

Evaluación Parcial 1

Iniciando un proyecto de Machine Learning

Sigla	Nombre Asignatura	Tiempo Asignado	% Ponderación
MLY0100	Machine Learning	4 horas	

1. Situación evaluativa

☐ Ejecución práctica

☒ Entrega de encargo

☒ Prueba escrita

☐ Presentación

2. Agente evaluativo

☒ Heteroevaluación

☐ Coevaluación

☐ Autoevaluación

3. Instrucciones para el/la estudiante

Esta es una evaluación que se compone de 2 partes. Una evaluación es de carácter teórico donde se medirán conocimientos técnicos y teóricos de manera individual, que tiene un 30% de ponderación sobre la nota parcial 1.

La segunda parte de la prueba parcial 1, corresponde a una entrega del caso semestral, donde deberán reunirse en parejas, y trabajar analizando y realizando las primeras etapas de la metodología CRISP-DM con el fin de implementar modelos de Machine Learning con contexto del caso entregado por tu docente. Esta entrega tendrá una ponderación del 70% sobre la nota parcial 1.

El tiempo para desarrollar esta evaluación es de 4 semanas y se realiza en parejas. La ponderación total de la prueba parcial 1 corresponde al 40% de la asignatura.

La evaluación consiste en:

- Hacer un Jupyter Notebook programado en lenguaje Python, con formato informe, donde se puedan ir detallando cada una de las fases de la metodología CRISP-DM y los procedimientos realizados para lograr el objetivo.
- Escribir y hacer uso de markdown que expresen y detallen los motivos que los llevaron a realizar ciertos procedimientos con la data, los descubrimientos y patrones que observan en los datos.
- Complementar con los aspectos del negocio y de los datos en el enunciado las decisiones tomadas durante el proceso.
- Hacer uso de componentes estadísticos y matemáticos revisados en el curso, con el fin de hacer más completo el análisis exploratorio.
- Realizar transformaciones y preprocesamientos a los datos, utilizando las mejores prácticas y combinando las mejores técnicas con el objetivo de tener buenos indicadores y métricas de desempeño al final del proyecto.

Entrega avance de Proyecto a nivel de Análisis Exploratorio inicial y Preprocesamiento de los datos, abordando las 3 primeras fases de CRISP-DM. Se deben analizar las entidades representadas en el conjunto de datos (tipos de datos, cantidad de datos, estadísticos, distribuciones, etc), detectar distribuciones para los campos que consideren relevantes, además de analizar la existencia de anomalías (valores atípicos) y de missing values. Proponer y ejecutar limpieza de los mismos, incluyendo las transformaciones que consideren necesarias.

4. Pauta de Evaluación

Categoría	% logro	Descripción niveles de logro
Muy buen desempeño	100%	Demuestra un desempeño destacado, evidenciando el logro de todos los aspectos evaluados en el indicador.
Buen desempeño	80%	Demuestra un alto desempeño del indicador, presentando pequeñas omisiones, dificultades y/o errores.
Desempeño aceptable	60%	Demuestra un desempeño competente, evidenciando el logro de los elementos básicos del indicador, pero con omisiones, dificultades o errores.
Desempeño incipiente	40%	Presenta importantes omisiones, dificultades o errores en el desempeño, que no permiten evidenciar los elementos básicos del logro del indicador, por lo que no puede ser considerado competente.
Desempeño insuficiente	20%	Presenta desempeño incorrecto.

Indicador de Evaluación	Categorías de Respuesta					Ponderación Indicador de Evaluación
	Muy buen desempeño 100%	Buen desempeño 80%	Desempeño aceptable 60%	Desempeño incipiente 30%	Desempeño no logrado 0%	
1. Utiliza dentro del formato Jupyter Notebook y durante el proceso las fases de la metodología CRISP-DM como un estándar para abordar el proyecto.						10%
2. Identifica el target adecuado para una tarea de regresión dado el						10%

contexto de negocio y la naturaleza de los datos.						
3. Identifica un el target adecuado para una tarea de clasificación dado el contexto de negocio y la naturaleza de los datos						10%
4. Utiliza librerías de python para Machine Learning como numpy, scikit learn, matplotlib, seaborn entre otras.						10%
5. Utiliza lenguaje Python sobre Jupyter notebook para realizar limpieza y preparación de datos de acuerdo con las buenas prácticas de la industria.						10%
6. Elabora documentación del proceso comparando los resultados obtenidos a partir de la información del caso y de los datos.						10%
7. Aplica tratamiento a los datos atípicos y missing values de acuerdo con la naturaleza de los datos.						10%
8. Integra el uso de estadísticos tanto de tendencia central como de						10%

dispersión para dar explicación a los datos.						
9. Propone técnicas como la normalización o estandarización, según la forma en la que están distribuidos los datos del caso.						10%
10. Establece dentro del Jupyter notebook los pasos realizados a través del uso de markdown la justificación a cada técnica utilizada durante el proceso.						10%