



Vigilada Mineducación

ANALIZANDO PATRONES DE ÉXITO EN YOUTUBE: UN SISTEMA DE  
RECOMENDACIÓN PARA CREADORES DE CONTENIDOS EDUCATIVOS

VANESSA OSORIO URREA

Tesis

Asesor

Santiago Ortiz Arias

Co Asesor

Francisco Javier del Castillo Cortázar

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2024

# Analizando patrones de éxito en YouTube: un sistema de recomendación para creadores de contenidos educativos

**Vanessa Osorio Urrea**

*C.C. 1128389743*

*vosoriou@eafit.edu.co*

**Director: Santiago Ortiz**

*sortiza2@eafit.edu.co*

*Escuela de Ciencias Aplicadas e Ingeniería, Universidad EAFIT, Medellín, Colombia*

**Co-Director: Francisco Javier del Castillo Cortázar**

*delcavideography@gmail.com*

*Director de Investigación y Diseño de UX para Latam,*

*Ipsos, Medellín, Colombia*

## Resumen

Los modelos de recomendación, son técnicas de Inteligencia Artificial que se utilizan para predecir y recomendar productos, servicios o contenidos que podrían interesar a un usuario en particular. Estos modelos, se basan en el uso de algoritmos de aprendizaje automático, en donde se analizan los datos recolectados y se generan recomendaciones personalizadas. Estos sistemas de recomendación, se utilizan comúnmente en plataformas de comercio electrónico, redes sociales y servicios de streaming, entre otros. YouTube, es una plataforma de contenidos digitales, que dentro de su clase, es la que cuenta con mayor popularidad. En ésta, interactúan dos tipos de personas: quienes consumen contenidos, y quienes los crean. Este documento se centrará en los usuarios que crean contenidos educativos para esta plataforma. A pesar de que existen muchos estudios y análisis sobre cómo funciona el algoritmo de YouTube, para los creadores de contenidos que monetizan sus canales, es incierto saber si sus contenidos serán exitosos y tendrán el retorno económico esperado con base en su esfuerzo y tiempo dedicado. Por lo tanto, este proyecto se enfoca en el análisis de la información pública de canales exitosos en YouTube dedicados al Diseño Gráfico, Audiovisual y Fotográfico, con el propósito de aplicar modelos de recomendación que ayuden al creador de contenidos DelcaVideography a tomar decisiones informadas sobre qué contenidos publicar. El objetivo, es identificar patrones y variables relacionadas con el interés de su audiencia en el mercado, donde se busca maximizar el impacto y el retorno de su trabajo.

*Palabras clave:* Sistemas de recomendación, Clustering, Filtrado colaborativo, YouTube, Canales educativos, Efecto de recomendaciones para crear contenidos audiovisuales.

## 1. Introducción

En el marco de la Inteligencia Artificial, se encuentran los Sistemas de Recomendación (RS por sus siglas en inglés) los cuales como propone Ricci et al. (2011a), se “han beneficiado de los resultados obtenidos en varios campos de la informática, especialmente el aprendizaje automático y la extracción de datos, la recuperación de información y la interacción humano-computadora”, al contribuir en el manejo de diferentes volúmenes de información,

ayudar en el almacenamiento y búsqueda de datos y permitir que las computadoras aprendan a realizar determinadas tareas de una forma óptima (Ricci et al., 2011a). Los RS son herramientas y técnicas de software que ofrecen a los usuarios recomendaciones de elementos que puedan serles útiles. Estas, se relacionan con diferentes tipos de decisiones como la selección de productos a comprar, elección de música a escuchar o noticias a leer (Ricci et al., 2011a).

El estudio de los RS es relativamente nuevo, en comparación de otras investigaciones tecnológicas. Su interés, como área de investigación, nace a mediados de la década de 1990 (Ricci et al., 2011a), en donde en 1994 Resnick et al. (1994) publicó el primer artículo de investigación moderno sobre sistemas de recomendación, en el cual, haciendo uso del concepto de Filtrado Colaborativo, expone un ejemplo para identificar los artículos más relevantes de un grupo de noticias (Abel et al., 2016). Para este objetivo, con su grupo de estudio, desarrollaron GroupLens, una arquitectura abierta para el filtrado colaborativo de noticias en la web, en donde se basaron en la idea de que es probable que las personas que hayan estado de acuerdo con una valoración subjetiva del pasado, puedan estarlo con otra del futuro (Resnick et al., 1994).

En 2006, aparece la iniciativa *The Netflix Prize*, la cual aportó a grandes avances en este campo durante sus 3 años de ejecución, como por ejemplo, el establecimiento de métodos de factorización de matrices como SVD, que se convirtió en el estado del arte de los modelos de recomendación (Abel et al., 2016). *The Netflix Prize* consistió en el reto que lanzó Netflix cuando publicó 100 millones de calificaciones anónimas de películas, para motivar a la comunidad analítica a desarrollar un sistema de recomendación que tuviera mejor desempeño frente a su sistema actual, Cinematch (Bennett and Lanning, 2007). Desde 2007, se realiza anualmente el principal evento de investigación y aplicación en la conferencia ACM Recommender Systems (RecSys), junto a la cual, también se lleva a cabo la competencia anual “RecSys Challenge”, que cada vez trae mayores contribuciones a las investigaciones de Sistemas de Recomendación (Abel et al., 2016). Para los siguientes años, y en la actualidad, el interés por estos sistemas continua siendo muy utilizado (Ricci et al., 2011a).

Cuando se habla de RS, existen diferentes técnicas que pueden aplicarse según el objetivo y las características de los datos que sirven como insumo, las cuales se mencionarán a continuación:

### **Basado en Contenido (Content-Based Recommender System o CBRS):**

El sistema aprende de elementos similares que le han gustado a un usuario en el pasado para hacer recomendaciones futuras (Ricci et al., 2011a). Las preferencias de los usuarios se comparan con el contenido de los elementos, las cuales pueden extraerse explícitamente de cuestionarios o aprendiendo de la interacción de un usuario con una plataforma (Guruge et al., 2021).

## **Filtro Colaborativo (Collaborative Filtering-Based Recommendation System o CFRS o CF):**

Es la técnica más popular e implementada de RS. Su implementación más simple, recomienda a un usuario los artículos que a otros usuarios con gustos similares, le han gustado en el pasado. Estas recomendaciones las hace basado en el historial de calificaciones de los usuarios (Ricci et al., 2011a). Para esto, requieren construir una matriz de elementos (perfil de los usuarios) con las preferencias para varios tipos de elementos, con el objetivo de identificar otros usuarios con gustos similares en el pasado y usar sus opiniones para recomendar elementos a otros usuarios con gustos similares (Guruge et al., 2021).

Según lo describen en el documento de Breese et al. (1998), esta técnica se divide en dos tipos:

- **Basadas en Memoria:** Se busca predecir los votos de un usuario particular con relación a un conjunto de elementos.
- **Basadas en Modelos:** En este caso, se espera predecir votos de un usuario sobre un elemento no conocido.

## **Demográfico:**

En este tipo, se hacen recomendaciones basado en el perfil demográfico del usuario. Se parte de la idea que se deben crear diferentes recomendaciones para diferentes nichos demográficos (Ricci et al., 2011a).

## **Basado en Conocimiento (Knowledge-Based Recommender System o KBRS):**

Estos utilizan el conocimiento especializado en un dominio específico para sugerir artículos que se ajusten a las necesidades y preferencias de los usuarios. Asociados a estos, se encuentran los sistemas basados en Restricciones que hacen uso de reglas explícitas para relacionar con las características del artículo a recomendar. Esta técnica tiende a funcionar muy bien al inicio, pero si no se acompaña por componentes de aprendizaje, puede ser superada por otras técnicas con CF (Ricci et al., 2011a). Su principal inconveniente es la dificultad de adquisición y representación del conocimiento del dominio en una estructura legible por una máquina y también necesita grandes esfuerzos para extraer conocimiento (Guruge et al., 2021).

## **Basado en la Comunidad:**

También conocido como Sistema de Recomendación social, se basa en las preferencias de los amigos de los usuarios para hacer recomendaciones. Se dice que las personas tienden a confiar más en las recomendaciones de sus amigos que en las de perfiles similares pero anónimos (Ricci et al., 2011a).

### **Enfoques de Minería de Datos (Data Mining o DM):**

Usan técnicas de ML y DL en donde se han identificado algoritmos compatibles con RS como enfoques bayesianos, árboles de decisión, redes neuronales, basados en vecinos, clústers y basados en descenso de gradiente. Un ejemplo de ML, se puede ver en la aplicación de *Independent Bayesian classifier combination* (IBBC por sus siglas en inglés). En esta, a pesar de la demanda computacional que exigen los modelos Bayesianos, usan la técnica *variational approximation* (que es mucho más eficiente) con el objetivo de identificar si un clasificador puede mejorar la gestión de los analistas de inversiones (Bew et al., 2019). También incluye el enfoque de Minería de reglas de asociación (*Association Rule Mining* o ARM) que a su vez incluye 3 categorías: ARM tradicional, ARM de soporte adaptativo (ASARM) y ARM evolutivo (Guruge et al., 2021).

### **Sistemas de Recomendación Híbridos:**

Se basa en la combinación de 2 o más de las técnicas descritas anteriormente, en donde trata de usar las ventajas de una para cubrir las desventajas de otra (Ricci et al., 2011a). Algunos ejemplos de combinaciones son CFRS y CBRs mediante métodos como la combinación ponderada, conmutada, mixta y en cascada. También se conocen combinaciones de DM y CFRS, en donde DM se usa para analizar y encontrar patrones ocultos en los comportamientos de los usuarios para crear recomendaciones personalizadas (Guruge et al., 2021).

También existen otros sistemas de recomendación como RS conversacional (CRS), RS basado en estadísticas y RS basado en recomendación de paquetes de artículos (Guruge et al., 2021).

