# UNIVERSIDAD TECNOLÓGICA DE CAMPECHE

## INGENIERÍA EN DESARROLLO Y GESTIÓN DE SOFTWARE

**ACTIVIDAD:**

PRÁCTICA

## ASIGNATURA:

EXTRACCION DE CONOCIMIENTOS DE BASE DE DATOS

**NOMBRE DE LOS ALUMNOS:** LINDA ARLENI JUÁREZ SÁNCHEZ ILHUÍCATL BÁEZ BENÍTEZ

FRANCISCO DANIEL SÁNCHEZ BETANZOS CRISTIAN GONZÁLES JIMÉNEZ

## NOMBRE DEL DOCENTE:

MARÍA LOURDES CÁRDENAS MALDONADO

## GRADO Y GRUPO:

9no “B”

## PERIODO:

MAYO – AGOSTO 2023

# TWITTER

**Justificación del algoritmo utilizado.**

El algoritmo de Twitter sigue siendo un componente clave en la plataforma para determinar qué contenido se muestra en el feed de cada usuario. El objetivo principal del algoritmo es ofrecer a los usuarios un feed personalizado y relevante. Esto significa que cada usuario puede tener una experiencia única en Twitter, ya que el algoritmo adapta los tweets mostrados según los intereses y comportamientos individuales. Este algoritmo utiliza una gran variedad de factores y señales para calcular la relevancia y el orden de los tweets que aparecen en el feed.

Uno de los factores que tiene en cuenta, se trataría de la interacción del usuario, como los likes, retweets y respuestas, **(para evaluar la popularidad y la calidad de un tweet)**. Además, considera la relación entre el usuario y las cuentas que sigue, así como la actividad reciente de esas cuentas.

A través del análisis de patrones de comportamiento, interacciones anteriores y contenido consumido, es cómo el algoritmo personaliza el feed para mostrar tweets que se ajusten a los intereses individuales de cada usuario, como hemos mencionado anteriormente.

Además de la relevancia, el algoritmo también busca ofrecer variedad en el contenido mostrado en el feed. Se esfuerza por evitar la saturación de tweets de una misma cuenta o tema, buscando equilibrar la diversidad de contenido para brindar una experiencia más enriquecedora.

# Descripción de los resultados del algoritmo

Es importante destacar que el algoritmo de Twitter se actualiza y mejora de forma regular para adaptarse a las necesidades cambiantes de los usuarios y combatir prácticas no deseadas, como el spam y el contenido inapropiado.

En 2023, el algoritmo de [Twitter](https://twitter.com/coostoes?lang=es) se ha convertido en una herramienta indispensable en el mundo de las redes sociales. El algoritmo de Twitter despliega su función principal al filtrar y mostrar los tweets más relevantes y valiosos en el feed de cada usuario. A medida que el contenido en Twitter sigue creciendo exponencialmente, este algoritmo se vuelve esencial para destacar entre el ruido y llegar a la audiencia adecuada.

## Cómo se eligen los tuits

Los tuits que llegan a la página de inicio de «Para ti», pasan por un proceso de selección y filtración de tres etapas:

* abastecimiento de candidatos.
* Clasificación de los tuits con modelos de aprendizaje automático.
* aplicación de métodos heurísticos.

En un proceso que se ejecuta unas 5.000 millones de veces al día, todo en un segundo y medio.

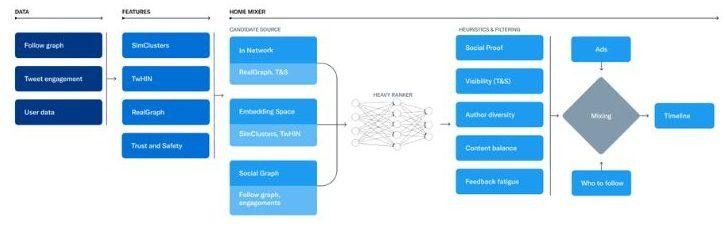


Diagrama que muestra los principales componentes con los que se construye una línea de tiempo

## Abastecimiento de candidatos

Este es un proceso con el que la plataforma recibe los mejores tuits provenientes de 3 fuentes de recomendación que forman una base de datos gigante: FollowGraph (que estima quien sigue a quien), las métricas de interacción con los tweets, y otros datos de usuarios relevantes, como usuarios que has bloqueado o contenidos que te interesan.

