

Desafío Antel: El Pulso del Streaming – Predecir el Éxito de la Sesión

Visión General

¡Bienvenidos al "Desafío Antel: El Pulso del Streaming"! Antel, líder en telecomunicaciones en Uruguay, te invita a sumergirte en el fascinante mundo del consumo de contenido digital. En esta era de streaming en auge, comprender el comportamiento del usuario es más crítico que nunca. ¿Qué impulsa a un usuario a quedarse y disfrutar de un contenido, o a abandonar una sesión?

Este desafío te propone analizar datos reales (anonimizados) de una aplicación de streaming híbrida (canales en vivo y contenido bajo demanda - VOD) para desentrañar los factores que definen una sesión exitosa. Tu misión será construir un modelo predictivo que Antel podrá utilizar para optimizar la experiencia del usuario, reducir el abandono y, en última instancia, aumentar la satisfacción y el engagement.

Objetivo del Desafío

El objetivo principal es desarrollar un modelo de clasificación binaria que, en base a datos de eventos de la sesión disponibles hasta el momento de la decisión de "play" (o el fin de la sesión si no hay play), pueda predecir si el usuario dará play a cualquier contenido (vivo o VOD) (es decir, si la sesión contendrá al menos un evento `click_play_button`) o si finalizará la sesión sin ninguna reproducción.

Más allá de la precisión predictiva, este desafío busca:

- **Identificar los drivers clave:** Descubrir cuáles son las variables (o combinaciones de variables y patrones de eventos) que más influyen en la decisión de un usuario de interactuar con el contenido.
- **Generar insights accionables:** Proponer recomendaciones concretas y estratégicas para Antel, basadas en tus hallazgos, que permitan mejorar la retención y el engagement.
- **Fomentar la explicabilidad:** Comprender el "porqué" detrás de las predicciones del modelo es tan importante como la predicción misma.

Descripción del Dataset

Se proporcionará un dataset de entrenamiento (y uno de testing para evaluación) que simula

eventos individuales recolectados de Google Analytics 4 (GA4). Cada fila representa un evento único, y la información de la sesión y el usuario se repite para cada evento dentro de la misma sesión.

Variables del Dataset:

Las variables disponibles corresponden a una extracción de datos de Google Analytics 4 y se presentarán con la siguiente estructura. Tenga en cuenta que muchos campos son anidados y la información relevante para su análisis se encuentra en combinación de event_params.key, event_params.value, user_properties.key, user_properties.value y items.

Nombre de la Variable	Tipo de Dato	Descripción
event_date	Date	Fecha del evento en formato AAAA MM DD.
event_timestamp	Integer (microsegundos)	Marca de tiempo del evento en microsegundos desde la época.
event_name	String	Nombre del evento (ej., user_engagement, session_start, screen_view, play).
event_params.key	String	Nombre de un parámetro del evento.
event_params.value.string_value	String	Valor de un parámetro del evento (tipo string).
event_params.value.int_value	Integer	Valor de un parámetro del evento (tipo entero).
event_params.value.float_value	Float	Valor de un parámetro del evento (tipo flotante).
event_params.value.double_value	Double	Valor de un parámetro del evento (tipo doble).
event_previous_timestamp	Integer (microsegundos)	Marca de tiempo del evento inmediatamente anterior.
event_bundle_sequence_id	Integer	Identificador de la secuencia de eventos dentro de un bundle.
event_server_timestamp_offset	Integer	Diferencia en milisegundos entre el timestamp del evento y el timestamp del servidor.
user_pseudo_id	String	ID pseudo-anónimo del usuario (identificador primario de GA4).