En la actualidad, como lo mencionan en el blog Hootsuite escrito por Kemp (2022), YouTube es una de las plataformas de contenidos digitales con mayor popularidad y audiencia a nivel mundial. En ella, interactúan millones de usuarios que consumen y crean contenidos de manera constante. Entre estos, se encuentran los creadores de contenidos educativos o edutubers. Los edutubers son creadores de contenido audiovisual educativo en YouTube (Pattier, 2022), quienes, a pesar de la incertidumbre que pueda existir en el éxito de sus videos, buscan generar ingresos a partir de sus contenidos y monetizar sus canales.

Según el estudio realizado por Oliva (2022), la producción de contenidos para YouTube es una tarea que requiere una gran cantidad de tiempo y esfuerzo por parte de los creadores. A pesar de la dedicación que ponen en la elaboración de contenidos con buena calidad en edición y producción, existe una gran incertidumbre respecto a su éxito en la plataforma. Esta incertidumbre se debe, en parte, a la alta competencia existente en el medio y a la gran cantidad de variables desconocidas que influyen en la aceptación y popularidad de un contenido en particular. Esto genera una preocupación constante entre los creadores de contenidos, quienes necesitan asegurarse de que sus esfuerzos sean recompensados con una buena acogida de su audiencia. Por lo tanto, el objetivo de este trabajo es identificar patrones y variables relacionadas con el éxito de los canales educativos más exitosos en YouTube, para desarrollar una sistema de recomendación que sirva como insumo para los creadores de contenidos. De esta manera, se busca entregarles una herramienta que les

permita aumentar la probabilidad de éxito de sus contenidos en la plataforma y, por ende, mejorar su retorno económico.

Jena et al. (2023) describe la aplicación de un RS de filtrado colaborativo (CF) para la recomendación de cursos *E-learning* (basadas en las preferencias de los usuarios), en donde de una gran cantidad de cursos disponibles, automatizan el proceso de selección para los usuarios de una plataforma digital. Esta aplicación es un ejemplo de como los modelos de recomendación pueden ser útiles en el contexto de creadores de contenidos digitales. De manera similar, pero desde el punto de vista de los creadores de contenidos de YouTube, los RS podrían ayudar a reducir la incertidumbre en la producción de contenidos al analizar patrones y variables que influyen en el éxito de los contenidos publicados por otros edutubers y realizar recomendaciones personalizadas para un usuario particular (en este caso un creador de contenidos), aumentando así la probabilidad de éxito en la plataforma. Por lo tanto, se evidencia la importancia de desarrollar un sistema de recomendación que permita a los edutubers aumentar la probabilidad de éxito de sus contenidos en YouTube, mejorar su retorno económico y, por ende, incentivar la producción de contenidos de alta calidad en la plataforma. De esta manera, se contribuiría a fortalecer el ecosistema de contenidos educativos en YouTube y se generarían beneficios tanto para los creadores como para los usuarios que los consumen.

En este trabajo se propone la aplicación de las técnicas de RS de filtrado Colaborativo a un grupo de Canales de YouTube, creadores de contenidos educativos de diseño gráfico, audiovisual y fotográfico con el objetivo de encontrar recomendaciones de contenidos a publicar por el Canal DelcaVideography. Para esto, el documento está organizado de la siguiente forma: En la Sección 2 se encuentra la Metodología usada, pasando por la descripción, preparación de los datos y aplicación de los modelos. En la Sección 3 se muestran los resultados de la aplicación de la metodología, en la Sección 4 se analizan y discuten los resultados, en la Sección 5 se concluyen los hallazgos y se mencionan algunos puntos por trabajar.

## 2. Metodología

### 2.1. Descripción de los Datos

Basados en el criterio experto del dueño del canal Delcavideography, se analizan y seleccionan 8 canales con características similares a las de él, es decir dedicados a la creación de contenidos educativos de Diseño gráfico, audiovisual y fotográfico. Estos canales en algunos casos son más exitosos o presentan comportamientos similares. Para obtener los datos, se desarrolla un proceso de Web Scraping con el cual se descarga la información pública de estos canales de forma diaria entre el 2 y el 31 de Agosto de 2023. Esta información contiene el título del video, fecha de creación, duración del video, la cantidad de visualizaciones, likes y dislikes para cada video.

Además se construye una tabla general con las características analizadas de cada canal en la revisión inicial, como lo son el nombre del canal, fecha de creación, país de origen,

cantidad de videos publicados, idioma y cantidad de suscriptores. También se construye un tercer insumo a partir de la información publicada en la página web Social Blade, en donde de forma general se pueden extraer la cantidad de visualizaciones, suscriptores e ingresos estimados para cada canal por cada uno de los días de la ventana de tiempo de análisis. Finalmente se crea un cuarto insumo al extraer los comentarios para cada uno de los videos de los canales seleccionados.

## 2.2. Preparación de Datos

Se consolida una estructura con las variables obtenidas en la vista general, tabla general y Scraping de videos descritos anteriormente. En cuanto a la información de Comentarios, se construyen 2 fuentes para complementar los datos, una que agrupa la cantidad de comentarios por cada video y otra en la que con el detalle de los comentarios se crea un Análisis de Sentimientos para obtener un resultado de polaridad por cada video, identificando si los comentarios fueron positivos, neutros o negativos.

Adicionalmente, con el objetivo de robustecer el conjunto de datos y tener información adicional que describa mejor las características de cada uno de los videos (ítems) se ajusta el contenido de la variable “*Duración video*” para expresarla en formato de tiempo. Así mismo, se realiza el cálculo de nuevas variables como lo son: día de la semana de publicación, días y meses de creación del video. Esta última variable se usa para excluir los contenidos con publicación menor o igual a 3 días, ya que son pocos los videos con estas características que se lograron identificar en la ventana de tiempo analizada, y que de acuerdo con la validación con el experto del canal, se sabe pueden tener métricas más altas y por lo tanto distorsionar las conclusiones finales. Esto sucede porque en la plataforma cuando un video nace y durante sus primeras 72 horas, YouTube los promociona a mayor cantidad de usuarios para aprender de estos contenidos nuevos e identificar la aceptación de la audiencia.

Con el uso de expresiones regulares de Python, se hace una análisis de texto para calcular el tema general del contenido basado en el título del video. Teniendo en cuenta que los `item_id` con los que se cuenta son identificadores únicos en YouTube, se usa el resultado del tema general del video para calcular un nuevo `item_id`, de forma que puedan encontrarse relaciones entre los usuarios (canales) y los contenidos (ítems), de la siguiente forma: Por cada canal y tema encontrado, se ordenan los contenidos de mayor a menor visualizaciones y se asigna un consecutivo para integrarlo al nuevo campo de `item_id`. Por ejemplo si encuentra un tema que sea LOGO, se tendrán diferentes `item_id` según la cantidad de contenidos publicados de este tema por cada canal de la siguiente forma: LOGO\_1, LOGO\_2... LOGO\_n, en donde LOGO\_1 será el contenido con mayores visualizaciones para cada canal.

Teniendo en cuenta que para aplicar los modelos de recomendación de filtrado colaborativo se necesita un rating para cada ítem (video) asociado a un usuario (canal), y tomando como referencia el libro de Jauset (2007) sobre la forma de calcular un rating, se realiza este cálculo basado en las visualizaciones de cada video (ítem) y el canal en general. Para esto se aplica la siguiente ecuación a cada punto de observación de los datos:

$$Rating = \frac{\text{Visualizaciones video}_t - \text{Visualizaciones video}_{t-1}}{\text{Universo de visualizaciones Canal}_t - \text{Universo de visualizaciones Canal}_{t-1}}$$

Donde  $t$  corresponde a la fecha actual y  $t - 1$  corresponde a la fecha anterior.

Adicionalmente, se realiza un Análisis de Componentes Principales (PCA por sus siglas en inglés) para identificar la importancia y peso de las diferentes variables en la varianza de los datos. Esto con el objetivo de complementar el set de datos para las recomendaciones finales (para mayor detalle ver Sección A del Material Suplementario).

De forma general y de acuerdo al análisis de PCA, con el dueño del canal Delcavideography, se identifican dos acciones importantes para tener en cuenta en la creación de contenidos: Incentivar a que los usuarios que ven los videos dejen sus comentarios, así mismo como explorar contenidos de interés para el público de los países India y Pakistán.

Finalmente, y teniendo en cuenta que para el entrenamiento de los modelos la combinación de usuarios y elementos no deben repetirse, se agrupan los datos calculando un rating mediano para cada canal—video, quedando así con 5,537 registros (videos).

### 2.3. Modelos Aplicados

Luego de la preparación de datos, se evalúan los modelos KNN Baseline, CoClustering, SVD y SVD++ de la librería Surprise de Python, con el objetivo de comparar sus resultados y elegir la mejor predicción.

Los algoritmos K-Nearest Neighbors (KNN por sus siglas en inglés) Baseline se utilizan para modelar y corregir las desviaciones sistemáticas de usuarios e ítems en las calificaciones, permitiendo así una estimación más precisa de las puntuaciones y mejorando la calidad de las recomendaciones en este tipo de sistemas (Ricci et al., 2011b).

En el documento de Koren (2010) se describe su ecuación de la siguiente forma:

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{j \in S^k(i;u)} S_{ij}(r_{uj} - b_{uj})}{\sum_{j \in S^k(i;u)} S_{ij}}$$

Donde  $\hat{r}_{ui}$ , es la calificación no observada por el usuario  $u$  para el ítem  $i$ .  $S^k(i; u)$  representa el conjunto de  $k$  vecinos, en la cual usando una medida de similitud se identifican los  $k$  ítems calificados por el usuario  $u$  que son más similares al ítem  $i$ . Finalmente, el valor predicho de  $\hat{r}_{ui}$  se toma como un promedio ponderado de las calificaciones de los ítems vecinos, al mismo tiempo que se ajusta por los efectos del usuario y del ítem a través de las estimaciones de la línea base  $b_{ui}$ .

Por su parte,  $b_{ui}$  está definido de la siguiente forma:

$$b_{ui} = \mu + b_u + b_i$$

Donde los parámetros  $b_u$  y  $b_i$  indican las desviaciones observadas del usuario  $u$  y del ítem  $i$ , respectivamente, respecto al promedio de calificaciones  $\mu$  (Koren, 2010).

