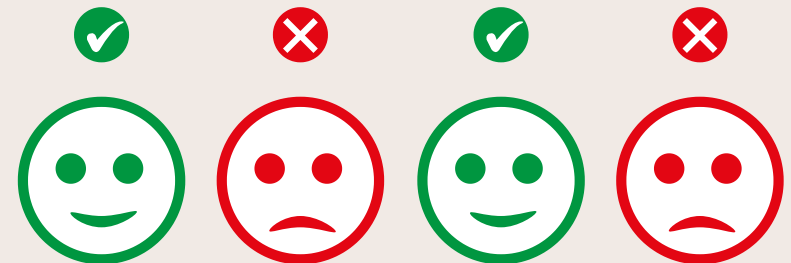


PROYECTO FINAL

(NLP)

CLASIFICACIÓN DE TEXTO - REVIEWS DE PELICULAS



JUSTIFICACIÓN



Identificar la
opinión general
de los
espectadores



Automatizar
el análisis de
grandes
volumenes de
reviews



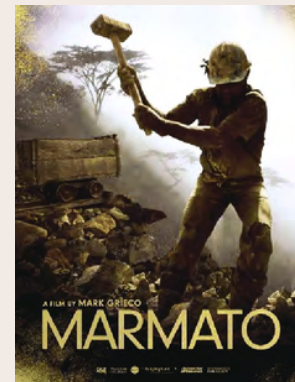
Apoyar la toma
de decisiones en
la industria
cinematográfica

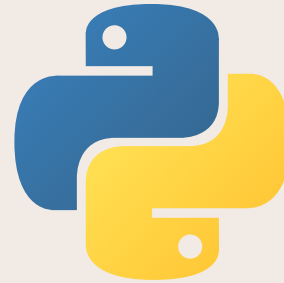


EXTRACCIÓN DE LOS DATOS

01

Se empleó la API de TMDB, la cual ofrece un amplio catálogo de películas y series, junto con información crítica relevante, para llevar a cabo análisis de sentimiento en el campo de NLP. Esta API brinda acceso detallado a cada película, incluyendo título, descripción, género, elenco, calificaciones de usuarios y críticos, entre otros detalles. Su utilidad se extiende tanto para analizar comentarios y críticas de usuarios y expertos en cine, como para predecir la calificación mediante la sinopsis de una trama.





