

Resumen Ejecutivo: Análisis de Regresión

Proyecto de clasificación de reclamaciones de TikTok

OVERVIEW

El equipo de datos de TikTok busca desarrollar un modelo de aprendizaje automático para ayudar en la clasificación de las reclamaciones de los envíos de los usuarios. Anteriormente, el equipo de datos observó que si un usuario está verificado, es mucho más probable que publique opiniones. Dado que el objetivo final es clasificar las afirmaciones y opiniones, es importante crear un modelo que muestre cómo predecir el comportamiento del tipo de cuenta (verificado) que tiende a publicar más opiniones. Por lo tanto, en esta parte del proyecto, el equipo de datos creó un modelo de regresión logística que predice `verified_status`.

PROJECT STATUS

La variable de `verified_status` se seleccionó para este modelo de regresión debido a la relación observada entre el tipo de cuenta verificada y el contenido del video. Se seleccionó un modelo de regresión logística debido al tipo y distribución de datos.

UN VISTAZO A LOS RESULTADOS DEL MODELO

El modelo de regresión logística logró una precisión del 69% y un recuerdo del 66% (promedios ponderados). Este modelo logró una precisión f1 del 66%. Los resultados de estos modelos informan sobre las funciones de video, que se analizan en "Información clave".

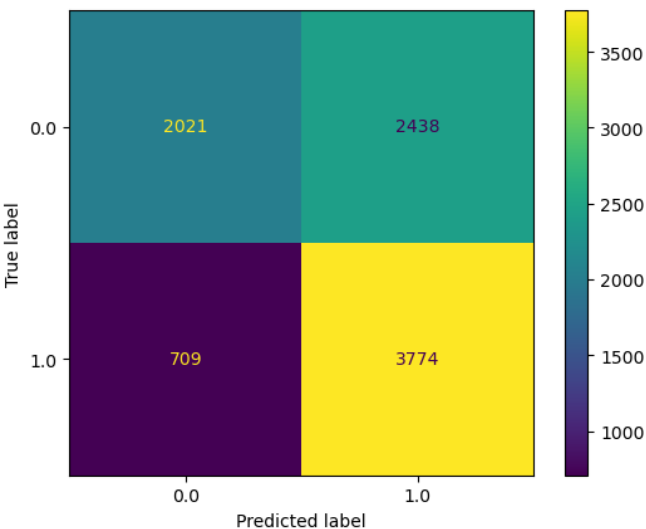
NEXT STEPS

El siguiente paso es construir un modelo de clasificación que prediga el estado de las reclamaciones realizadas por los usuarios. Ese es el proyecto final y la expectativa original del equipo de TikTok. Ahora, hay suficiente información para analizar los resultados de ese modelo con un contexto útil en torno al comportamiento del usuario.

KEY INSIGHTS

Según los coeficientes estimados del modelo a partir de la regresión logística, los videos más largos tienden a asociarse con mayores probabilidades de que el usuario sea verificado. Otras características de video tienen coeficientes estimados pequeños en el modelo, por lo que su asociación con el estado verificado parece ser pequeña. Como resultado, otras funciones de video además de la duración del video no parecen estar asociadas con el estado verificado.

Matriz de confusión para el modelo de regresión logística



Arriba a la izquierda: el número de videos publicados por cuentas no verificadas. Arriba a la derecha: el número de videos publicados por cuentas no verificadas. Abajo a la izquierda: el número de videos publicados por cuentas verificadas. Abajo a la derecha: el número de videos publicados por cuentas verificadas.