

IDENTIFICACIÓN DEL PROBLEMA:

La empresa de neighborhood necesita el desarrollo de una herramienta de software para realizar un análisis de usuarios en la red social de Twitter, específicamente quieren saber sus gustos en ciertas categorías y cómo influyen esos gustos en las personas que conocen.

Nombre	R-1: Añadir un usuario al programa
Descripcion	Se agrega a la estructura de datos apropiada para poder analizar sus atributos en conjunto con los otros usuarios.
Entradas	La información del perfil de Twitter en datos brutos.
Salidas	Ninguna.

Nombre	R-2: Consultar clasificación de los usuarios por categorías
Descripcion	Permite ver un listado de los usuarios con su respectiva puntuación de una categoría específica.
Entradas	Categoría a consultar (Deportes, Ciencia o Política).
Salidas	Un listado ordenado de mejor a peor puntaje de los usuarios del programa cada usuario (Nombre, puntaje).

Nombre	R-3: Consultar usuarios que coinciden por Hashtags (#)
Descripción	Permite conocer cuales usuarios están conectados por los mismos hashtags.
Entradas	Ninguna.
Salidas	Listado de los usuarios que comparten mismos hashtags.

Nombre	R-4: Consultar usuarios que coinciden por arroba(@)
Descripción	Permite conocer cuales usuarios están conectados por los mismos arrobas.
Entradas	Ninguna.
Salidas	Listado de los usuarios que comparten mismos arrobas.

Nombre	R-5: Consultar los usuarios que superen un puntaje y su grado de parentesco desde otro usuario.
Descripcion	Muestra todos los usuarios que tengan un puntaje superior en alguna de las categorías (Ciencia, deportes o política) y muestran el grado en que se encuentran desde un usuario.
Entradas	Puntaje a buscar, categoría y usuario de inicio.
Salidas	Listado de los usuarios y el grado de parentesco.

Nombre	R-6: Consultar la futura relación más probable desde un usuario
Descripcion	El programa analiza cuál de los usuarios que estén en el programa y no se relacionen con un usuario específico, se puede relacionar a futuro con este usuario con mayor probabilidad.
Entradas	Un usuario a relacionar.
Salidas	El usuario con más probabilidad de relacionarse.

RECOPIACIÓN DE INFORMACIÓN:

Red Social:

Una red social es una estructura capaz de comunicar entre sí a personas o instituciones. A través de internet se pueden establecer relaciones que creen grupos o comunidades con intereses comunes.

Las redes sociales suponen un contacto ilimitado y a tiempo real. Esto se consigue gracias a la interactividad, uno de sus rasgos más distintivos y novedosos. Antes de las redes sociales, las empresas y los medios no tenían prácticamente formas de poderse comunicar con su público y así conocer su opinión. Las redes permiten establecer un contacto mutuo entre emisor y receptor.

Entramos en una forma de comunicar en la que se rompen las barreras de tiempo y espacio. La comunicación sucede de manera casi inmediata gracias a dichas plataformas.

Marketing en redes sociales:

Hace referencia a todas aquellas actividades realizadas con el fin de promover una marca a través de redes sociales

Realizar marketing en redes sociales se esta volviendo una importante tendencia para la relevancia de cualquier empresa hoy en día. Representa una oportunidad de conectarse con posibles clientes de manera rápida y efectiva.

Uno de los pasos claves para hacer buen marketing en redes sociales parte en encontrar de manera efectiva los potenciales clientes entre todos los usuarios de la plataforma. Esta tarea en particular puede ser facilitada con herramientas de software cómo la que se está tratando de desarrollar en este informe, poder caracterizar usuarios de manera rápida y efectiva resulta entonces útil y muy probablemente lucrativo.

Twitter:

Es una plataforma de comunicación basada en el microblogging bidireccional con naturaleza de red social porque permite elegir con quien te relacionas, las interacciones se generan por medio de tweets mensajes de no más de 280 caracteres.