RECUPERACIÓN, LIMPIEZA Y MODELADO DE DATOS



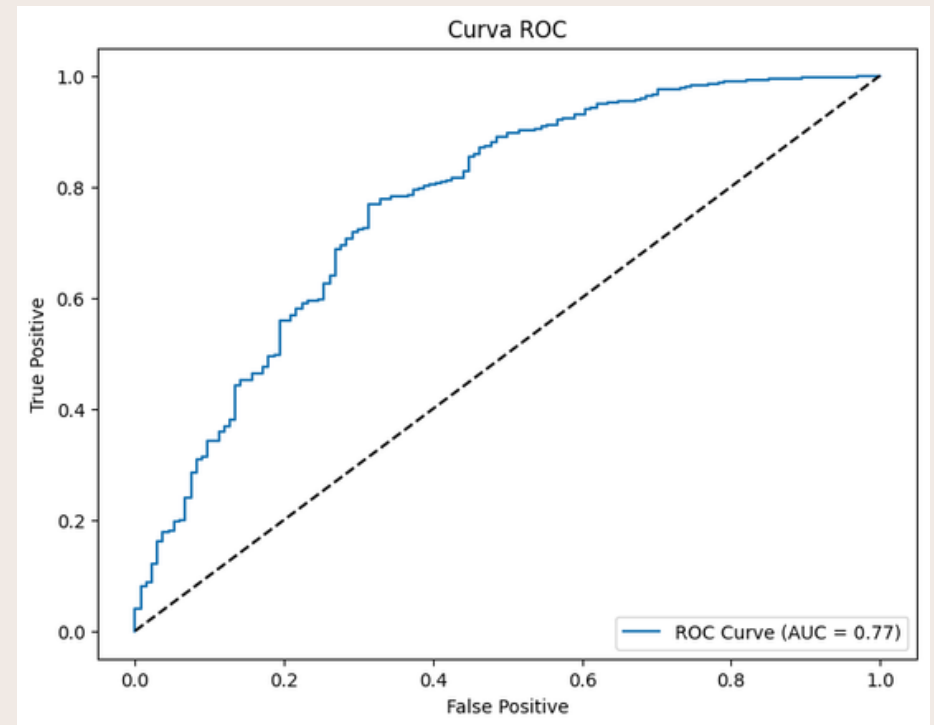
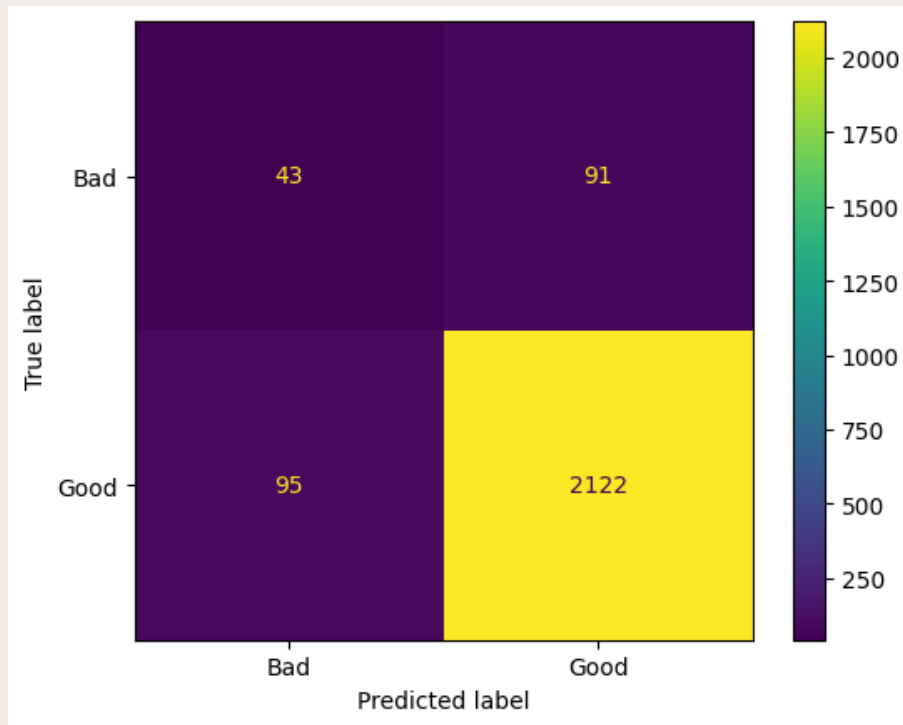
02