En cuanto al algoritmo CoClustering este es un algoritmo de filtrado colaborativo que se basa en una técnica llamada co-clustering. Esta técnica se utiliza para manejar de manera eficiente datos que son de alta dimensionalidad o escasos. La idea principal de este algoritmo es agrupar al mismo tiempo las columnas y las filas de una matriz. A diferencia de los métodos de agrupamiento tradicionales, el co-clustering busca identificar grupos o bloques de filas y columnas que están relacionados y muestran un rendimiento similar en un subconjunto específico de columnas, o viceversa (Nguyen et al., 2023). Los clústeres se asignan utilizando un método de optimización directa, similar al enfoque utilizado en el algoritmo k-means (Hug, 2015b).

Su ecuación está definida de la siguiente forma (Hug, 2015b):

$$\hat{r}_{ui} = \bar{C}_{ui} + (\mu_u - \bar{C}_u) + (\mu_i - \bar{C}_i)$$

Donde  $\hat{r}_{ui}$  es la calificación que un usuario  $u$  daría a un elemento  $i$ .  $\bar{C}_{ui}$  representa el promedio de calificación de la co-clusterización  $\bar{C}_{ui}$ .  $\bar{C}_u$  es el promedio de calificación del clúster de usuario  $u$ .  $\bar{C}_i$  es el promedio de calificación del clúster de elemento  $i$ . Si el usuario  $u$  es desconocido (es decir, no se tiene información sobre él), la predicción es igual al promedio de calificación de todos los elementos, que es representado por  $\mu_i$ . Si el elemento  $i$  es desconocido (es decir, no se tiene información sobre él), la predicción es igual al promedio de calificación de todos los usuarios en el mismo clúster de usuario que  $u$ , que es representado por  $\mu_u$ . Si tanto el usuario  $u$  como el elemento  $i$  son desconocidos, la predicción es igual al promedio de calificación global de todas las calificaciones disponibles, que es representado por  $\mu$ .

Como lo describe Ricci et al. (2011b), el algoritmo *Singular Value Decomposition* (SVD por sus siglas en inglés) se basa en la factorización de matrices, asignando usuarios y elementos a un espacio de factores latentes compartido de dimensionalidad  $f$ . Los modelos de factores latentes son un enfoque alternativo que intenta explicar las calificaciones asignadas a elementos al describir tanto a los elementos como a los usuarios en función de un conjunto de factores ocultos y que se derivan de los patrones de calificación observados en el sistema (Koren et al., 2009). Las interacciones entre usuarios y elementos se representan como productos internos en este espacio para explicar las calificaciones. Cada elemento tiene un vector de características  $q_i$ , y cada usuario tiene un vector que indica su interés en esas características  $P_u$ . Estos vectores reflejan la relación entre usuarios y elementos en términos de factores específicos. La ecuación del algoritmo se define en función de estos vectores y productos internos de la siguiente forma (Ricci et al., 2011b):

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T P_u$$

Donde el producto punto  $q_i^T P_u$  representa la interacción entre el usuario  $u$  y el elemento  $i$ , reflejando el interés general del usuario en las características del elemento. La calificación

final se calcula agregando predictores de referencia que dependen solo del usuario o del elemento. Para aprender los parámetros del modelo ( $b_u$ ,  $b_i$ ,  $P_u$  y  $q_i$ ), se minimiza el error cuadrado regularizado, de la siguiente forma:

$$\min_{b^*, q^*, p^*} \sum_{(u,i) \in \kappa} (r_{ui} - \mu - b_i - b_u - q_i^T P_u)^2 + \lambda_4 (b_i^2 + b_u^2 + \|q_i\|^2 + \|P_u\|^2)$$

En donde  $\lambda_4$  controla el grado de regularización. La minimización se realiza normalmente mediante el descenso de gradiente estocástico o el método de mínimos cuadrados alternantes.

Por su parte el algoritmo SVD++, propuesto por Koren (2008) el cual combina características de modelos de vecindad y factores latentes, es una mejora del algoritmo SVD al considerar también la retroalimentación implícita, la cual entrega indicios adicionales de las preferencias de los usuarios y por lo tanto ofrece una precisión superior a la del algoritmo SVD. Esto lo hace incluyendo otro conjunto de factores para los elementos representados por vectores  $y_i \in \mathbb{R}^f$ . Estos factores de elementos adicionales se emplean para describir a los usuarios según los elementos que han evaluado. El modelo se define de la siguiente manera (Ricci et al., 2011b):

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T (p_u + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j)$$

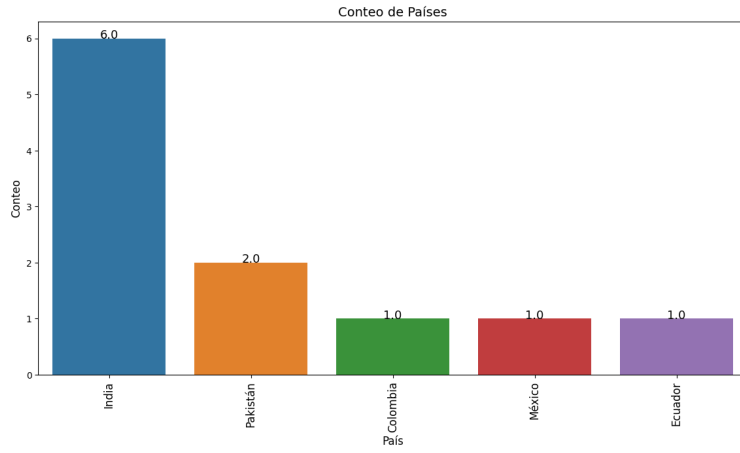
En donde el conjunto de elementos calificados por un usuario  $u$  se denota como  $R(u)$ . Un usuario  $u$  se modela como  $p_u + |R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j$ . El cual usa un vector de factores de usuario libre  $p_u$ , que se aprende a partir de las calificaciones explícitas proporcionadas. Este vector se complementa con la suma  $|R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} y_j$ , que representa la perspectiva de la retroalimentación implícita. Dado que los valores de  $y_j$  están centrados alrededor de cero (debido a la regularización), la suma se normaliza por  $|R(u)|^{-\frac{1}{2}}$ , con el fin de estabilizar su variabilidad a lo largo del rango de valores observados de  $|R(u)|$ .

De forma similar al algoritmo SVD, los parámetros del modelo se determinan minimizando la función de error cuadrado regularizado asociada mediante la técnica de descenso de gradiente estocástico.

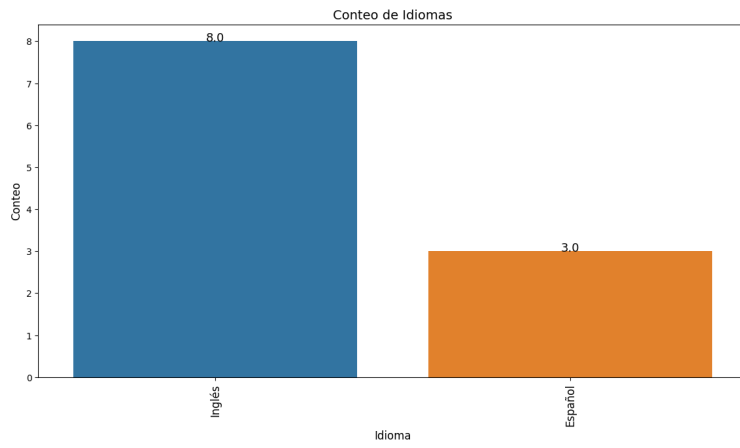
## 3. Resultados

### 3.1. Preparación de datos

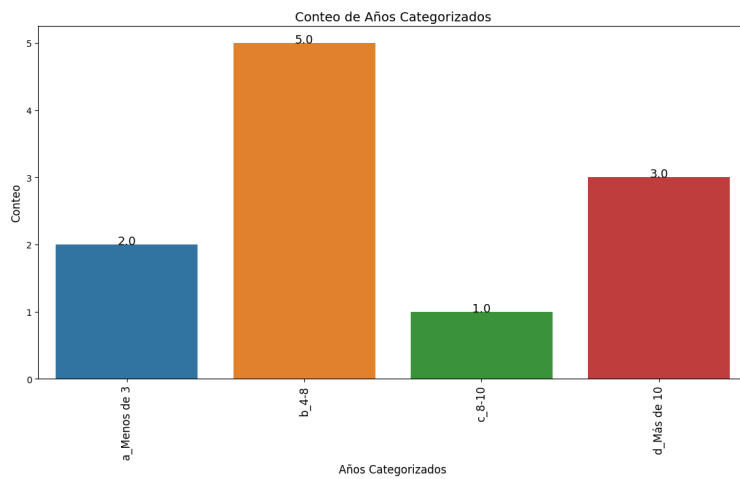
De acuerdo a los criterios descritos en la Sección 2 con la metodología para la selección de los canales, se identifica que la mayoría se encuentran localizados en la India y el idioma principal de los contenidos publicados es Inglés. Adicionalmente, se identifica que los canales seleccionados son canales con una buena trayectoria en YouTube, contando principalmente con más de 4 años de creación, como se observa en la Figura 1.



(a) Países Canales



(b) Idiomas Canales



(c) Años Creación Canales

Figura 1: Información Canales

Inicialmente se valida el comportamiento de las visualizaciones por cada canal tal cual como se descargó la información de YouTube y se logra identificar que el canal de Aslam\_Ahmed es el de mayores visualizaciones. También se hace un monitoreo del comportamiento de los 10 principales videos en el conjunto de datos, observando comportamientos muy lineales a lo largo del tiempo (para mayor detalle ver Sección B del Material Suplementario).

De acuerdo a lo observado en la información original, y teniendo en cuenta que la información de YouTube se descarga acumulada desde el momento en que se creó cada contenido, se valida el comportamiento del delta de visualizaciones por cada video y canal, encontrando los resultados por canales que se obvervan en la Figura 2.

