

**Brandon Garrido**  
**Carnet : 25001202**

## **Reporte con comparación de métricas y conclusiones sobre la evolución de los datos.**

Se realizaron cuatro corridas del pipeline utilizando versiones modificadas del dataset original, con el objetivo de evaluar la sensibilidad del modelo gradient\_boosting ante distintos escenarios de calidad y estructura de datos:

- **Versión 1:** dataset original sin modificaciones
- **Versión 2 (simulación de pérdida de columna):** se interpreta como si se hubiera eliminado una variable predictiva relevante
- **Versión 3:** se eliminaron registros para simular datos faltantes
- **Versión 4:** se eliminó una columna completa para evaluar pérdida estructural

### **Comparación de métricas por versión**

	<b>Versión</b>	<b>MAE</b>	<b>RMSE</b>	<b>R<sup>2</sup></b>	<b>Modelo ganador</b>
	v1	0.33175	0.49399	0.81378	gradient_boosting
	v2	0.31881	0.47921	0.80452	gradient_boosting
	v3	0.32488	0.48171	0.82679	gradient_boosting
	v4	0.32950	0.48525	0.82346	gradient_boosting

### **Análisis de evolución**

- **Versión 1** sirvió como línea base, con métricas estables y buen desempeño general.
- **Versión 2**, simulando la pérdida de una columna, mostró un leve aumento en RMSE y una mejora marginal en R<sup>2</sup>. Esto sugiere que la columna eliminada podría no ser crítica y que el modelo logró compensar su ausencia con otras variables.
- **Versión 3**, con registros eliminados, presentó las mejores métricas globales, lo que indica que el modelo se benefició de una depuración de datos posiblemente ruidosos.
- **Versión 4**, con una eliminación estructural real, mantuvo métricas competitivas, confirmando la robustez del modelo ante pérdida de información.

## **Conclusiones**

- El modelo gradient\_boosting fue consistente en todas las versiones, adaptándose bien a cada modificación del dataset.
- La versión 3 ofreció el mejor balance entre precisión y capacidad explicativa, siendo la más favorable en términos de métricas.
- La versión 2, simulando la pérdida de una columna, mostró que el modelo puede mantener estabilidad incluso ante la ausencia de variables.
- La versión 4 confirmó que el pipeline es resistente a la pérdida estructural, manteniendo métricas cercanas a las mejores obtenidas.
- El uso de DVC permitió comparar de forma trazable y reproducible cada experimento, facilitando la validación empírica del impacto de los cambios.