MODELO DE REGRESIÓN LOGÍSTICA

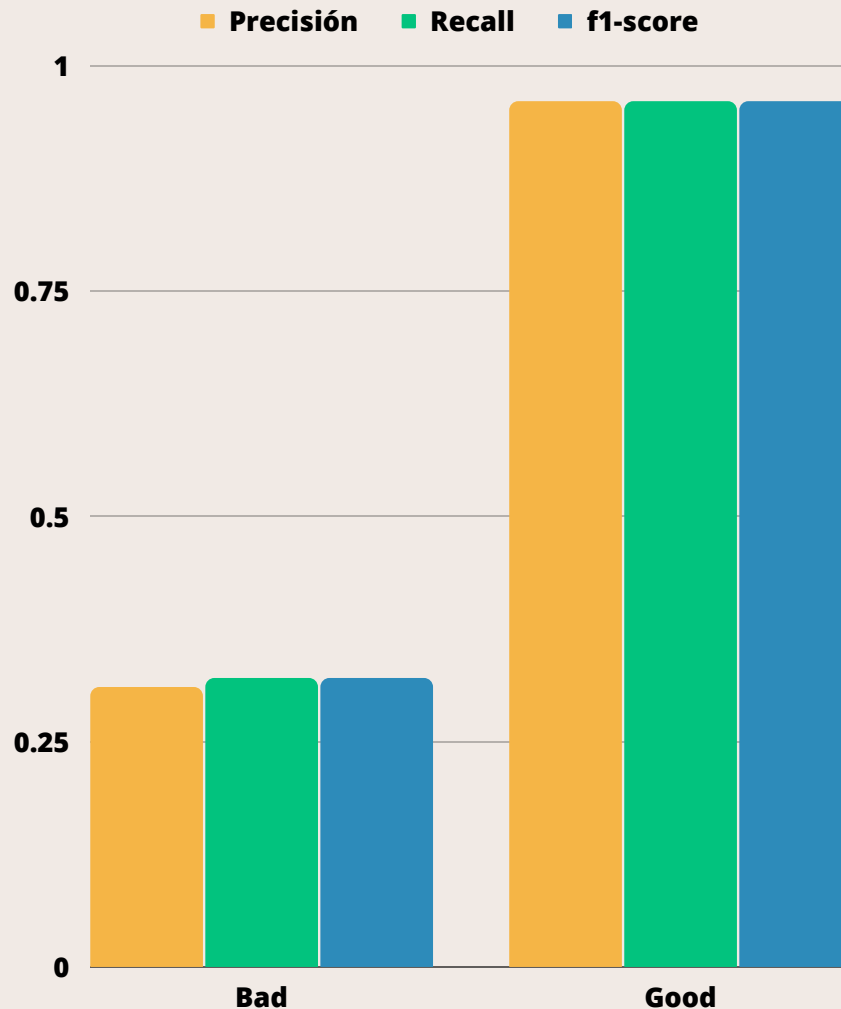
Se implementó un modelo de regresión logística en primer lugar porque proporciona una solución eficiente y de interpretación sencilla para abordar problemas de clasificación binaria. Además, su capacidad para estimar probabilidades de clasificación resulta útil en la toma de decisiones basadas en la confianza del modelo. La simplicidad de su implementación y bajo requerimiento de recursos también fueron factores clave en su elección inicial.