Las relaciones que se crean son asimétricas. En Twitter los dos extremos de la relación no se ponen simplemente en contacto el uno con el otro, sino que se diferencian entre “seguidores” (“followers”) y “seguidos” (“followed”).

Si sigues a alguien, quiere decir que verás sus “tweets” (los mensajes que escribe en Twitter) en tu “cronología”. Un usuario de Twitter decide a quien seguir, pero la persona a la que sigue no necesariamente tiene que seguirla de vuelta.

Ejemplo de herramientas que analizan información de redes sociales:

Definimos el análisis de redes sociales como la recopilación de datos en redes sociales y otro contenido en línea para guiar las decisiones de negocios. Toma muchas formas, pero siempre se trata de obtener información social procesable para guiar su estrategia empresarial.

Usualmente los datos obtenidos en el análisis resultan beneficiosos cuando la empresa que los obtiene tiene unos objetivos claros que quiere alcanzar con dicha información, cuando determina que clase de indicadores va a usar para determinar si se tuvo éxito o no (como compromiso, sentimiento, conversión).

Algunos ejemplos de herramientas que realizan esta recopilación de datos son:

- Hootsuite ha crecido para ofrecer planes para marcas de tamaño empresarial, así como para individuos y pymes. Con ideas basadas en sentimientos, mapas de conversación, paneles de control personalizables e informes, es un candidato digno para las marcas que necesitan y pueden pagar más.
- Meltwater cuenta con un motor social impulsado por la inteligencia artificial que automatiza la escucha social y los esfuerzos de participación para que las marcas y agencias puedan "convertir billones de puntos de datos propios y obtenidos en perspectivas y predicciones contextuales que impulsen decisiones empresariales más sólidas.
- Sprout Social es otra herramienta sólida para la escucha social que guía los esfuerzos de marketing y marca. Si bien no se mencionan las capacidades de inteligencia artificial, hay muchas cosas que me gustan del lado competitivo e inteligente de la marca, así como integraciones de la bandeja de entrada, monitoreo de tendencias y hashtag, etc.

Impactos contraproducentes del análisis de usuarios en redes sociales:

Cambridge Analytica era una empresa con sede en Londres que usaba el análisis de datos para desarrollar campañas para marcas y políticos que buscan "cambiar el comportamiento de la audiencia", según indica su sitio web.

En concreto, describen como "decisivo" su trabajo en la campaña presidencial de Trump y de otros candidatos republicanos al Congreso de Estados Unidos.

La obtención de perfiles de 50 millones de usuarios de Facebook no fue obra de Cambridge Analytica, sino que se atribuye al profesor de la Universidad de Cambridge Aleksandr Kogan.

A modo de proyecto personal, Kogan desarrolló en 2013 una prueba de personalidad en formato de aplicación de Facebook. Unos 265.000 usuarios completaron la prueba que requería permiso para acceder a información personal y de la red de amigos, sin el consentimiento de estos últimos. Fue así como Kogan se hizo de actualizaciones de estado, "me gusta" y hasta mensajes privados de más del 15% de la población de EE. UU los cuales luego vendió a la empresa.

Los datos obtenidos por Kogan se cruzaron con la información de Facebook para inferir perfiles psicológicos de cada usuario. Así, Cambridge Analytica logró saber cuál debía ser el contenido,

tema y tono de un mensaje para cambiar la forma de pensar de los votantes de forma casi individualizada.