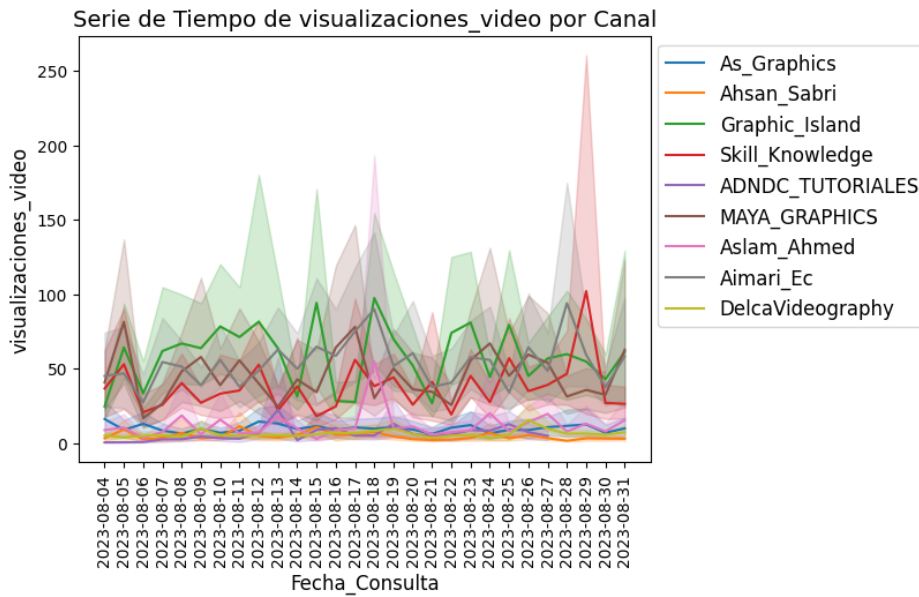


Figura 2: Delta Visualizaciones video por Canal

Para este caso, se identifican comportamientos más similares entre los canales, aún así destacándose el comportamiento del canal Graphic\_Island. También se grafican los 10 principales videos encontrando el comportamiento que se observa en la Figura 3.

Con lo anterior se identifican comportamientos diferenciados y no lineales en comparación con los datos acumulados de la historia, los cuales ayudarán a identificar de una forma más precisa los contenidos exitosos en una ventana de tiempo reciente. De acuerdo con esto, se decide utilizar esta información para continuar el entendimiento de los datos y planteamiento de los modelos.

Por otro lado, también se identifica que algunos canales tienen una velocidad de publicación más alta que el resto, y que incluso pueden llegar a crear varios contenidos durante el mismo día, este es el caso del canal Ashan\_Sabri que claramente excede la cantidad de contenidos publicados por el resto de canales analizados, como se observa en la Figura 4.

Al analizar la información de comentarios, se encuentra que la mayoría de videos (63%) tienen menos de 2 comentarios. Adicional, se identifica que el 80.2% de los comentarios

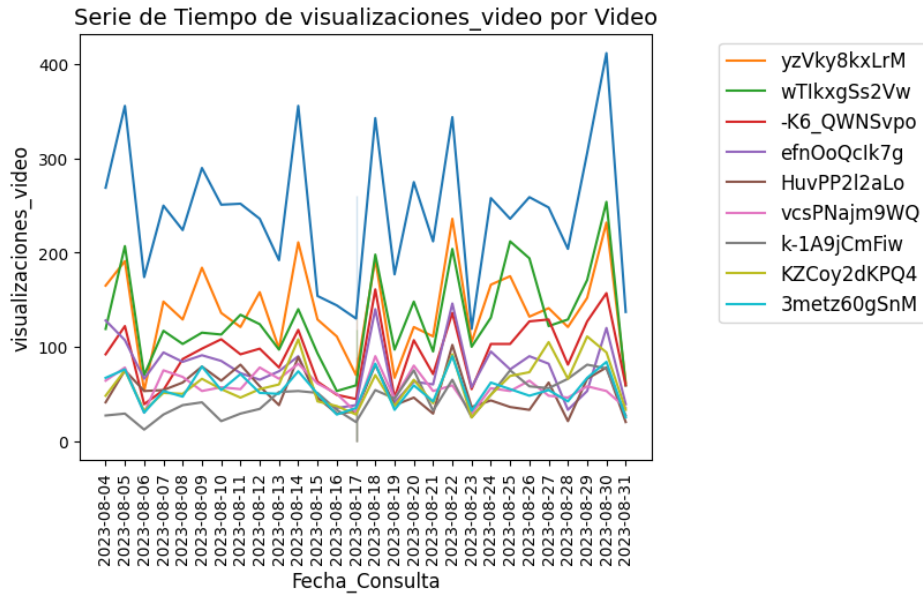


Figura 3: Delta Visualizaciones por video

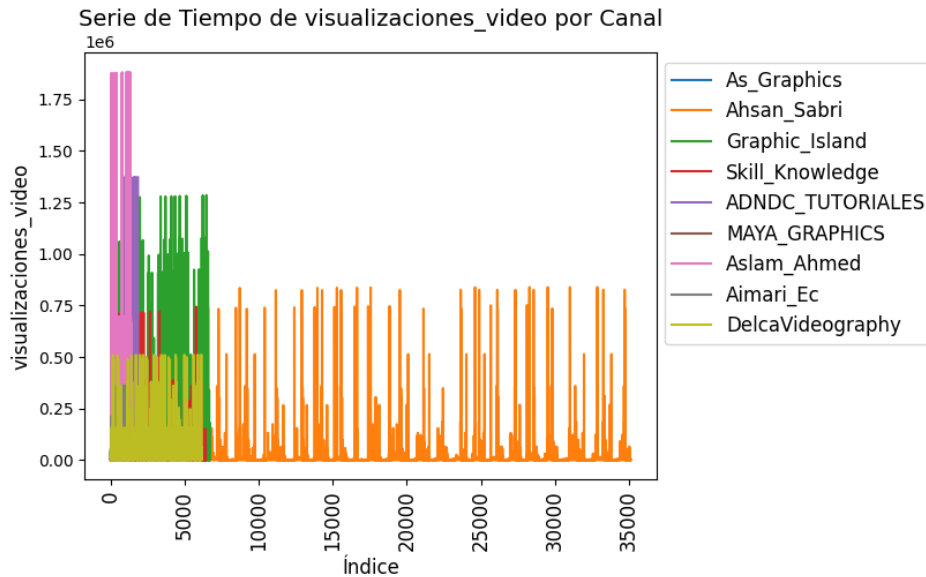
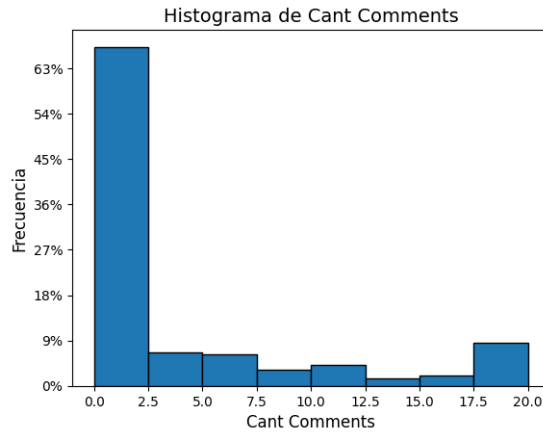
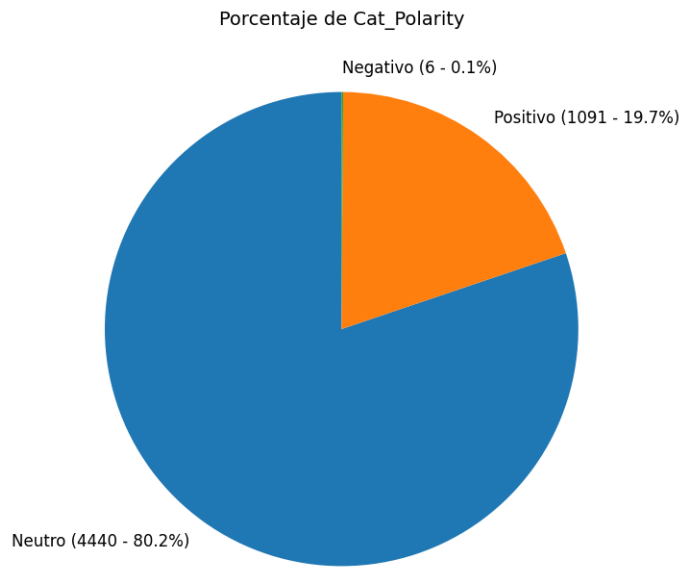


Figura 4: Visualizaciones videos y cantidad de contenidos

son Neutros y que el porcentaje de comentarios negativos es muy bajo (0.1%). También se observa que la mayor cantidad de comentarios por contenido es 20 y se identifica que cerca del 9% de los contenidos presentan esta característica, lo que podría dar un indicio de contenidos que serán más recomendados por YouTube, ya que según las validaciones con el dueño del canal DelcaVideography, en la mayoría de los casos las personas sólo ven los contenidos y pocos los comentan. Esto generalmente sucede cuando hay un real interés por el contenido, como se observa en la Figura 5.