MODELO DE REGRESIÓN LOGÍSTICA - RESULTADOS



MODELO DE REGRESIÓN LOGÍSTICA - RESULTADOS



ACCURRACY

0.92

PRECISION

0.96

ROC AUC

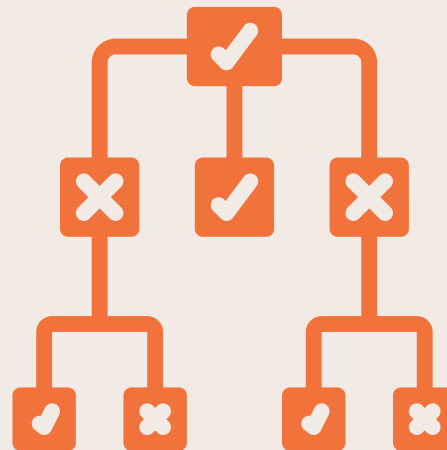
0.77



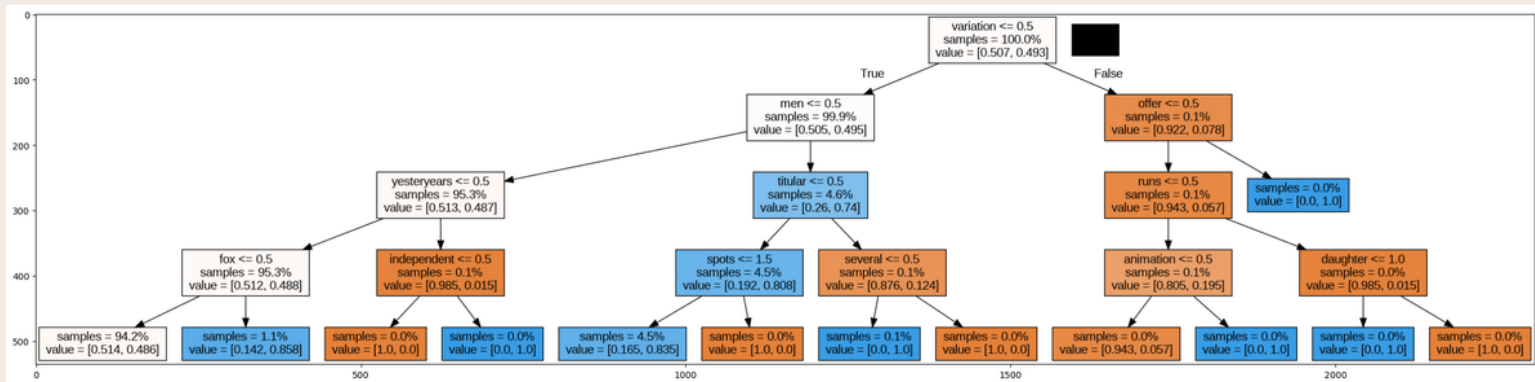
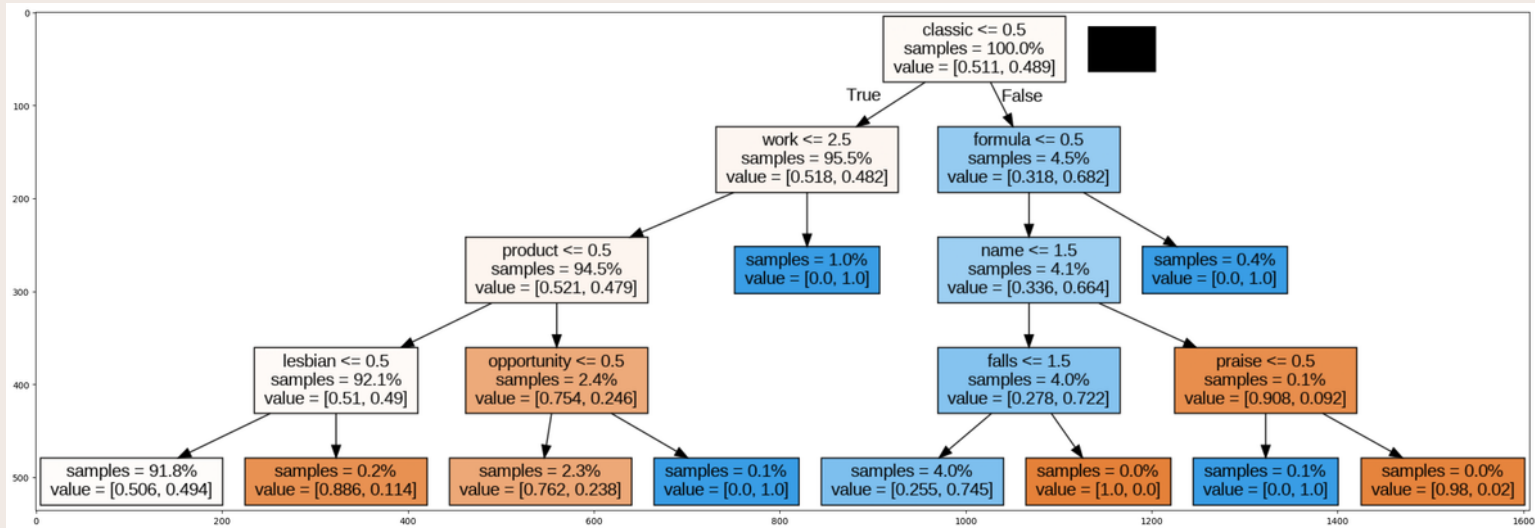
03

