

THEBRIDGE - DATASCIENCE BOOTCAMP

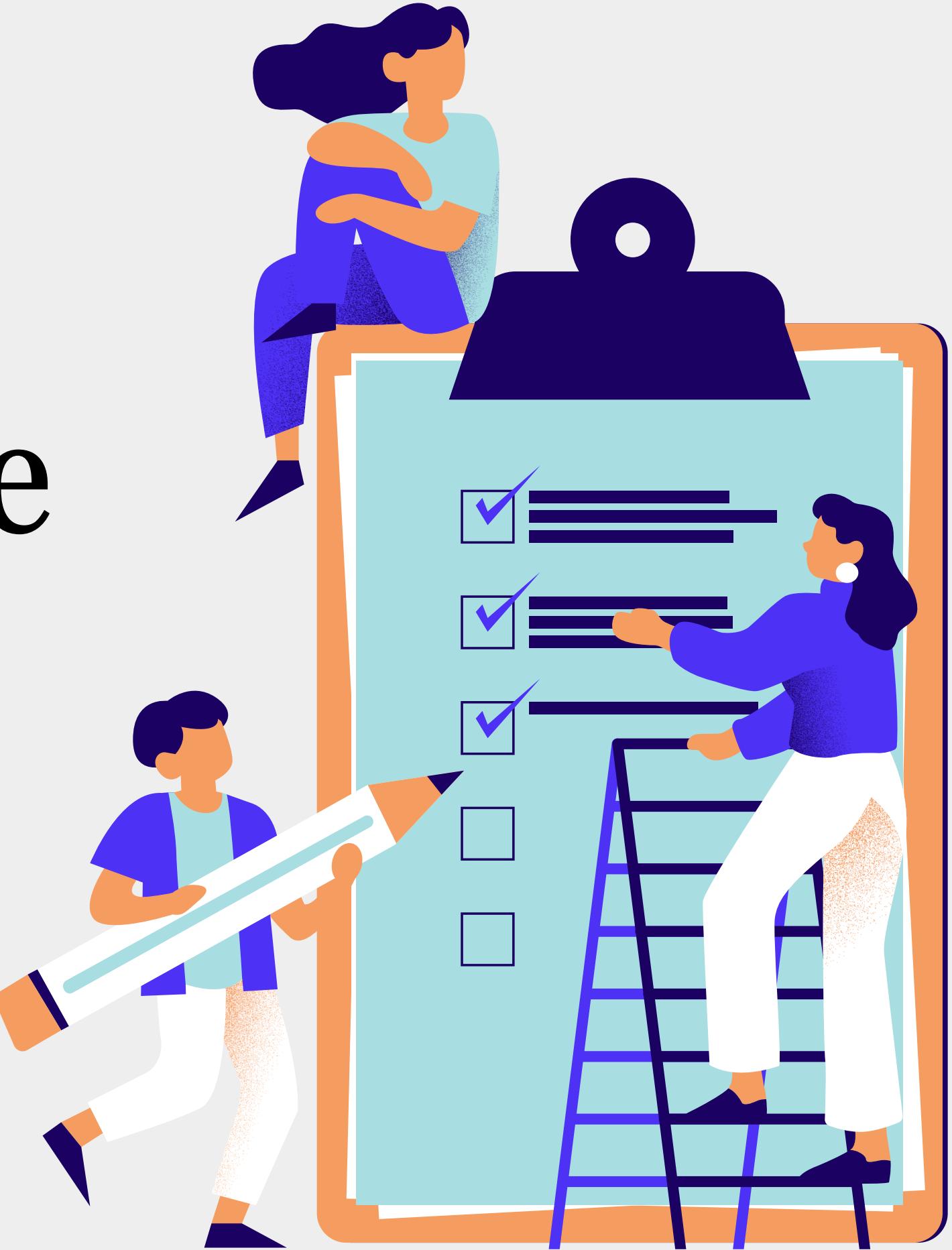
Machine Learning:

Workplace Absence

Alert



DATE: JULY, 2025
BY: EMMA GARCIA



“Anticipar el absentismo: hacia una gestión más eficiente”

- Impacto absentismo
- Importancia análisis predictivo
- Objetivo del proyecto



¿Cómo lo abordamos desde el machine learning?

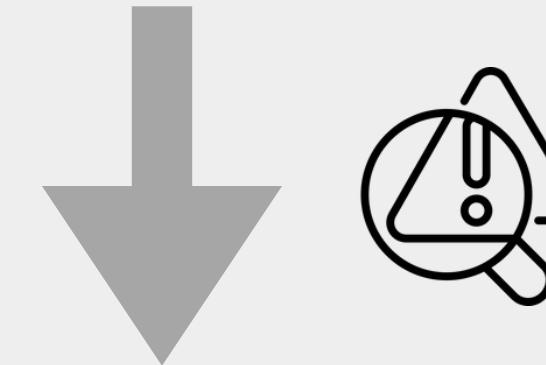
Objetivo empresarial:

Anticipar casos de absentismo significativo

Relevancia de la distinción:

Diferenciar entre ausencias leves ($\leq 3h$) y significativas ($> 3h$)