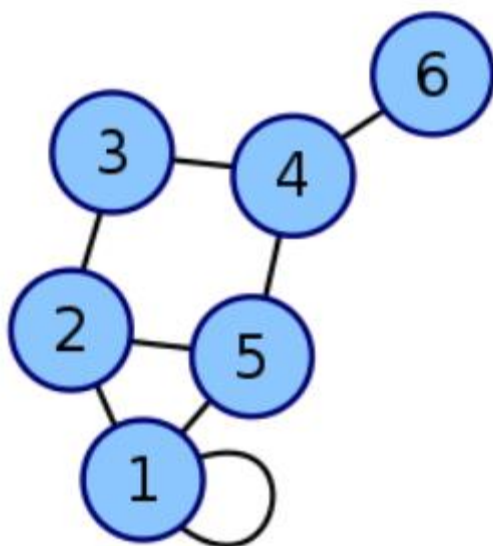
Para Christopher Wylie, un científico de datos y expleado de Cambridge Analytica es claro que: "Si empiezas a deformar la percepción de los votantes sin su consentimiento o conocimiento, esa es una violación básica de su autonomía para tomar decisiones libres, porque están votando en función de cosas que creen que son reales, pero no necesariamente lo son"

Este es un gran ejemplo de la responsabilidad que tenemos como ingenieros de crear herramientas de manera responsable y siempre teniendo en cuenta los posibles efectos que las herramientas que creemos pueden tener en la sociedad.

Grafo:

Un grafo es un conjunto, no vacío, de objetos llamados vértices y una selección de pares de vértices, llamados aristas que pueden ser orientados o no. Típicamente, un grafo se representa mediante una serie de puntos conectados por líneas.

Diferentes situaciones en las que pueden identificarse objetos y relaciones que satisfagan la definición de grafo pueden verse como grafos y así aplicar la Teoría de Grafos en ellos.



Un grafo puede verse como una pareja de conjuntos $G = (V, A)$ donde V es el conjunto de vértices y A es un conjunto de aristas, A es un conjunto formado por parejas de la forma (u, v) tal que $u, v \in V$.

Muchas redes de uso cotidiano pueden ser modeladas con un grafo: una red de carreteras que conecta ciudades, una red eléctrica o la red de drenaje de una ciudad.

BUSQUEDA DE SOLUCIONES CREATIVAS:

Hay muchas maneras de representar las relaciones que se pueden generar entre los usuarios de Twitter.

La estrategia general que usaremos para representar esta situación son los grafos. Ahora bien, hay muchas maneras de representar un grafo, y dependiendo de las operaciones que vayamos a realizar sobre él hay una serie posible de representaciones que resultan más beneficiosas.

Además de la representación del grafo debemos tener en cuenta las características que van a tener las operaciones dado que dentro de cada posible implementación de la solución.

Consideraciones generales:

Dado que debemos almacenar una serie de palabras características para clasificar las preferencias de cada usuario en las 3 categorías, tomaremos la decisión de almacenar dichas palabras en un diccionario dado que su tiempo de búsqueda es constante provee una solución interesante. Cada usuario debe tener un puntaje en cada una de las categorías

Representación de cada usuario: Cada usuario contendrá un HashMap con todos sus tweets, y algunas características fundamentales del perfil de tweeter, además contendrá un puntaje para cada una de las categorías de interés. Al final se deben tener en cuenta de manera particular los hashtags y arrobas que contengan sus tweets, aquí algunas ideas al respecto:

1. **Representación 1:** Cada usuario tendrá un arreglo de hashtags y otro de arrobas, y a la hora de comparar y generar las relaciones entre usuarios recurriríamos a dichas estructuras cuyo tiempo de búsqueda es lineal.
2. **Representación 2:** Cada usuario tendrá dos árboles binarios de búsqueda balanceados organizados alfabéticamente de hashtags y arrobas, y a la hora de comparar y generar las relaciones entre usuarios recurriríamos a dichas estructuras cuyo tiempo de búsqueda es logarítmico.
3. **Representación 3:** Cada usuario tendrá un HashMap de hashtags y arrobas, y a la hora de comparar y generar las relaciones entre usuarios recurriríamos a dichas estructuras cuyo tiempo de búsqueda y recepción son constantes.