(a) Histograma Comentarios



(b) Polaridad Comentarios

Figura 5: Información Comentarios

### 3.2. Modelos Aplicados

De acuerdo con lo descrito anteriormente, se usa el conjunto de datos de 5,537 registros compuesto por 9 usuarios (`user_id`) y 4,055 elementos (`item_id`) y se aplica cada uno de los modelos (descritos en la sección 2.3) creando una malla de hiperparámetros, en donde se encuentra que los óptimos en cada caso son:

KNN Baseline:

- Número de vecinos (`k`): 10
- Métrica de similitud: Mean Squared Difference (MSD por sus siglas en inglés)

- Filtro basado en usuarios (User\_based) = False, es decir se aplica el filtro basado en elementos

Co-Clustering:

- Número de clústeres para usuarios (n\_cltr\_u): 2
- Número de clústeres para ítems o elementos (n\_cltr\_i): 2

SVD:

- Número de factores latentes (n\_factors): 50
- n\_epochs: 20. Número de épocas de entrenamiento durante el procedimiento de descenso de gradiente estocástico (SGD) para aprender los factores latentes del modelo de recomendación.
- Tasa de aprendizaje (lr\_all): 0.005
- Término de regularización (reg\_all): 0.1

SVD++:

- Número de factores latentes (n\_factors): 20
- n\_epochs: 30.
- Tasa de aprendizaje (lr\_all): 0.01
- Término de regularización (reg\_all): 0.1

Para la evaluación de los modelos se usaron dos métricas para medir su precisión: Error Absoluto Medio (MAE por sus siglas en inglés) y Raíz del Error Cuadrático Medio (RMSE por sus siglas en inglés).

MAE encuentra la media de los errores absolutos de la diferencia entre el puntaje real y el puntaje predicho, con la siguiente ecuación (Hug, 2015a):

$$MAE = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

RMSE calcula la raíz cuadrada de los errores medios de la diferencia entre el puntaje real y el puntaje predicho, de la siguiente forma (Hug, 2015a):

$$RMSE = \sqrt{\frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} (r_{ui} - \hat{r}_{ui})^2}$$

Donde,  $\hat{R}$  es el número total de ratings,  $r_{ui}$  es el puntaje real asignado por el usuario  $u$  al elemento  $i$ , y  $\hat{r}_{ui}$  es el puntaje predicho para el elemento  $i$  (Hug, 2015a).

Algoritmo	RMSE	MAE	% RMSE	% MAE
KNNBaseline	0.018536481	0.006246393	4.65 %	1.57 %
CoClustering	0.018118556	0.006100008	4.54 %	1.53 %
SVD	0.02194188	0.009717125	5.50 %	2.44 %
SVD++	0.016717564	0.005810385	4.19 %	1.46 %

Cuadro 1: Resultados de los algoritmos

### 3.3. Resultados Modelos

Para los modelos mencionados en la Sección 3.2, se aplican las dos métricas descritas anteriormente MAE y RMSE. Los resultados de ambas métricas están expresados en las unidades de medida del valor que se está analizando, en este caso el rating de visualizaciones de los videos. Con lo cual, para dimensionar sus resultados de forma porcentual, se calcula la relación del valor obtenido respecto al valor de rating máximo en el conjunto de datos, que en este caso es de 0.398993365. Encontrando los resultados que se muestran en el Cuadro 1.

Teniendo en cuenta que el objetivo de ambas métricas es buscar minimizar el valor obtenido y que el RMSE podría ser más apropiado para evaluar modelos cuando la distribución del error sea Gaussiano o que puede ser una métrica más sensible a valores atípicos (Chai and Draxler, 2014), se da mayor importancia al resultado del MAE. Razones por las que se decide seleccionar el algoritmo SVD++ como el de mejor desempeño para el caso de estudio.

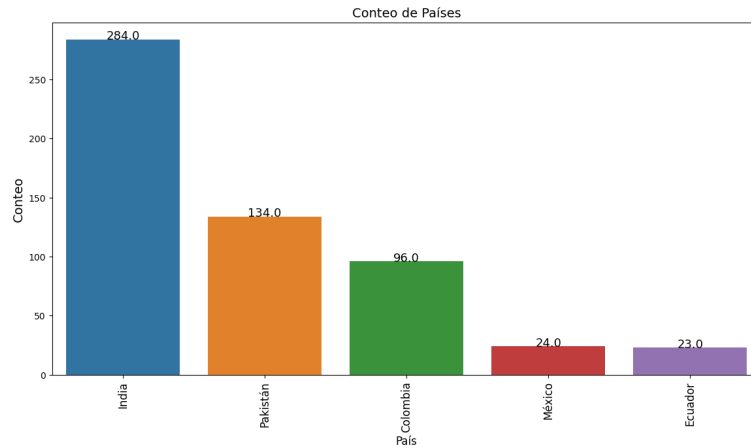
Estos resultados están alineados con lo descrito por Koren (2008) y su propuesta de los modelos SVD++, en donde cuenta que con la experiencia en la participación en *The Netflix Prize*, aprendieron que los modelos de vecindad y de factores latentes se aplican de forma muy diferente según las estructuras de datos, y que por lo tanto ninguno de ellos es óptimo por si solo. Esto debido a que los modelos de vecindad son eficaces detectando relaciones muy localizadas. Mientras que los modelos de factores latentes son eficaces estimando la estructura general de los elementos, pero son pobres en detectar relaciones fuertes en conjuntos pequeños de datos.

Como se describió anteriormente, el mejor algoritmo seleccionado es el SVD++, el cual es una versión mejorada del modelo SVD que realiza una factorización matricial teniendo en cuenta calificaciones y relaciones implícitas entre los usuarios y elementos tenidos en cuenta. Como resultado de este modelo, se tienen varios tipos de recomendaciones: basadas en usuarios y basadas en elementos. Las recomendaciones basadas en usuarios se hacen teniendo en cuenta las preferencias que pueda tener un usuario de acuerdo a interacciones similares en el pasado. Mientras que las recomendaciones basadas en elementos se hacen basadas en elementos similares con los que se haya interactuado anteriormente. Se validan ambos tipos de recomendaciones y se decide trabajar con las recomendaciones basadas en elementos, ya que proporcionan más información para el interés de identificar que contenidos son exitosos en el mercado. Además, se encuentra relación directa con lo descrito en el documento de Sarwar et al. (2001), quienes realizaron varios experimentos aplicando modelos de filtrado colaborativo y llegaron a la conclusión de que los algoritmos basados

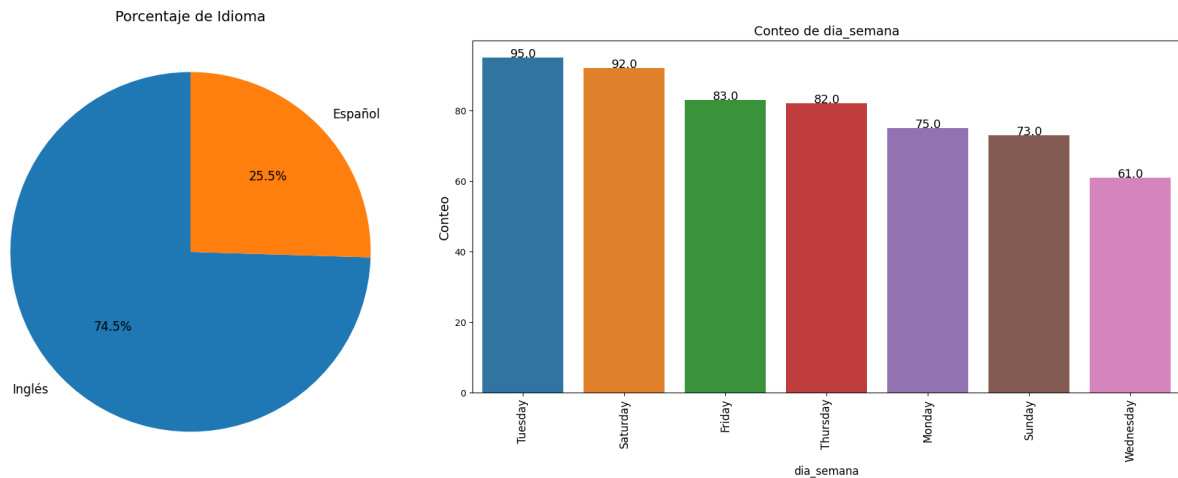
en elementos presentan mejor rendimiento que los algoritmos basados en usuarios.

Enfocando el análisis a las recomendaciones basadas en elementos, también se pueden tener diferentes tipos de respuestas: elementos valorados por el usuario, elementos no valorados por el usuario y recomendaciones generales que incluyen las 2 anteriores. En este caso se seleccionan las recomendaciones generales puesto que, según validaciones con el dueño del canal DelcaVideography, se identifica que además de contenidos similares a los que tiene considerados en su canal, se pueden identificar otros tipos de contenidos exitosos no necesariamente valorados por el usuario principal, es decir se amplía el panorama de posibles contenidos a publicar, lo cual agrega un valor adicional para los objetivos esperados.

Este resultado cuenta con 561 recomendaciones, en donde como se observa en la Figura 6, la mayoría de los contenidos son publicados por canales de la India, en idioma inglés y principalmente se publican los días Martes y Sábado.



(a) Países Canales en Resultado



(b) Idiomas Canales en Resultado

(c) Día publicación en Resultado

Figura 6: Información Canales en Resultado

Revisando el detalle de los canales, la mayoría de los contenidos recomendados han sido publicados por el canal Graphic Island (Figura 7). Llama la atención que en cuarto lugar se encuentran contenidos del canal DelcaVideography, por lo cual se analizan los deltas de visualizaciones de los 4 principales canales y se identifica, que si bien los contenidos de DelcaVideography son de interés para la audiencia, no tienen tanto impacto porque los deltas de visualizaciones son los más bajos respecto a los otros 3 canales (para mayor detalle ver Sección C del Material Suplementario).