Problema de negocio



ML

Enfoque inicial:

Problema de regresión

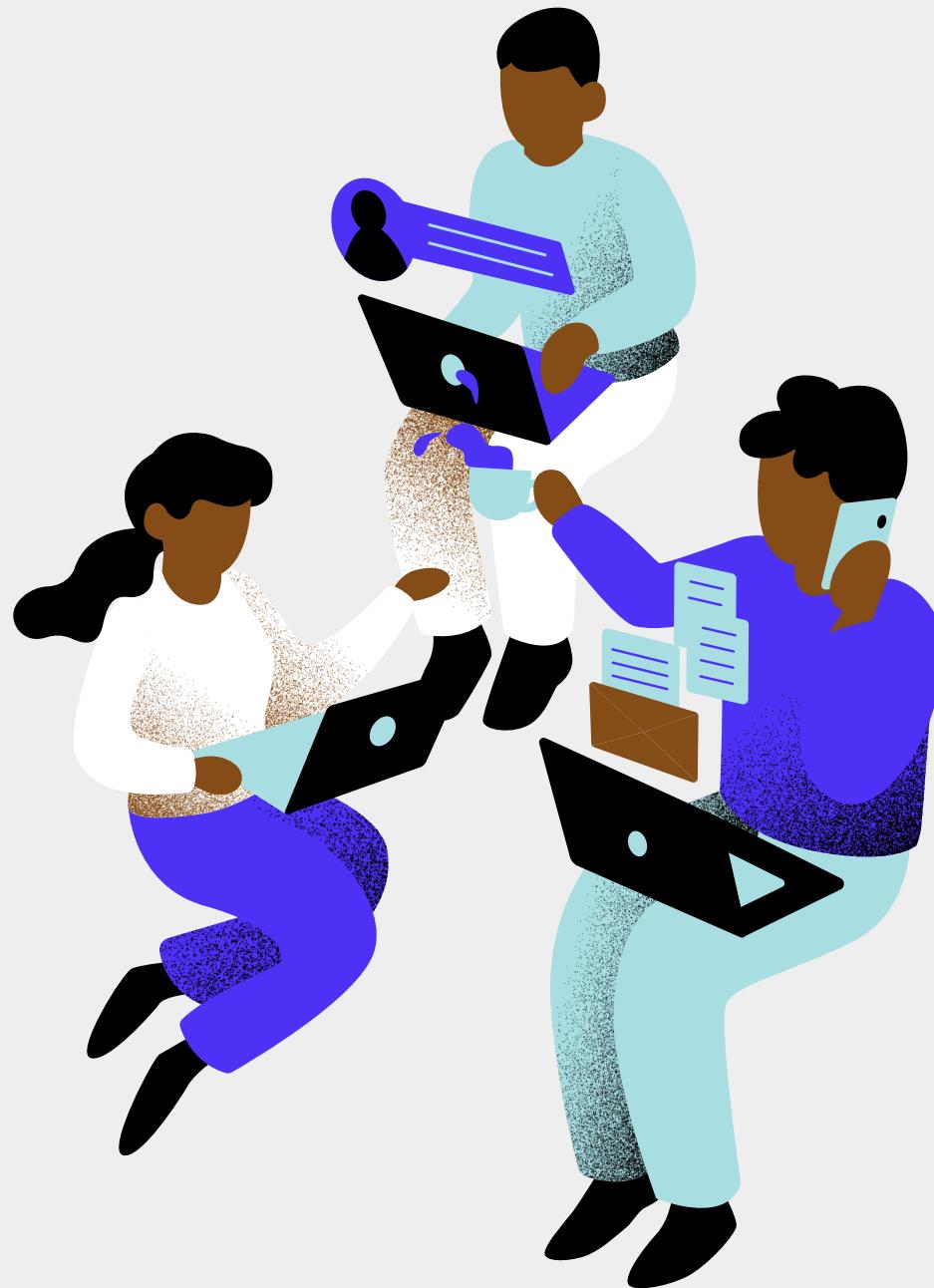
Retos encontrados:

Baja correlación y tamaño de muestra limitado

Reformulación:

Problema en una clasificación binaria (ausente / no ausente)