privacy_info.analytics_storage	String	Indica si el almacenamiento de analytics está permitido (Yes/No).
privacy_info.ads_storage	String	Indica si el almacenamiento de anuncios está permitido (Yes/No).
privacy_info.uses_transient_token	String	Indica si se usa un token transitorio para la sesión (Yes/No).
user_properties.key	String	Nombre de una propiedad de usuario.
user_properties.value.string_value	String	Valor de una propiedad de usuario (tipo string).
user_properties.value.int_value	Integer	Valor de una propiedad de usuario (tipo entero).
user_properties.value.float_value	Float	Valor de una propiedad de usuario (tipo flotante).
user_properties.value.double_value	Double	Valor de una propiedad de usuario (tipo doble).
user_properties.value.set_timestamp_micros	Integer (microsegundos)	Marca de tiempo en que se estableció la propiedad de usuario.
user_first_touch_timestamp	Integer (microsegundos)	Marca de tiempo de la primera interacción del usuario.
device.category	String	Categoría del dispositivo (ej., 'mobile', 'smart tv', 'desktop').
device.mobile_brand_name	String	Marca del dispositivo móvil.
device.mobile_model_name	String	Modelo del dispositivo móvil.
device.mobile_marketing_name	String	Nombre de marketing del dispositivo móvil.
device.mobile_os_hardware_model	String	Modelo de hardware del sistema operativo móvil.
device.operating_system	String	Sistema operativo del dispositivo.
device.operating_system_version	String	Versión del sistema operativo.
device.language	String	Idioma configurado en el dispositivo.
device.is_limited_ad_tracking	String	Indica si el seguimiento de anuncios está limitado (Yes/No).
device.time_zone_offset_seconds	Integer	Desplazamiento de la zona horaria del dispositivo en segundos.

geo.city	String	Ciudad de donde proviene la actividad del usuario.
geo.country	String	País de donde proviene la actividad del usuario.
geo.continent	String	Continente de donde proviene la actividad del usuario.
geo.region	String	Región (estado/provincia) de donde proviene la actividad del usuario.
geo.sub_continent	String	Subcontinente de donde proviene la actividad del usuario.
app_info.id	String	Identificador de la aplicación.
app_info.version	String	Versión de la aplicación.
app_info.install_store	String	Tienda de instalación de la aplicación (ej., 'Google Play').
app_info.firebase_app_id	String	ID de la aplicación en Firebase.
app_info.install_source	String	Fuente de instalación de la aplicación.
traffic_source.name	String	Nombre de la fuente de tráfico.
traffic_source.medium	String	Medio de la fuente de tráfico (ej., 'organic', 'cpc').
traffic_source.source	String	Fuente de tráfico (ej., 'google-play', 'direct').
stream_id	String	ID del stream de datos.
platform	String	Plataforma del dispositivo (ej., 'ANDROID', 'IOS', 'WEB').
is_active_user	Boolean	Indica si el usuario es un usuario activo.
batch_event_index	Integer	Índice del evento dentro de un lote.
items.item_id	String	ID del elemento interactuado (ej., ID de contenido).
items.item_name	String	Nombre del elemento interactuado (ej., nombre del canal o VOD).
items.item_category	String	Categoría del elemento interactuado (ej., 'CANAL', 'VIDEO').

Dataset de Ejemplo (Archivo train_data_sample.csv):