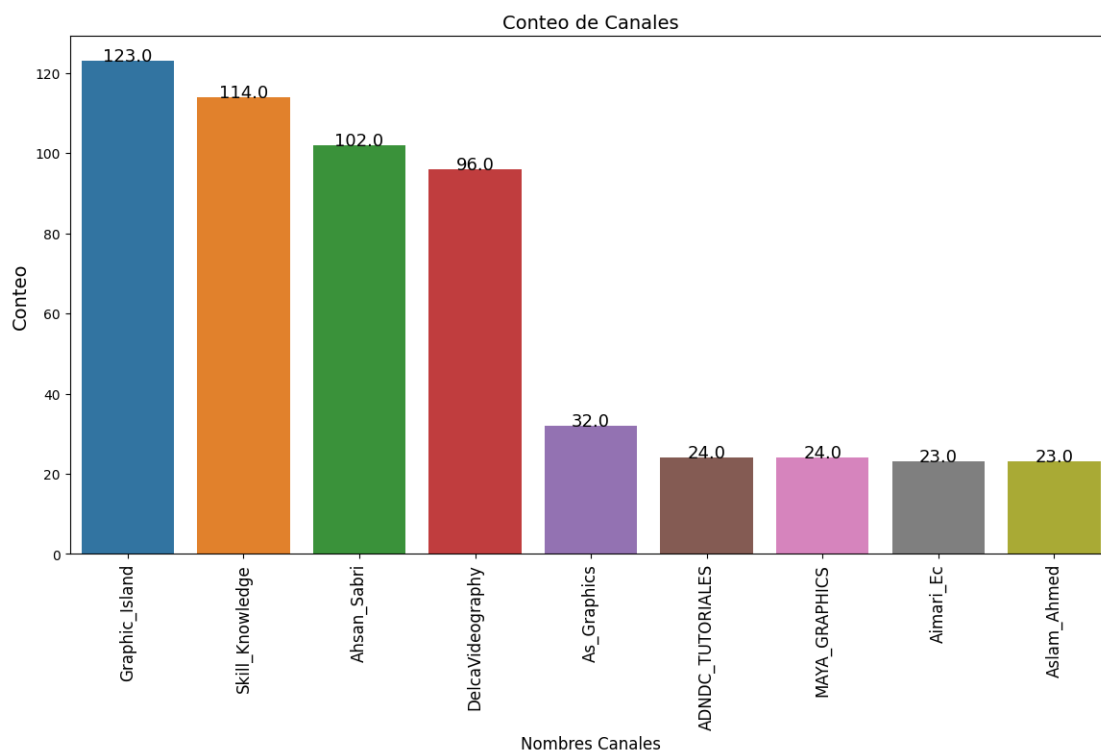


Figura 7: Composición Canales en Resultado

Adicionalmente se identifica que sobre la categorización de temas, el de CorelDRAW es el que mayor participación tiene en las recomendaciones (Figura 8). Según las validaciones hechas, esto se corresponde con la información en los dashboards analíticos que entrega YouTube Studio al canal DelcaVideography.

Con relación a la información de comentarios, se identifica que se redujo la cantidad de contenidos con menos de 2 comentarios, pasando de un 65% (Figura 5) a un 31% en las recomendaciones. Adicionalmente, se encuentra que cerca del 60% de los contenidos presenta más de 5 comentarios, y un 22% presenta más de 17. Lo cual, según lo descrito anteriormente sobre el comportamiento de los usuarios ante el hecho de registrar comentarios, indica que los contenidos recomendados si están generando un mayor interés para la comunidad al interactuar con el contenido, con lo cual se da retroalimentación al dueño del canal y por ende al algoritmo de YouTube al identificar una satisfacción con el contenido (Insider, 2023) (Figura 9).

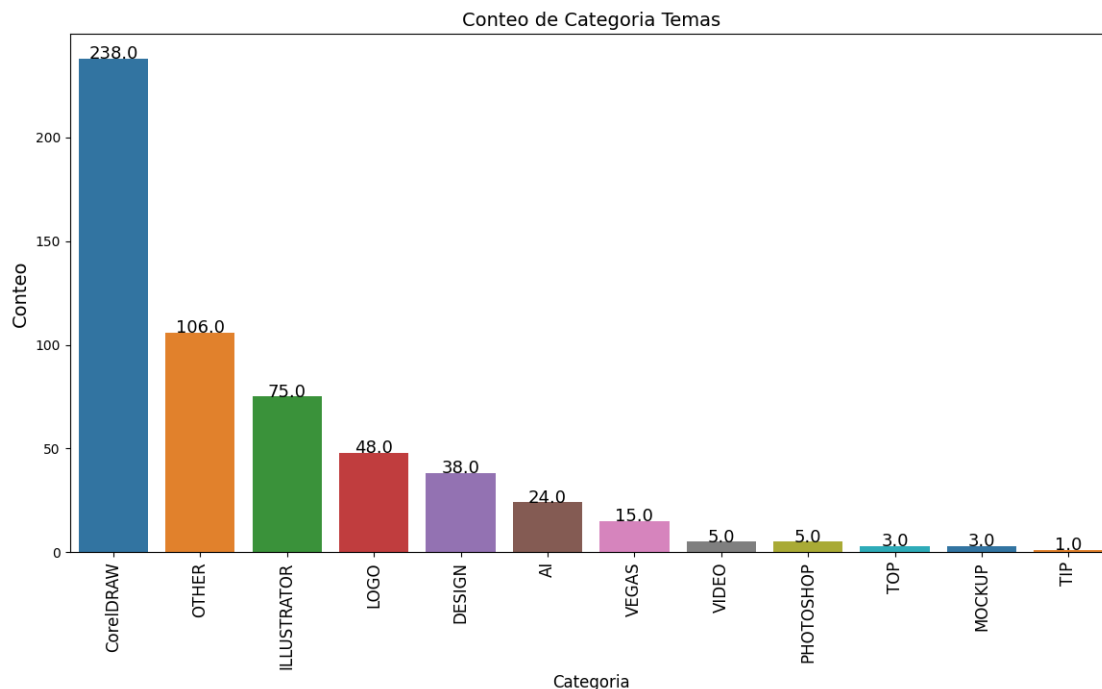


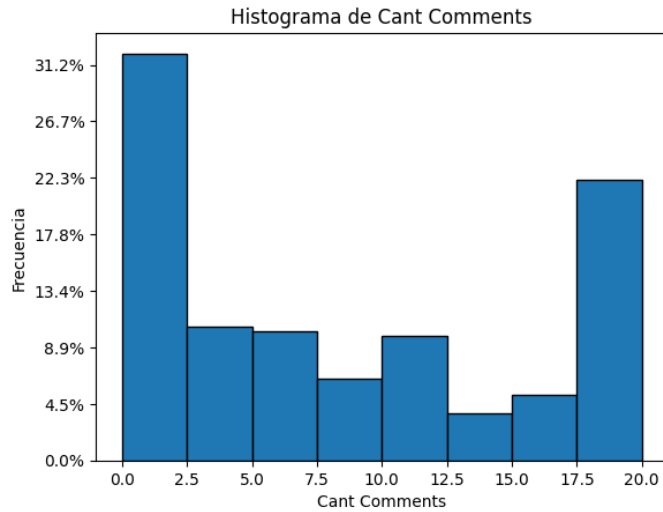
Figura 8: Categorías de Temas en Resultado

Adicionalmente como se puede ver en la Figura 9, se observa que la mayoría de los comentarios son Positivos con un 35.8%, lo cual es superior al universo de contenidos descrito al inicio del documento, en la Figura 5, con un 19.7%.

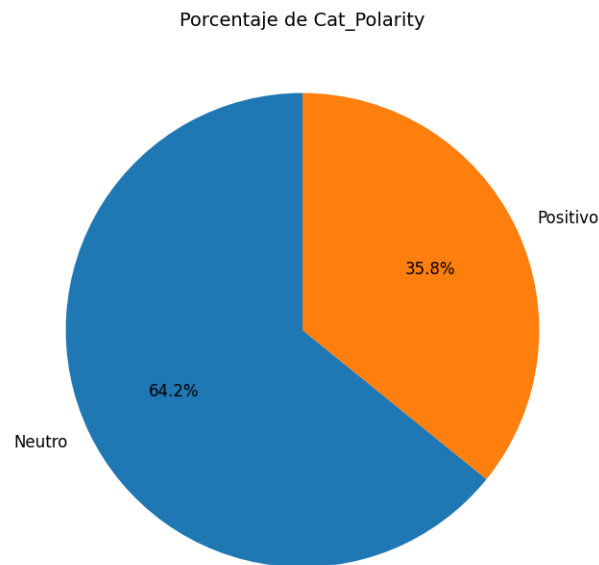
## 4. Discusión

Antes de comenzar con el análisis de resultados, es importante tener en cuenta los términos que se usan para clasificar el ciclo de vida del comportamiento de un video en el tiempo, que como los llama Fons (2021), un Youtuber especialista en analítica y SEO (*Search Engine Optimization* por sus siglas en inglés) de esta plataforma, son: 1) *Sprint* se refiere a las primeras 72 horas de un video, en donde YouTube promociona el contenido a una gran variedad de audiencia (nueva y de cada canal) para identificar su respuesta ante ese contenido. 2) *Evergreen* es el rendimiento que tiene un contenido a lo largo de su vida, que puede ser bueno o malo. El cual puede madurar temprano o tarde a través del tiempo. Por ejemplo si un video tiene buen desempeño en la fase de *Sprint*, será más recomendado por YouTube y tendrá un mejor desempeño de forma temprana. Sin embargo, esto no necesariamente ocurre en esta primera fase, porque hay videos que su rendimiento se incrementa tiempo después en la fase *Evergreen*.

Retomando el resultado descrito en la Sección 3.3, se tiene una lista de videos recomendados, los cuales se ordenan descendientemente por los contenidos con mayores deltas de visualizaciones, es decir mayor velocidad de visualización, ya que son los contenidos de mayor interés para el usuario DelcaVideography. Adicionalmente, parte de estos resultados



(a) Histograma de Comentarios en Resultado



(b) Polaridad de Comentarios en Resultado

Figura 9: Información Comentarios en Resultado

asociados a recomendaciones del usuario DelcaVideography, se contrastan con las recomendaciones privadas hechas por YouTube a este canal, y se identifica una alta coincidencia, con lo cual podría esperarse que las recomendaciones obtenidas de los demás usuarios (canales) hagan parte de estas recomendaciones privadas que hace YouTube en cada caso.