Representaciones de las relaciones entre usuarios:

1. Representación de los usuarios en una matriz de incidencias:

Para una matriz de incidencias se analiza la posibilidad de ver las relaciones entre usuarios como una matriz con los nodos en las filas y cada conexión posible entre ellos en las columnas. Así pues, si un par de usuarios están conectados se marcarían de manera distintiva sus posiciones en la misma columna y para representar otra posible conexión se mueva a la columna del lado derecho.

2. Representación de los usuarios en una matriz de adyacencias:

En esta representación los usuarios ocupan las filas y las columnas, para representarlas conexiones entre dos usuarios, de manera direccionada buscamos el usuario de la adyacencia en las filas y a partir de esa coordenada buscamos el usuario incidente en las columnas. En caso de no ser un grafo dirigido determinaríamos que el proceso se hace en las dos direcciones.

3. Representación de los usuarios con listas de adyacencias:

A cada usuario se le asigna una lista de duplas de objetos, con la cantidad de usuarios que tengan en común, en cada dupla hay un usuario al que está conectado y la información relacionada con esa conexión.

Una de las operaciones importantes que realizamos sobre esta estructura será ejecutar el algoritmo de Dijkstra para encontrar las rutas con el peso más bajo entre un usuario y todos los demás. Hay muchas maneras de implementar este algoritmo, aquí hay un par de ideas:

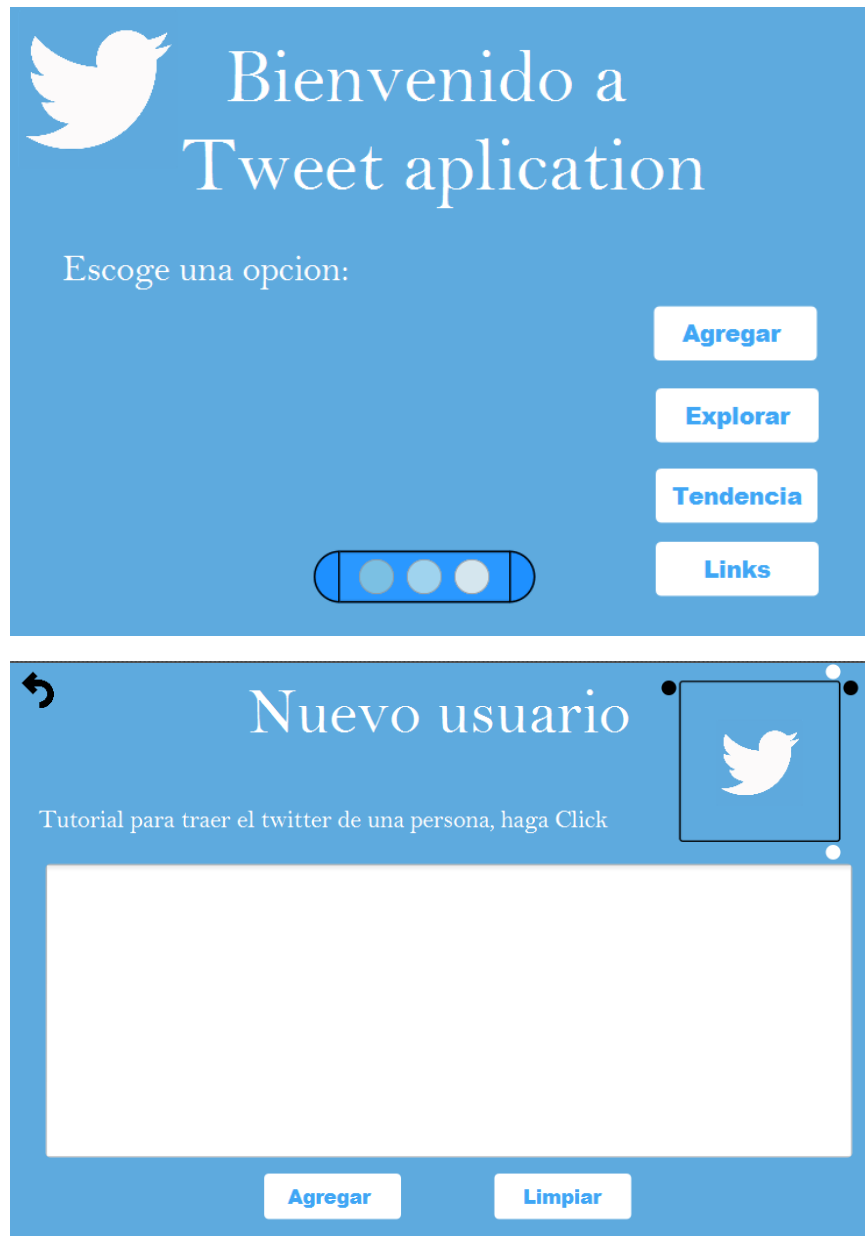
a. Para recorrer el grafo podemos tener un arreglo de booleanos donde cada posición represente un nodo y vamos marcando los nodos recorridos y mientras todos los nodos no estén marcados seguiremos buscando, lo anterior nos daría una complejidad cuadrática para recorrer todas las posibles aristas en el grafo.

