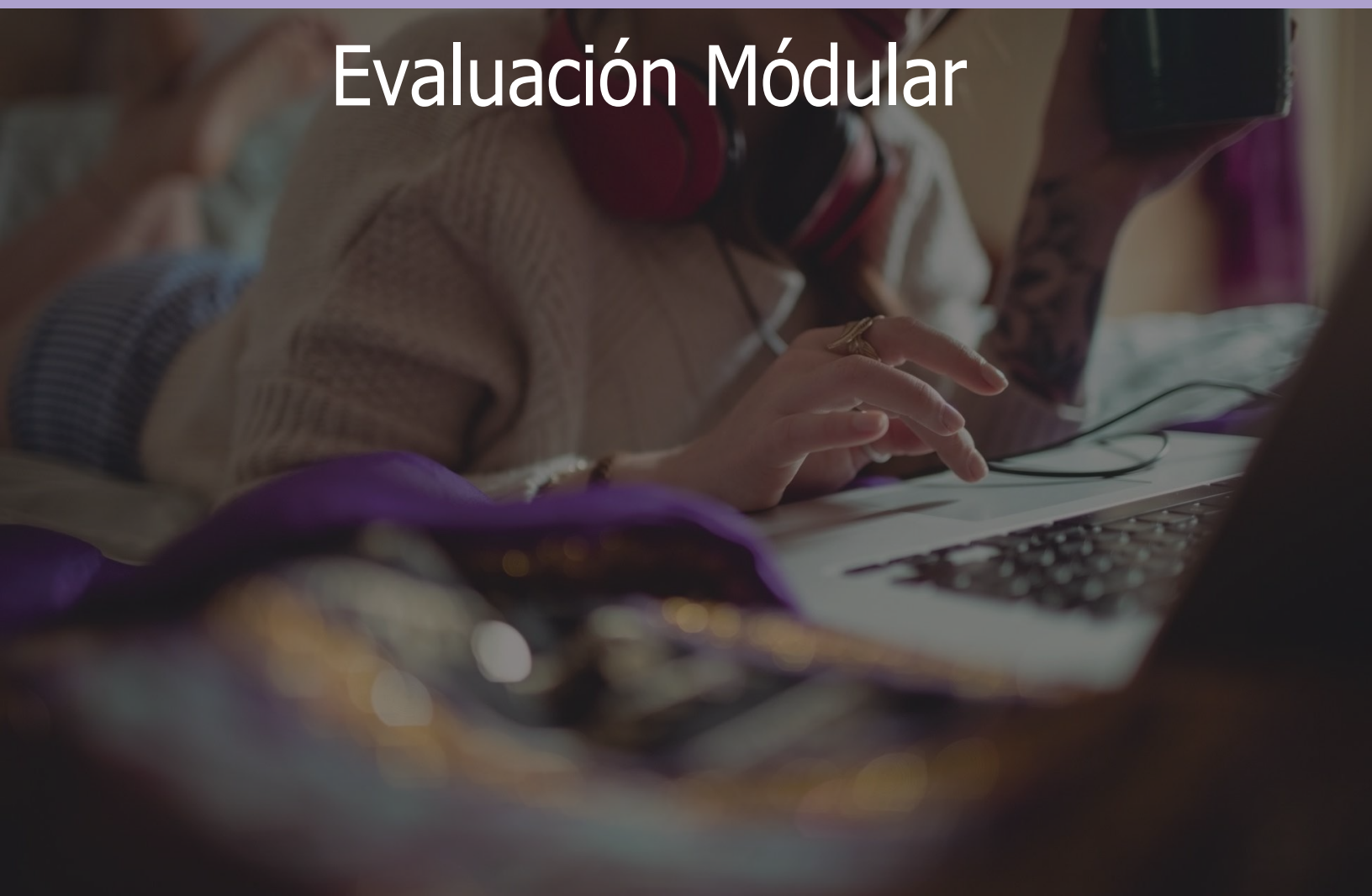


# Módulo 6

Evaluación Modular



# ACTIVIDAD:

## Desarrollo de Modelos Predictivos y Clasificatorios con Validación Cruzada

### Objetivo:

Aplicar los conocimientos adquiridos durante el módulo para construir modelos de regresión y clasificación usando un dataset realista. La solución debe demostrar habilidades en preprocesamiento de datos, entrenamiento de modelos supervisados, evaluación mediante métricas apropiadas y validación cruzada, junto con la presentación de un informe técnico que comunique adecuadamente los resultados obtenidos.