Fragmento de código

```
event_date,event_timestamp,event_name,event_params.key,event_params.value.string_value,event_params.value.int_value,device.category,device.mobile_brand_name,device.mobile_model_name,device.operating_system,device.operating_system_version,geo.city,geo.country,geo.region,app_info.id,app_info.version,app_info.install_store,traffic_source.medium,traffic_source.source,platform,user_pseudo_id,user_id,items.item_id,items.item_name,items.item_category
20250522,1747900872000000,user_engagement,firebase_screen_class,MainActivity,,,,Android,Android 12,Montevideo,Uruguay,Montevideo
Department,uy.com.antel.androidtv.veratv,2.1.1,,(none),(direct),ANDROID,67948b2823fb2c9ec841e278d8e85eac,
,,,,
20250522,1747904795000000,session_start,ga_session_id,,1747904795,mobile,Honor,WDY-LX3,Android,Android id 14,Paysandu,Uruguay,Paysandu
Department,uy.com.adinet.adinetv,3.4.2,com.android.vending,organic,google-play,ANDROID,ed7f0a9e6571ed77504ba77f006b2211,,,,
20250522,1747904795000000,screen_view,firebase_screen_class,SplashActivity,,,,Android,Android 14,Paysandu,Uruguay,Paysandu
Department,uy.com.adinet.adinetv,3.4.2,com.android.vending,organic,google-play,ANDROID,ed7f0a9e6571ed77504ba77f006b2211,,,,
20250522,1747904795000000,screen_view,firebase_screen_class,MainActivity,,,,Android,Android 14,Paysandu,Uruguay,Paysandu
Department,uy.com.adinet.adinetv,3.4.2,com.android.vending,organic,google-play,ANDROID,ed7f0a9e6571ed77504ba77f006b2211,,,,
20250522,1747904795000000,screen_view,firebase_screen_class,SinglePlayerActivity,,,,Android,Android 14,Paysandu,Uruguay,Paysandu
Department,uy.com.adinet.adinetv,3.4.2,com.android.vending,organic,google-play,ANDROID,ed7f0a9e6571ed77504ba77f006b2211,,,,
20250522,1747904795000000,play,item_id,,,,Android,Android 14,Paysandu,Uruguay,Paysandu
Department,uy.com.adinet.adinetv,3.4.2,com.android.vending,organic,google-play,ANDROID,ed7f0a9e6571ed77504ba77f006b2211,,2snj7,Canal 10 | Uruguay,CANAL
20250522,1747904825000000,session_start,ga_session_id,,1747904825,mobile,Apple,iPhone 11,iOS,iOS 18.1.1,Montevideo,Uruguay,Montevideo
Department,uy.com.antel.VeraTV,3.4.2,iTunes,,(direct),IOS,B669E51C00654D11A762EBF898C2BDFC,,,,
20250522,1747904825000000,screen_view,firebase_screen_class,SplashViewController,,,,iOS,iOS 18.1.1,Montevideo,Uruguay,Montevideo
Department,uy.com.antel.VeraTV,3.4.2,iTunes,,(direct),IOS,B669E51C00654D11A762EBF898C2BDFC,,,,
20250522,1747904825000000,user_engagement,firebase_screen_class,SplashViewController,,,,iOS,iOS 18.1.1,Montevideo,Uruguay,Montevideo
Department,uy.com.antel.VeraTV,3.4.2,iTunes,,(direct),IOS,B669E51C00654D11A762EBF898C2BDFC,,,,
20250522,1747904825000000,screen_view,firebase_screen_class,HomeViewController,,,,iOS,iOS
```

```

18.1.1, Montevideo, Uruguay, Montevideo
Department, uy.com.antel.VeraTV, 3.4.2, iTunes,, (direct), IOS, B669E51C00654D10D4647A,,,,,
20250522, 1747904825000000, user_engagement, firebase_screen_class, HomeViewController,,,,, iOS, iOS
18.1.1, Montevideo, Uruguay, Montevideo
Department, uy.com.antel.VeraTV, 3.4.2, iTunes,, (direct), IOS, B669E51C00654D10D4647A,,,,,
20250522, 1747904819000000, session_start, ga_session_number,, 105,, Android, Android 10,, Uruguay, Colonia
Department, uy.com.adinet.adinetTV, 3.4.2, com.android.vending, organic, google-play, ANDROID, 3c662cd33796a83
9af005b06e70d1915,,,,,
20250522, 1747904819000000, screen_view, firebase_screen_class, ChannelDetailsActivity,,,,, Android, Android
10,, Uruguay, Colonia
Department, uy.com.adinet.adinetTV, 3.4.2, com.android.vending, organic, google-play, ANDROID, 3c662cd33796a83
9af005b06e70d1915,,,,,
20250522, 1747904819000000, view_item, item_id, 2snj7,,,,, Android, Android 10,, Uruguay, Colonia
Department, uy.com.adinet.adinetTV, 3.4.2, com.android.vending, organic, google-play, ANDROID, 3c662cd33796a83
9af005b06e70d1915, 2snj7, Canal 10 | Uruguay, CANAL