MODELO DE ARBOLES DECISORIOS

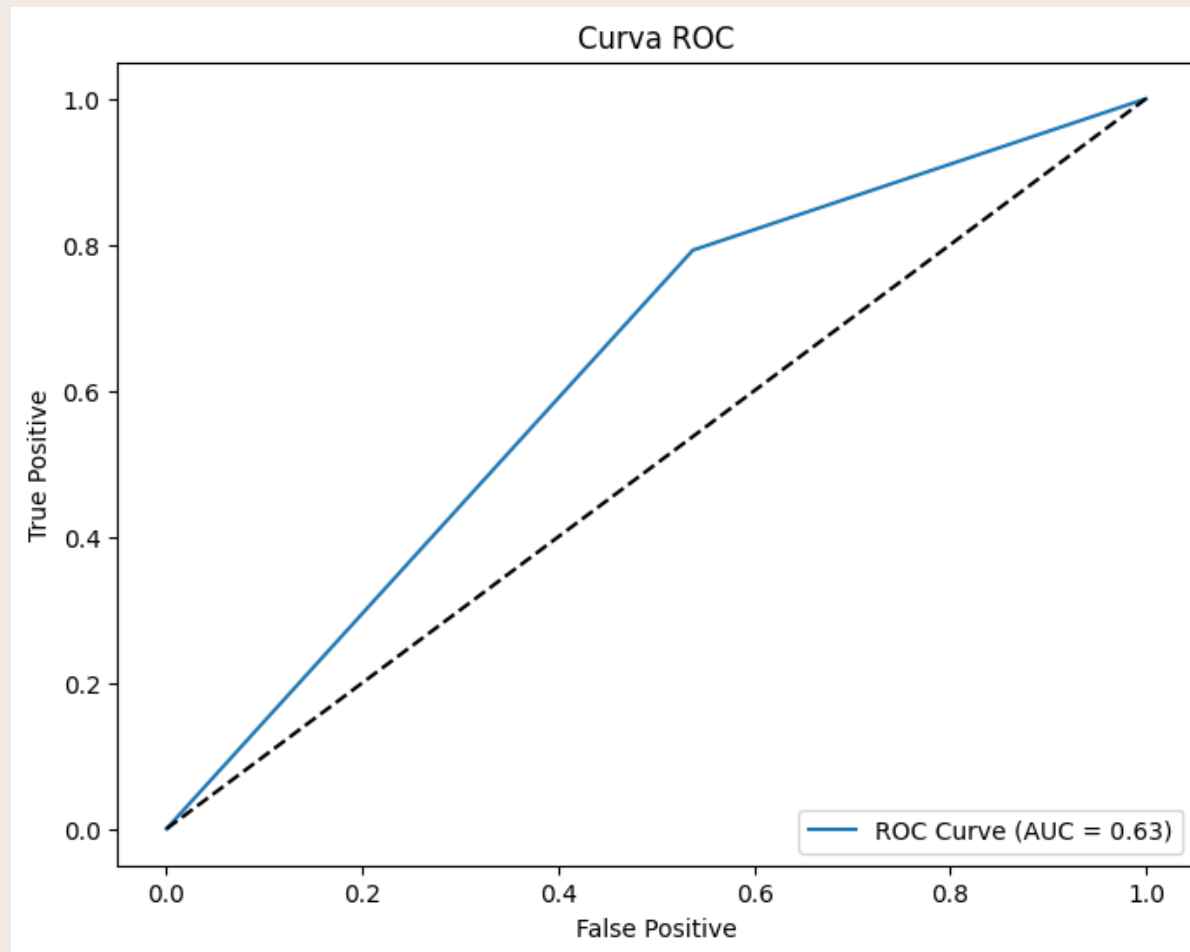
Se eligió el modelo de árboles decisorios por su interpretación sencilla, eficiencia en el entrenamiento y predicción, y su capacidad para manejar características numéricas y categóricas sin transformaciones adicionales.