### Contexto:

En la actualidad, los ingenieros de datos no solo deben gestionar la infraestructura de datos, sino también ser capaces de generar valor mediante modelos de Machine Learning. Durante este módulo, se ha abordado el uso de técnicas supervisadas para resolver problemas de regresión y clasificación a partir de datos estructurados.

Esta evaluación busca consolidar esos aprendizajes mediante la resolución de un problema con datos de seguros médicos. Se espera que el estudiante logre entrenar modelos predictivos, realizar análisis comparativos entre técnicas de validación cruzada y presentar sus hallazgos de forma estructurada y reflexiva.



# Instrucciones:

El entregable debe contener los siguientes componentes:

1. **modelo\_regresion.ipynb**: Desarrollo de un modelo de regresión para predecir el costo del seguro (charges).
2. **modelo\_clasificacion.ipynb**: Desarrollo de un modelo de clasificación binaria (por ejemplo, si charges > 10000).
3. **validacion\_cruzada.ipynb**: Aplicación y comparación de al menos dos técnicas de validación cruzada (Holdout, K-Fold, LOOCV).
4. **informe\_final.docx o .pdf**: Documento con análisis, reflexiones, justificaciones técnicas y capturas del proceso.
5. **esquema\_modelo.png (opcional)**: Diagrama explicativo del flujo general de trabajo.

## Recomendaciones:

- Utiliza el dataset insurance.csv incluido en la carpeta ([Puedes descargar la carpeta acá](#)).
- Aplica preprocesamiento adecuado: codificación de variables categóricas y escalado de variables numéricas.
- Usa modelos como LinearRegression, LogisticRegression o KNeighborsClassifier.
- Evalúa los modelos con métricas correspondientes:
  - Regresión: MAE, RMSE,  $R^2$ .
  - Clasificación: matriz de confusión, precisión, recall, ROC-AUC.
- Compara y reflexiona sobre los resultados obtenidos mediante distintas técnicas de validación cruzada.
- El informe debe estar bien redactado, con justificaciones claras y capturas del proceso realizado.

## Entregable opcional (para nota máxima):

Simulación de despliegue en Colab que incluya:

- Una función interactiva de predicción con input() o widgets.
- Demostración de carga del modelo y predicción con nuevos datos.

## Rúbrica:

Criterio	Insuficiente (0-20%)	Por lograr (21-40%)	Medianamente logrado (41-60%)	Logrado (61-80%)	Sobresaliente (81-100%)
1. Implementación de modelos	Sin modelos funcionales o incorrectos	Modelos incompletos o con errores graves	Modelos básicos funcionales con fallas menores	Modelos correctamente implementados	Modelos robustos, bien documentados y explicados
2. Preprocesamiento y limpieza de datos	No realizado o inadecuado	Parcialmente correcto	Correcto pero con omisiones	Correcto y completo	Correcto, detallado y justificado
3. Validación cruzada y análisis comparativo	No implementado	Implementado pero mal interpretado	Implementado con análisis limitado	Comparación clara con reflexión	Comparación crítica y profunda con aportes relevantes
4. Informe final	Ausente o muy pobre	Incompleto o con errores	Presenta estructura básica	Claro, estructurado y comprensible	Redacción excelente, visual, bien argumentado
5. Presentación y organización general	Archivos mal organizados o faltantes	Entrega parcial y desordenada	Entrega completa con estructura básica	Entrega clara y coherente	Entrega impecable, bien estructurada y profesional