Sobre esta base de datos, se aplican dos cálculos para obtener los tweets candidatos para cada usuario: RealGraph, que calcula la probabilidad de que te interese un tweet publicado por un usuario, aún si nunca antes has interactuado con él; y TweepCred, que determina tu autoridad en la plataforma, según la autoridad que tienen tus seguidores.

Gracias a estos cálculos, se extraen los mejores 1.500 tweets provenientes de un grupo de cientos de millones de tuits de las fuentes de recomendación, tomando en consideración tanto las cuentas que sigues como las que no sigues. «Hoy en día, la cronología de Para ti consiste en un 50% de tuits dentro de la red y un 50% de tuits fuera de la red en promedio, aunque esto puede variar de un usuario a otro», explica la red social.

Los tuits candidatos fuera de tu red se eligen mediante: Social Graph o gráficos sociales, que busca usuarios afines a ti, estudiando los tuits con los que han interactuado los usuarios que sigues y los usuarios que tienen gustos parecidos a los tuyos; y Embedding Spaces o espacios de incrustaciones, que también busca usuarios afines, pero por medio de números de cadenas de texto que determinan cuánto se parecen dos tweets o conjuntos de tweets.

Dentro de estos espacios de incrustación, existe SimCluster, con el que se pueden descubrir comunidades ancladas por un grupo de [influencers](https://marketing4ecommerce.net/el-influencer-y-su-importancia-en-el-marketing-web/). Existen alrededor de

145.000 comunidades, actualizadas cada tres semanas, que pueden variar en tamaños, desde unos pocos miles de usuarios hasta cientos de millones. Mientras más popularidad tenga un tuit dentro de una comunidad, más se asociará ese tipo de contenido a la comunidad.

Clasificación de cada tuit

En este punto del proceso, el sistema ya ha seleccionado aproximadamente 1.500 tuits candidatos que pueden ser relevantes. Estos son sometidos a modelos de

aprendizaje automático, que predicen la importancia que tiene cada una de las publicaciones candidatas.

La encargada de esta clasificación es una red neuronal de 48 millones de parámetros, que se entrena constantemente en datos de interacciones de tweets (Me gusta, Retweets y Respuestas) para lograr obtener el mejor resultado de interacción posible. En este momento, el sistema genera diez etiquetas que dan una puntuación a cada tweet y se establece la probabilidad de que interactúes con estas publicaciones.

## Heurísticas y filtros

Estando ya clasificada toda la información, se aplican filtros que permiten una recomendación más precisa, como eliminar las sugerencias de cuentas bloqueadas, evitar muchos tweets consecutivos del mismo usuario, reducir la puntuación de una publicación si has dejado comentarios negativos en ella, entre otras acciones.

## Mezclar y servir

Con las publicaciones de sugerencias escogidas, Twitter activa el Home Mixer, «la columna vertebral del software que conecta diferentes fuentes candidatas, funciones de puntuación, heurísticas y filtros», según explican desde la misma empresa. Este servicio es el encargado de enviar las sugerencias a cada dispositivo, mientras que el sistema de clasificación las combina con otro tipo de contenido, como anuncios y recomendaciones de seguimiento de otras cuentas.

# Reporte De Evaluación Y Optimización De La Implementación Del Algoritmo.

Aunque los medios de comunicación tratan poco a poco de no depender tanto de terceros y establecer una relación más directa con sus potenciales lectores, las redes sociales siguen siendo una herramienta necesaria no sólo para llegar a nuevas audiencias, sino para transmitir la imagen de marca y el valor y la calidad del periodismo que realizan.

Y, aunque en estos momentos Twitter vive momentos de cambio, y nadie sabe cuál será el futuro de la red, mientras éste se deﬁne de manos de su nuevo propietario es importante conocer cómo funciona el algoritmo de la red de Elon Musk para tratar de maximizar el esfuerzo, en una época en la que, además, hay que hacer más con menos.

“los tuits que probablemente te importen más aparecerán primero en tu línea de tiempo”.