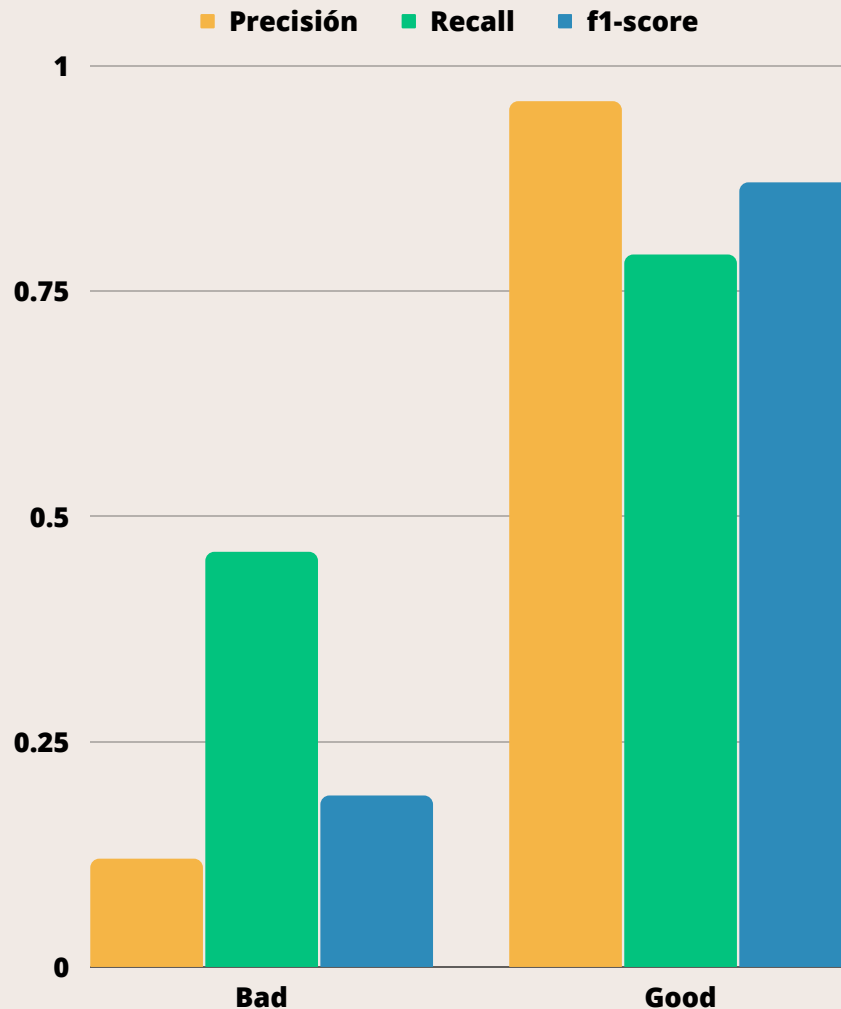
MODELO DE ARBOLES DECISORIOS - RESULTADOS