Para el creador del canal DelcaVideography, si bien la mayoría de contenidos recomendados están relacionados con Corel Draw (siendo este el contenido principal de este canal), se logra identificar un nuevo tema (con altos deltas de visualización) que puede considerarse

de interés para la creación de logos, en este caso en Word. Este contenido es analizado por el dueño del canal. Posteriormente se publica uno similar basado en estas recomendaciones. Una vez publicado el contenido (con id “IExd1LLGOfo”) se monitorean sus estadísticas públicas durante 1 semana (para analizar su comportamiento en la fase de *Sprint*). En la Figura 10 se muestran los resultados en visualizaciones en comparación con otros videos con la misma ventana de tiempo de publicación para este canal.

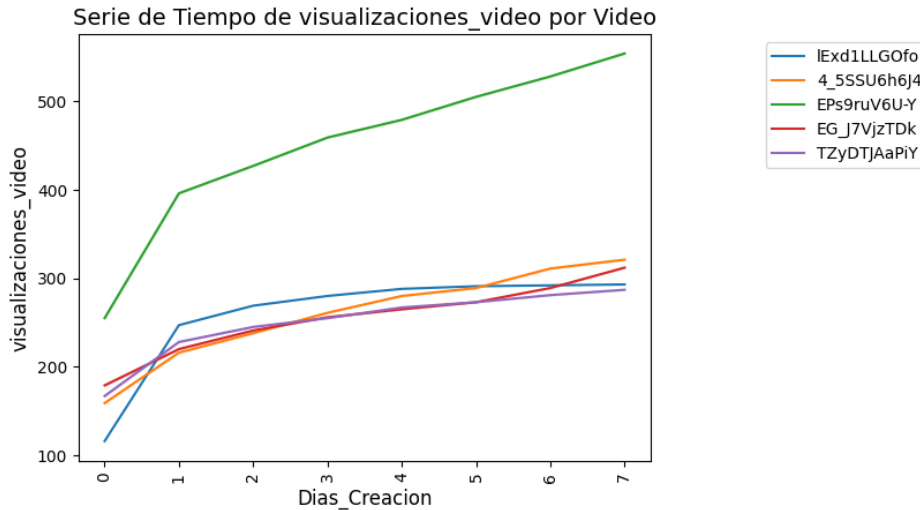


Figura 10: Comportamiento video de prueba publicado

Al analizar este resultado, en la Figura 10 se observa que en los primeros días de su publicación, el vídeo en cuestión obtuvo el segundo lugar en métricas de rendimiento en comparación con otros videos. Sin embargo, esta tendencia no se mantiene constante a lo largo del tiempo, y a partir del cuarto día de publicación, se puede notar una meseta en la cantidad de visualizaciones. En consecuencia, se puede concluir que las recomendaciones proporcionadas por el modelo, inicialmente permitieron la identificación de un nuevo tema de interés para la audiencia del canal DelcaVideography. No obstante, parece que dicho contenido no ha logrado mantener un nivel atractivo sostenido, lo que sugiere la necesidad de seguir explorando otras recomendaciones y validar la viabilidad de contenidos que puedan generar un impacto mayor.

Para el dueño del canal, lo ocurrido con el video de Word que se publicó, se comportó de tal manera debido a que la base de su audiencia no está en su canal por temas diferentes a los tradicionales. Esto significa que si se quiere implementar un diseño, logo o cualquier otro video relacionado con lo anterior, pero en un programa diferente por los cuales la audiencia está siguiéndolo, no va a tener la misma respuesta de ella y abandonará rápidamente el interés por estos videos. No obstante, este video podrá posicionarse en el tiempo hacia una audiencia externa que no esté suscrita (o no conozca previamente a DelcaVideography).

En resumen, es importante señalar que dadas las múltiples variables que influyen en el éxito de un contenido, se recomienda complementar estos análisis con datos internos específicos del canal DelcaVideography, con el fin de comprender mejor las características y el perfil

de su audiencia y así afinar más la creación de los contenidos recomendados a publicar.

## 5. Conclusión

Aunque los modelos de recomendación se han usado normalmente para recomendaciones de productos a comprar o consumir (por ejemplo en *E-commerce* web) y pese a que la información insumo para este trabajo corresponde sólo a las estadísticas generales publicadas en YouTube, con el desarrollo de esta investigación se demuestra que es posible aplicar las técnicas de modelos de recomendación en contextos diferentes, esto gracias al ajuste de los datos para poder aplicarlas. En este caso se aplica a la creación de contenidos de interés para la publicación de videos de Diseño gráfico, audiovisual y fotográfico. Sin embargo, si se hubiera tenido acceso a la información privada de los canales, se hubieran tenido muchas más herramientas para mejorar los resultados de las recomendaciones.

En el día a día consumimos datos, donde plataformas como redes sociales (Instagram, YouTube, Netflix, entre otras) usan modelos de recomendación para personalizar la experiencia de los consumidores, sin embargo con este estudio se concluye que es posible usar estas técnicas en el sentido opuesto (ingeniería inversa), es decir no con el fin de consumir sino de crear contenidos. Ya que con el apoyo de la ciencia de datos, en este caso con los modelos de recomendación, se pueden identificar los intereses de los usuarios y por lo tanto ayudar a optimizar la creación de contenidos.

Si bien algunas de las recomendaciones concuerdan con los reportes que entrega YouTube a cada uno de los canales, en este caso al canal DelcaVideography, los resultados de este modelo entregan un mayor detalle o características más específicas de los contenidos que debería tener en cuenta para seguir publicando, como lo son identificar nuevos temas que podrían ser de interés para la audiencia general. Por ejemplo: Logos en Word. Esto conecta un tema de interés para la audiencia como lo son los Logos con una herramienta nueva, en este caso Word.

Aunque con los resultados obtenidos existe una gran variedad de contenidos recomendados para que el canal continúe aplicando, se evidencian algunos otros puntos a desarrollar que podrían mejorar los resultados e interpretación:

- Realizar pruebas con otras recomendaciones para validar el desempeño del modelo y tener conclusiones más precisas.
- Mejorar la técnica de análisis de texto para crear recomendaciones más precisas de los temas que se deberían publicar.
- Construir una herramienta que realice análisis de video para entregar información adicional a tener en cuenta en la producción de los contenidos recomendados.
- Construir una metodología para complementar el análisis de las recomendaciones con las estadísticas internas del canal al que se le estén aplicando, en este caso DelcaVideography.

- Trabajar en la mejora del *Look and feel* (apariencia) del resultado, ya que actualmente se entrega una tabla de Excel en bruto, lo que dificulta su consumo y entendimiento para el creador de contenido que quiera consultarla. Por lo tanto, se propone entregar los datos visualmente más atractivos y fáciles de consumir para mejorar la experiencia de usuario en la toma de decisiones.

De forma similar a como se hizo en este trabajo, se podrán analizar diferentes usos de los sistemas de recomendación en áreas de investigación en campos diferentes a los tradicionalmente aplicados. Esto teniendo en cuenta que será necesario adaptar los datos según las necesidades.

## Referencias

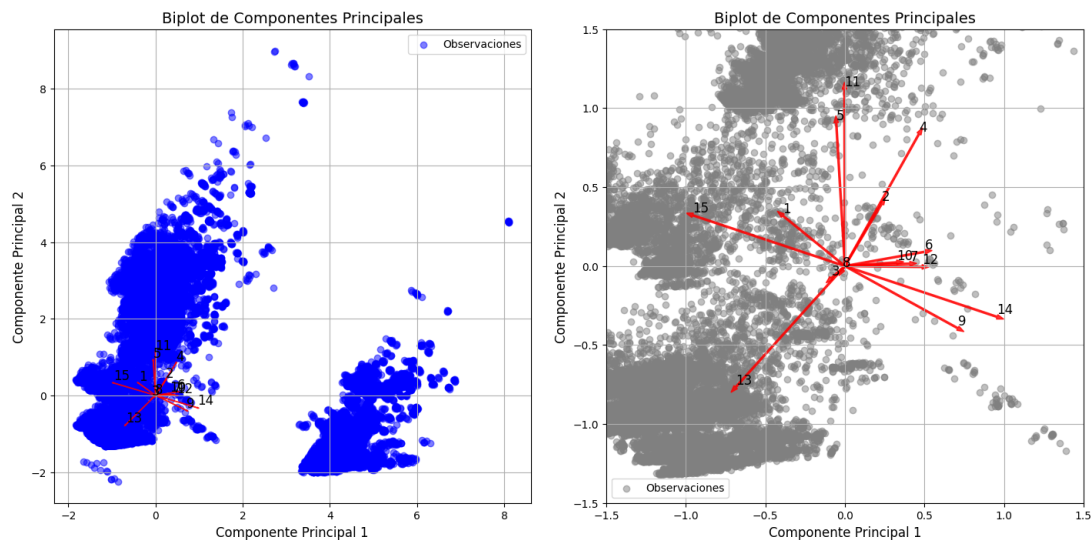
- Abel, F., Benczúr, A., Kohlsdorf, D., Larson, M., and Pálovics, R. (2016), “RecSys challenge 2016: Job recommendations,” Association for Computing Machinery, Inc.
- Bennett, J. and Lanning, S. (2007), “The Netflix Prize,” *In Proc. of KDD Cup and Workshop*, 3–6.
- Bew, D., Harvey, C. R., Ledford, A., Radnor, S., and Sinclair, A. (2019), “Modeling Analysts’ Recommendations Via Bayesian Machine Learning,” *The Journal of Financial Data Science*, 75–98.
- Breese, J. S., Heckerman, D., and Kadie, C. (1998), “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” *Microsoft Research*, 43–52.
- Chai, T. and Draxler, R. R. (2014), “Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature,” *Geoscientific Model Development*, 7, 1247–1250.
- Fons, R. (2021), “CRECETUBE — Curso Especializado para Acelerar Tu Crecimiento en YouTube® 5. Fase Evergreen,” URL <https://vimeo.com/496539735>.
- Guruge, D. B., Kadel, R., and Halder, S. J. (2021), “The state of the art in methodologies of course recommender systems—a review of recent research,” *Data*, 6, 1–30.
- Hug, N. (2015a), “Surprise Documentation. Accuracy module,” URL <https://surprise.readthedocs.io/en/stable/accuracy.html#module-surprise.accuracy>.
- (2015b), “Surprise Documentation. Co-clustering,” URL [https://surprise.readthedocs.io/en/stable/co\\_clustering.html#surprise.prediction\\_algorithms.co\\_clustering.CoClustering](https://surprise.readthedocs.io/en/stable/co_clustering.html#surprise.prediction_algorithms.co_clustering.CoClustering).
- Insider (2023), “YouTube Algorithm 2023: What Creators Need to Know,” URL <https://www.youtube.com/watch?v=67TkpHKv9tE>.
- Jauset, J. A. (2007), *Estadística para periodistas, publicitarios y comunicadores*, Editorial UOC.
- Jena, K. K., Bhoi, S. K., Malik, T. K., Sahoo, K. S., Jhanjhi, N. Z., Bhatia, S., and Amsaad, F. (2023), “E-Learning Course Recommender System Using Collaborative Filtering Models,” *Electronics (Switzerland)*, 12.
- Kemp, S. (2022), “Reels Grew By 220M Users in Last 3 Months (And Other Jaw-Dropping Stats),” URL [https://blog.hootsuite.com/simon-kemp-social-media/#Top\\_10\\_takeaways](https://blog.hootsuite.com/simon-kemp-social-media/#Top_10_takeaways).
- Koren, Y. (2008), “Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model,” *ACM*, 426–434.