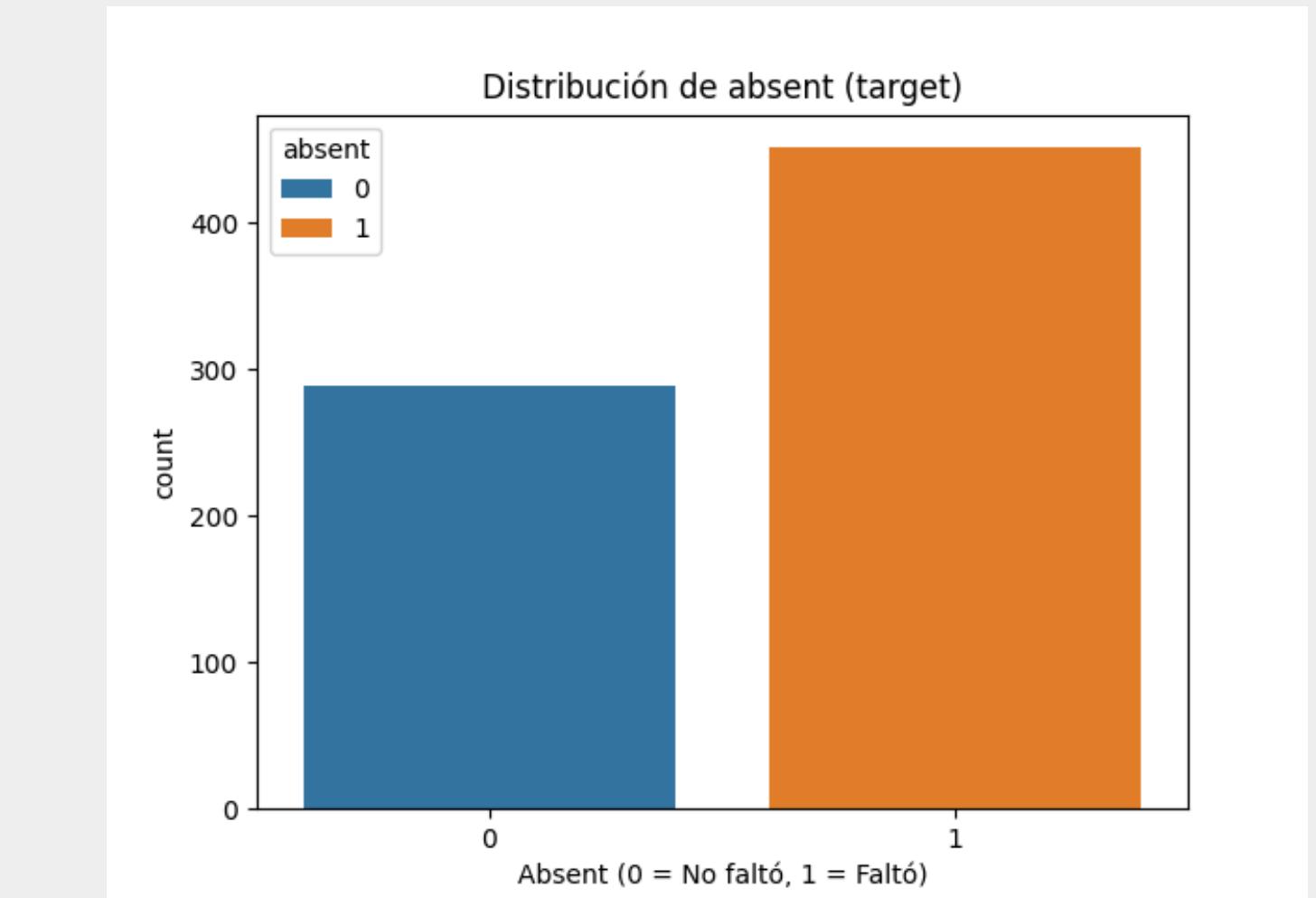
Dataset: *Absenteeism at work*



| Variables | Transformaciones |
|---|---|
| Absent | 0 : faltó menos o 3 horas vs 1 : faltó más de 3 horas |
| Educación | Instituto y Educación Superior |
| Mascotas | Tiene o no mascotas |
| Hijos | Tiene o no hijos |
| ID, Height, Weight y Disciplinary failure | Eliminar |

Data →

740 registros
Compañía Brasil



Preprocessing →

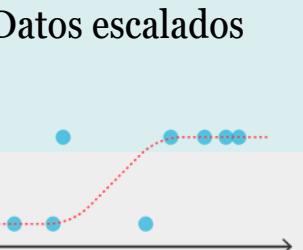
- Transformaciones numéricas
- Transformaciones categóricas

Model

Siguiente diapositiva

Modelos

Logistic Regression



XGBoost Classifier

Datos no escalados

XGBoost

Tras GridSearch



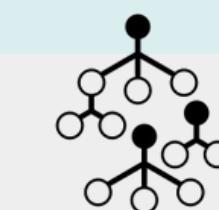
LGBM Classifier

Datos no escalados

 **LightGBM**

Random Forest Classifier

Datos no escalados



Resultados y evaluación



Modelo: LGBM Classifier

Recall 60% clase 1

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.37 | 0.32 | 0.34 | 85 |
| 1 | 0.54 | 0.60 | 0.57 | 115 |
| accuracy | | | 0.48 | 200 |
| macro avg | 0.46 | 0.46 | 0.46 | 200 |
| weighted avg | 0.47 | 0.48 | 0.47 | 200 |

Capacidad para detectar a quienes no
faltaron fue menor



El modelo tiende a sobrepredecir
ausencias

Conclusiones y futuras acciones

Anticipación: El modelo predice casos significativos de absentismo.

Aplicación: Puede integrarse en sistemas de planificación de RRHH.

Valor: Ayuda a identificar perfiles con mayor propensión a faltar.

- Mejora 1: Ampliar los datos históricos de ausencias.
- Mejora 2: Añadir factores contextuales (como clima, eventos, etc.).





¡Gracias!

