

Resumen Ejecutivo: Resultados de Pruebas Estadísticas

Proyecto de clasificación de reclamaciones de TikTok

Project 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. En esta parte del proyecto, el equipo de datos realizará una prueba de hipótesis para analizar la relación entre `verified_status` y `video_view_count`.

Key Insights

- El análisis muestra que hay una diferencia en el número de visualizaciones entre los videos de TikTok publicados por cuentas verificadas y los videos de TikTok publicados por cuentas no verificadas.
- Como resultado, estos hallazgos sugieren que podría haber diferencias de comportamiento fundamentales entre estos dos grupos de cuentas: verificadas y no verificadas.
- Sería interesante investigar la causa raíz de esta diferencia de comportamiento. Por ejemplo, considere:
 - ¿Las cuentas no verificadas tienden a publicar videos más atractivos? ¿Es ese contenido atractivo una afirmación u opinión?
 - ¿O las cuentas no verificadas están asociadas con bots de spam que ayudan a inflar los recuentos de vistas?

Details

El equipo de datos de TikTok consideró la relación entre `verified_status` y `video_view_count`.

Uno de los enfoques que se llevó a cabo fue examinar los valores medios de `video_view_count` para cada grupo de `verified_status` en los datos de la muestra. Los resultados mostraron que la mayoría de las cuentas no estaban verificadas. 265.663 cuentas no fueron verificadas y 91.439 cuentas fueron verificadas.

```
verified_status
not verified    265663.785339
verified        91439.164167
Name: video_view_count, dtype: float64
```

El segundo enfoque fue una prueba de hipótesis de dos muestras. Alineado con los resultados preliminares de los valores medios, este análisis estadístico muestra que cualquier diferencia observada en los datos de la muestra se debe a una diferencia real en las medias de la población correspondiente.

Next Steps

El equipo sugiere avanzar y construir un **modelo de regresión** sobre el estado verificado.

Un modelo de regresión para `verified_status` puede ayudar a analizar el comportamiento de los usuarios en este grupo de usuarios verificados. A continuación, este contexto se puede utilizar para considerar los resultados de un modelo de clasificación de notificaciones que se creará posteriormente.