

Evaluación Parcial 2

Nombre: Técnicas de Minería de Datos

Sigla	Nombre Asignatura	Tiempo Asignado	% Ponderación
BIY7121	Minería de Datos	6 semanas	35%

1. Situación evaluativa

☐ Ejecución práctica

☒ Entrega de encargo

☐ Prueba escrita

☐ Presentación

2. Agente evaluativo

☒ Heteroevaluación

☐ Coevaluación

☐ Autoevaluación

3. Tabla de Especificaciones

Resultado de Aprendizaje	Indicador de Logro (IL)	Indicador de Evaluación (IE)	Ponderación Indicador de Evaluación	Ponderación Indicador Logro
RA2 Analizar los datos mediante la aplicación de modelos estadísticos, utilizando el lenguaje Python.	IL 2.1 Utiliza el lenguaje de programación Python para la limpieza y transformación de datos de acuerdo a sus características: como tipo de dato, orden de magnitud o reemplazo por un valor válido.	Aplica estrategias de detección de valores atípicos en los datos siguiendo buenas prácticas de la industria.	20%	25%
		Realiza la búsqueda de valores inexistentes en algunas de las características del conjunto de datos.	20%	
	IL 2.2 Aplica modelos matemáticos sobre los datos, utilizando bibliotecas especializadas de Python, para obtener información significativa del negocio.	Realiza la transformación de variables de acuerdo con la implementación de modelos matemáticos siguiendo buenas prácticas de la industria.	20%	25%
	IL 2.3 Interpreta los resultados obtenidos considerando las necesidades y características del negocio.	Interpreta los resultados obtenidos del trabajo con los datos, de acuerdo con las necesidades y características del negocio.	20%	25%
	IL 2.4 Analiza las características de los datos y del negocio a través de la identificando insight, permitiendo entregar valor y apoyar la toma de decisiones.	Analiza las características de los datos y del negocio identificando insights de alto impacto.	20%	25%
Total				100%

4. Instrucciones para el/la estudiante

Esta es una evaluación que corresponde a una **Entrega de encargo sin presentación** y tiene un **35%** de ponderación sobre la nota final de la asignatura.

El tiempo para desarrollar esta evaluación es de 6 semanas y se realiza en parejas de trabajo fuera del taller de alto computo.

La evaluación consiste en realizar un informe técnico en **formato Jupyter Notebook** que contenga las tres primeras fases de la metodología CRISP-DM aplicada a los datos del clima en Australia. Para el desarrollo de la evaluación utilice ***“Evaluación Parcial2_Anexo2”*** y ***“weatherAUS.xlsx”***.

Instrucciones:

Realizar un informe técnico que contenga las siguientes fases de la metodología CRISP-DM:

- En la fase de **Data Preparation** **realizar** las transformaciones necesarias a los datos para evitar inconsistencias y problemas en la generalización del modelo como overfitting o underfitting.
- Dado la anterior, prepare los datos con las técnicas revisadas en clases, con el objetivo de entrenar y testear en modelos tanto supervisados como no supervisados.
- Además, debe realizar las transformaciones de datos (**Data Transformations**), para que los futuros modelos funcionen de forma adecuada.
- Interprete los resultados obtenidos del trabajo con los datos, de acuerdo con las necesidades y características del negocio.
- Analice las características de los datos y del negocio identificando insights de alto impacto.

****Considerar uso de bloques de códigos debidamente comentados y realizar análisis de la información en bloques Markdown.***

5. 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	30%	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 no logrado	0%	Presenta ausencia o incorrecto desempeño.

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%	
Aplica estrategias de detección de valores atípicos en los datos siguiendo buenas prácticas de la industria.	Aplica estrategias avanzadas de detección de valores atípicos en Python, con métodos estadísticos y técnicas específicas según características del dato.	Aplica estrategias básicas de detección de valores atípicos en Python, utilizando métodos convencionales con rango intercuartílico o	Aplica estrategias de detección de valores atípicos en Python, con aplicaciones básicas y sin considerar la totalidad e las características específicas del dato.	Aplica estrategias de detección de valores atípicos en Python, pero no realiza manejo de los datos para su tratamiento.	No aplica estrategias de detección de valores atípicos en Python o estas no son insuficientes o inexistentes.	20%

		visualización de boxplots.				
Realiza la búsqueda de valores inexistentes en algunas de las características del conjunto de datos.	Detecta todas las características con valores inexistentes.	Detecta la mayoría de las características con valores inexistentes.	Detecta la mitad de las características con valores inexistentes.	Detecta menos de la mitad de las características con valores inexistentes.	No detecta las características con valores inexistentes.	20%
Realiza la transformación de variables de acuerdo con la implementación de modelos matemáticos siguiendo buenas prácticas de la industria.	Realiza la transformación del total de las variables que lo requieran, de manera precisa y ajustada a las buenas prácticas de la industria.	Realiza la transformación de la mayoría de las variables que lo requieran, de manera precisa y ajustada.	Realiza la transformación de la mitad o menos de las variables que lo requieran, de manera precisa y ajustada.	Realiza la transformación solo de algunas de las variables que lo requieran.	No realiza la transformación de las variables.	20%
Interpreta los resultados obtenidos del trabajo con los datos, de acuerdo con las necesidades y características del negocio.	Interpreta la totalidad de los resultados, identificando conexiones estratégicas clave con las necesidades y características del negocio.	Interpreta la mayoría de los resultados, identificando conexiones estratégicas clave con las necesidades y características del negocio.	Interpreta la mitad de los resultados, identificando conexiones relevantes con las necesidades del negocio.	Interpreta menos de la mitad de los resultados y no logra establecer conexiones con las necesidades del negocio.	No interpreta los resultados.	20%
Analiza las características de los datos y del negocio identificando insights de alto impacto.	Realiza un análisis de completo de las características de los datos y del negocio, identificando insights de alto impacto.	Realiza un análisis completo de las características de los datos y del negocio, identificando algunos insights de forma consistente.	Realiza un análisis completo de las características de los datos, pero no identifica insights.	Realiza un análisis incompleto de las características de los datos, no logrando identificar insight.	No demuestra competencia en el análisis significativo de las características de los datos y del negocio.	20%
						100%