- (2010), “Factor in the neighbors: Scalable and accurate collaborative filtering,” *ACM Transactions on Knowledge Discovery from Data*, 4.
- Koren, Y., Bell, R., and Volinsky, C. (2009), “MATRIX FACTORIZATION TECHNIQUES FOR RECOMMENDER SYSTEMS,” *IEEE Computer Society*, 42–49.
- Nguyen, L. V., Vo, Q. T., and Nguyen, T. H. (2023), “Adaptive KNN-Based Extended Collaborative Filtering Recommendation Services,” *Big Data and Cognitive Computing*, 7.
- Oliva, M. (2022), “Me encanta mi trabajo, pero es un trabajo’: creadores de contenido en redes sociales e imaginarios laborales,” *Palabra Clave*, 25.
- Pattier, D. (2022), “Diseño y validación de instrumento para analizar canales educativos de YouTube,” *Icono14*, 20.
- Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994), “GroupLens: An Open Architecture for Collaborative Filtering of Netnews,” .
- Ricci, F., Rokach, L., and Shapira, B. (2011a), *Introduction to Recommender Systems Handbook*, Springer US.
- Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B. (2011b), *Recommender Systems Handbook*, Springer US.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001), “Item-Based Collaborative Filtering Recommendation Algorithms,” *ACM*, 285–295.

# Material Suplementario

## A. Análisis PCA

De acuerdo a los datos obtenidos en la preparación de datos, y comparando los componentes 1 y 2, el resultado de PCA se muestra en la Figura 11. En donde cada número de variable está definido en el Cuadro 2.



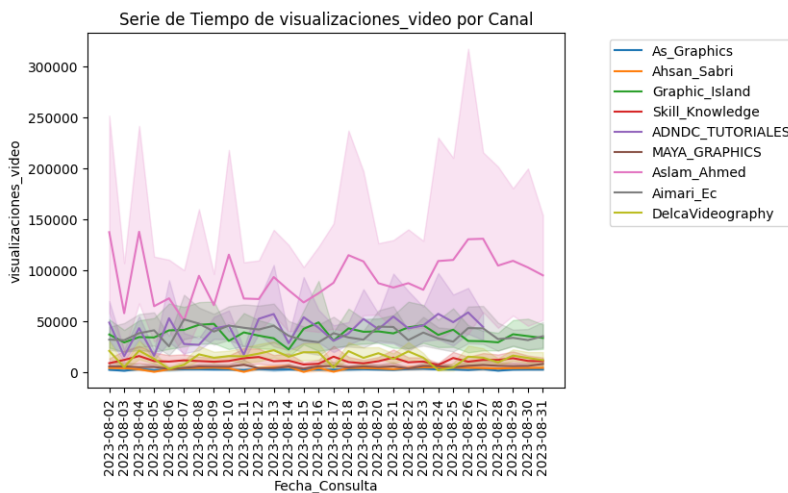
(a) Biplot Componentes Principales General (b) Biplot Componentes Principales Enfoque

Figura 11: Biplot Componentes Principales

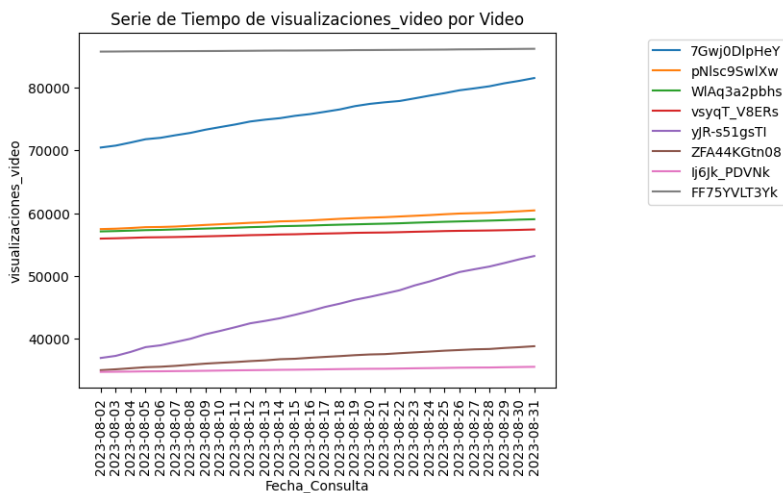
#	Variable
1	IngresosEst_Prom
2	likes
3	rating
4	Comment
5	Result_Polarity
6	meses_creacion
7	duracion_video_min
8	dia_semana
9	País_Colombia
10	País_Ecuador
11	País_India
12	País_México
13	País_Pakistán
14	Idioma_Español
15	Idioma_Inglés

Cuadro 2: Lista de Variables

## B. Comportamiento de Visualizaciones en información original



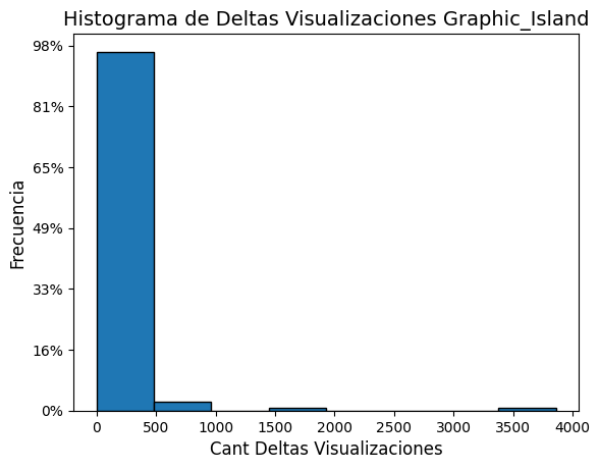
(a) Visualizaciones video por Canal



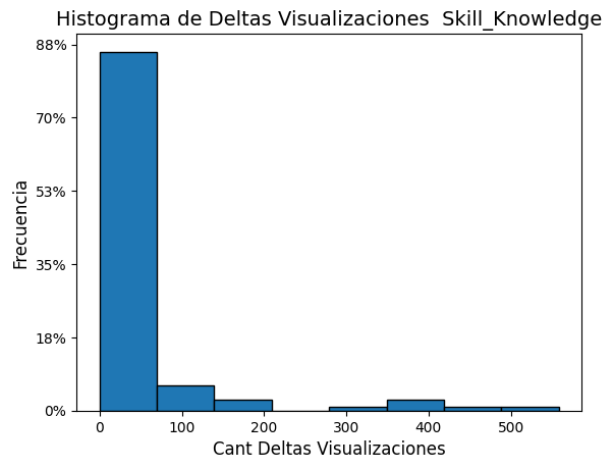
(b) Visualizaciones video

Figura 12: Visualizaciones en información original

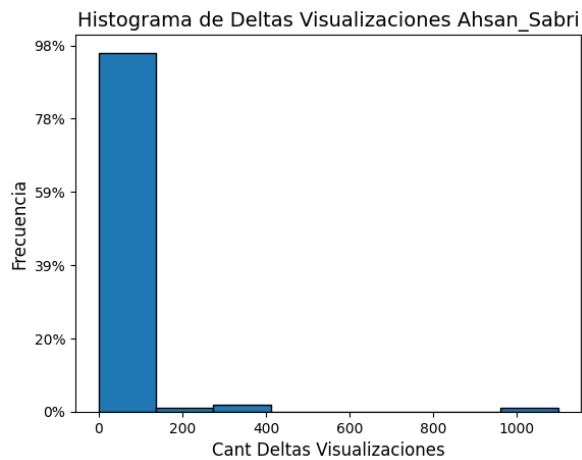
## C. Análisis Resultados Modelos



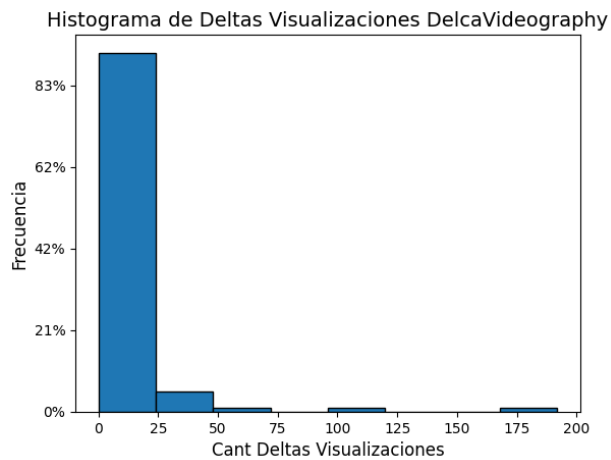
(a) Deltas Visualizaciones Graphic\_Island



(b) Deltas Visualizaciones Skill\_Knowledge



(c) Deltas Visualizaciones Ahsan\_Sabri



(d) Deltas Visualizaciones DelcaVideography

Figura 13: Deltas Visualizaciones en Resultado