```

Notas Importantes sobre el Dataset y el Desafío:

- Estructura de Eventos:** El dataset contiene una secuencia de eventos por sesión. Deberán realizar Feature Engineering para transformar esta secuencia de eventos en características a nivel de sesión que su modelo pueda utilizar. Por ejemplo, podrían contar el número de eventos antes de un play, la última categoría vista antes de la decisión, o el tiempo transcurrido hasta la primera interacción significativa. Las propiedades de usuario (user_properties) y los parámetros de evento (event_params) son campos anidados que contienen información valiosa para la construcción de features. Por ejemplo, ga_session_id (event_params.value.int_value cuando event_params.key es 'ga_session_id') y user_pseudo_id pueden usarse para identificar sesiones únicas. Las categorías de contenido y los IDs de los elementos interactuados se encuentran en los campos items.item_id, items.item_name y items.item_category.
- Definición de dio_play (Variable Objetivo):** La variable objetivo dio_play no está incluida directamente en el dataset. En su lugar, se considerará que una sesión dio_play (valor 1) si contiene al menos un evento con event_name igual a play (o click_play_button si este evento es el indicador en los datos proporcionados). Si una sesión no contiene este evento, se considerará que no dio_play (valor 0).
- Límite de Datos:** Para el dataset de entrenamiento, se incluyen todos los eventos de la sesión hasta el momento en que se produce un evento de play (si ocurre) o todos los

eventos de la sesión si no se produce ningún play. Esto asegura que no haya data leakage de eventos futuros.

- **Dataset de Testing:** El dataset de testing tendrá la misma estructura de eventos por sesión, pero no incluirá la columna `dio_play`. Su modelo deberá predecir esta variable para cada combinación única de `user_pseudo_id` y el `ga_session_id` correspondiente en el testing set.

Métricas de Evaluación

La evaluación de los equipos se realizará mediante una combinación de métricas automáticas y una evaluación cualitativa por un jurado experto.

1. Métrica Principal (Automática)

F1 Score: El F1 Score será la métrica principal para clasificar las predicciones. Es una métrica robusta para problemas de clasificación binaria con posibles desbalances de clases, ya que considera tanto la precisión como el *recall*.

- **Entrega:** Cada equipo deberá enviar un archivo CSV con las predicciones para el dataset de testing (identificador de sesión, `dio_play_predicho`). El código utilizado para generar las predicciones también debe ser entregado (ej., un notebook Jupyter o scripts de Python/R).
- **Ponderación en el Puntaje Final:** El F1 Score representará el **70% del puntaje total final**.

2. Evaluación Adicional (por Jurado)

Un panel de expertos de Antel evaluará los trabajos basándose en los siguientes criterios, que representan el **30% restante del puntaje total final**:

- **Calidad del Modelo y Enfoque de Machine Learning:**
 - Uso apropiado de técnicas de ML.
 - Robustez y eficiencia del modelo.
 - Creatividad en el Feature Engineering.
- **Análisis de Datos y Descubrimiento de Insights:**

- Profundidad del análisis exploratorio (EDA).
- Relevancia y originalidad de los insights.
- Claridad en la presentación de hallazgos.

- **Recomendaciones Accionables y Estratégicas:**

- Viabilidad y aplicabilidad de las recomendaciones.
- Impacto potencial para Antel.
- Coherencia con los objetivos del negocio.

- **Calidad de la Presentación y Código:**

- Organización y claridad del notebook/código.
- Documentación y comentarios.
- Reproducibilidad de los resultados.