En términos simples, el algoritmo de Twitter es el sistema que determina lo que Twitter piensa que “probablemente te importe más” y, por lo tanto, lo primero que veas cuando inicies sesión en tu red social.

Frescura

Twitter prioriza las Tendencias en función de temas que son populares en ese preciso instante, y otorga menos relevancia a otras tendencias más sostenidas en el tiempo, e incluso diarias, pero no tan apegadas al momento.

Relevancia

Twitter crea un perﬁl de cada usuario en función de acciones anteriores, como los propios tuits y los tuits con los que se ha interactuado.

La aplicación también priorizará las cuentas con las que se interactúa a menudo y los temas que sigue y con los que interactúa más.

La ubicación también es un elemento clave en el contenido que se enviará al usuario.

Twitter también realiza un seguimiento de la cantidad de tuits que ha realizado que están relacionados con un tema.

# Entrega El Modelo De Agrupación Y Reducción De Dimensionalidad En Un Repositorio Según La Herramienta De Software Utilizada.

Aplicar agrupación y reducción de dimensionalidad en un repositorio de datos de Twitter puede ser muy útil para analizar y comprender mejor los patrones y temas dentro de los tweets.

Paso 1: Recopilación de Datos

Utiliza la API de Twitter o herramientas de terceros para recopilar tweets relevantes para tu análisis. Puedes especificar palabras clave, ubicaciones, fechas, usuarios, etc., para obtener un conjunto de datos significativo.

Paso 2: Pre Procesamiento de Texto

El preprocesamiento de texto es esencial para limpiar y transformar los tweets en un formato adecuado para el análisis. Esto podría incluir la eliminación de menciones, hashtags, enlaces, signos de puntuación y palabras vacías. También puedes realizar la tokenización y la matización para simplificar las palabras a su forma base.

Paso 3: Representación Vectorial

Convierte los tweets preprocesados en una representación numérica que pueda ser utilizada por los algoritmos de agrupación y reducción de dimensionalidad. Una técnica común es usar la representación TF-IDF (Frecuencia de Término-Inversa de Frecuencia de Documento).

Paso 4: Agrupación (Clustering)

Aplica técnicas de agrupación como K-Means o DBSCAN para agrupar los tweets similares en temas o categorías. Esto te permitirá identificar las conversaciones más relevantes en tu conjunto de datos.

Paso 5: Reducción de Dimensionalidad

Dado que los datos de texto a menudo tienen una alta dimensionalidad, la reducción de dimensionalidad es útil para la visualización y análisis. Puedes aplicar el Análisis de Componentes Principales (PCA) u otras técnicas como t-SNE para proyectar los datos en un espacio de menor dimensión mientras se conserva la estructura.

Paso 6: Visualización

Visualiza los resultados de los pasos anteriores para comprender mejor los patrones y temas dentro de los tweets. Puedes usar gráficos de dispersión, mapas de calor o diagramas de dendrogramas para representar clusters y relaciones entre ellos.

Paso 7: Interpretación

Interpreta los resultados de la agrupación y la reducción de dimensionalidad para obtener información significativa. Examina los clusters y las palabras clave más importantes en cada uno. Esto te permitirá extraer ideas y tendencias relevantes de tus datos de Twitter.

Paso 8: Ajuste y Optimización

Experimenta con diferentes parámetros en los algoritmos de agrupación y reducción de dimensionalidad para obtener resultados más precisos y relevantes para tu análisis específico.

Recuerda que la aplicación real puede variar según tus objetivos y los datos específicos que estás analizando. También puede ser útil aplicar técnicas de procesamiento de lenguaje natural (NLP) más avanzadas si deseas un análisis más profundo de los contenidos de los tweets.

# Conclusión

Twitter utiliza una variedad de técnicas de agrupación, algoritmos y reducción de dimensionalidad para gestionar y analizar los datos generados por sus usuarios. La agrupación es útil para organizar grandes volúmenes de datos en grupos o clústeres basados en similitudes y los algoritmos son conjuntos de instrucciones que ayudan a procesar datos y tomar decisiones automáticamente. Los datos generados en Twitter son multidimensionales y pueden ser difíciles de manejar eficientemente por lo que la reducción de dimensionalidad se utiliza para disminuir la complejidad de los datos mientras se conserva