MODELO DE ARBOLES DECISORIOS - RESULTADOS



MODELO DE ARBOLES DECISORIOS - RESULTADOS



ACURRACY

0.77

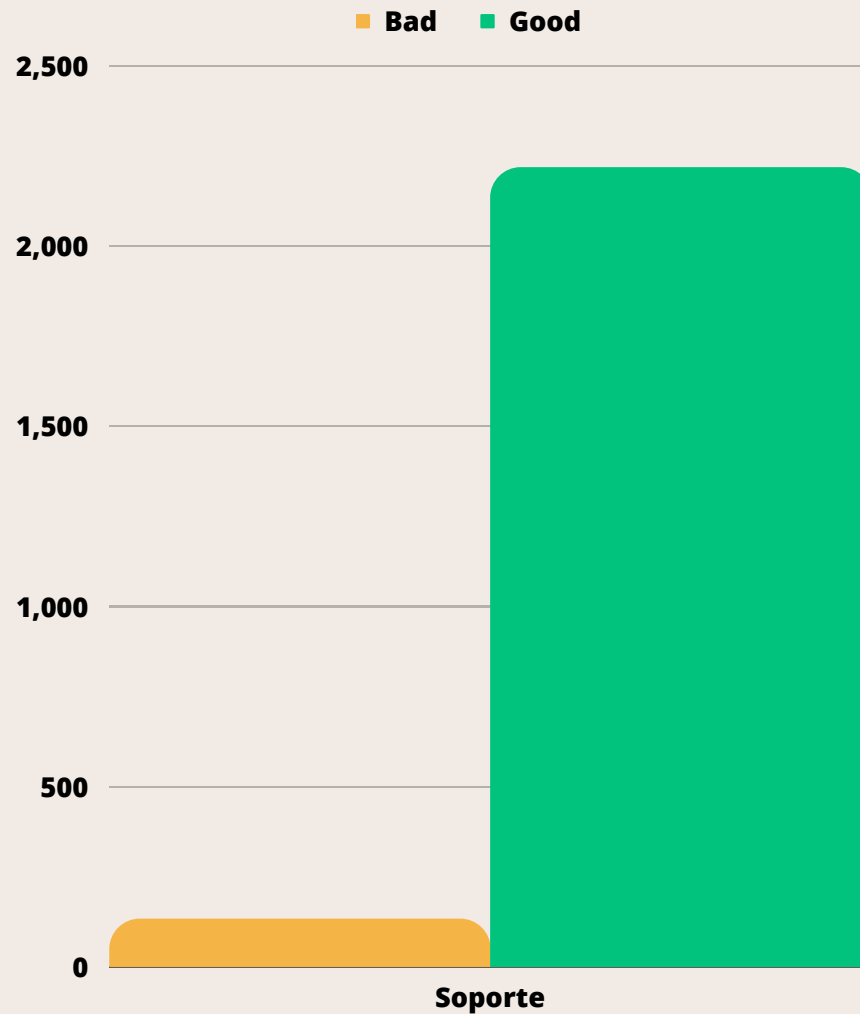
PRECISSION

0.79

ROC AUC

0.63

MODELO DE ARBOLES DECISORIOS - RESULTADOS



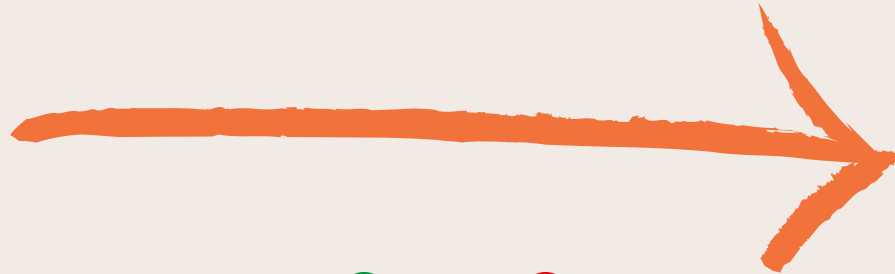
**SOPORTE
CLASE BAD**

134

**SOPORTE
CLASE GOOD**

2217

REEVALUACIÓN



REGRESIÓN LOGISTICA



ACCURRACY
0.93

+0.8%

PRECISSION
0.96

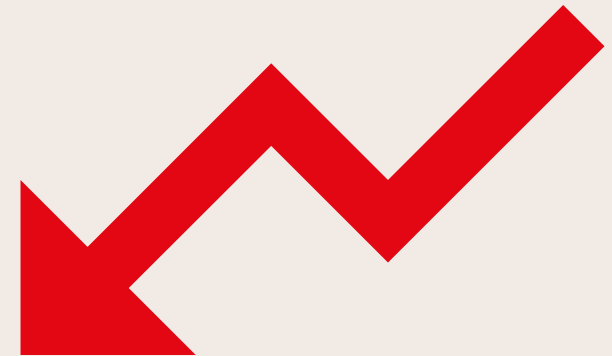
+1.3%

ROC AUC
0.80

+3.1%



ARBOLES DECISORIOS



+1.8%

ACCURRACY
0.79

+0.4%

PRECISSION
0.80

-1.9%

ROC AUC
0.61

MEJORAS PARA FUTUROS MODELOS

Utilizar un conjunto de datos con clases más balanceadas entre sí.



Investigar/Aplicar otras técnicas de manejo de desequilibrio de clases, como oversampling o undersampling, para mejorar la capacidad del modelo para predecir reseñas "Bad".



CONCLUSIONES

El modelo de Regresión Logística ofrece un rendimiento generalmente superior.



La Regresión Logística logra una alta precisión y recall promedio.



El modelo de Regresión Logística también muestra una mejor capacidad para identificar correctamente las reseñas "malas" en comparación con el modelo de Árboles Decisorios.



Ambos modelos tienen un rendimiento desigual en términos de clasificación de las reseñas "malas". Esto sugiere que hay margen de mejora en la identificación de este tipo de reseñas para ambos enfoques.