b. Para recorrer el grafo podemos tener un priority Queue que ordene los nodos a partir del valor del peso que se tiene actualmente para llegar a ese nodo en particular. El nodo que tenga el peso más pequeño será desencolado y analizado, recorrer todo el grafo de esta forma da una complejidad $Aristas * \log_2(Vertices)$.

TRANSICIÓN DE IDEAS A DISEÑOS PRELIMINARES:

Se opta por usar la opción 3 (**Representación de los usuarios con listas de adyacencias**) puesto que el grafo presenta todas las herramientas para cumplir con los requerimientos y se considera como la implementación que va hacer los métodos de manera más eficiente en un menor costo de tiempo.

algunos Mockups de la interfaz que se espera construir son:





EVALUACIÓN Y SELECCIÓN DE LA MEJOR SOLUCIÓN:

En esta sección evaluaremos dos aspectos importantes para la solución del problema como modelar los usuarios y como modelar las relaciones entre los usuarios.

Criterios para los usuarios:

Criterio 1 – Eficiencia temporal en búsquedas:

[7]- Constante

[5]- Logarítmica

[3]- Lineal

[1]- Logarítmica*Lineal

Criterio 2 – Eficiencia temporal para compararse:

[8]- Lineal

[4]- Logarítmica*Lineal

[1]- Mayor o igual a cuadrática

Criterio 3 – Eficiencia espacial:

[5]- Lineal, solo guardando elementos necesarios

[3]-Lineal, guardando espacios extra

[1]- Mayor o igual a cuadrática

	Criterio 1	Criterio 2	Criterio 3	Total
Representacion 1	3	1	5	9
Representacion 2	5	4	5	14
Representacion 3	7	8	3	18

Criterios para las relaciones entre usuarios:

Criterio 1 – Eficiencia temporal en DFS y BFS:

- [5]- Vértices + Aristas**
- [3]- Vértices * Aristas**
- [1]- Mayor a cuadrática**

Criterio 2 – Eficiencia temporal para Dijkstra:

- [7]- Lineal**
- [5]- Logarítmica*Lineal**
- [3]- Cuadrática**
- [1]- Mayor a cuadrática**

Criterio 3 – Eficiencia temporal para Prim:

- [7]- Lineal**
- [5]- Logarítmica*Lineal**
- [3]- Cuadrática**
- [1]- Mayor a cuadrática**

Criterio 4 – Eficiencia espacial:

- [7]- Lineal, solo guardando elementos necesarios**
- [5]-Lineal, guardando espacios extra**
- [3]- Cuadrática**
- [1]- Mayor a cuadrática**

Criterio 5 - Facilidad de implementación métodos de recorrido:

- [5]- Moderada**
- [3]- Trabajosa**
- [1]– Desafiante**

	Criterio 1	Criterio 2	Criterio 3	Criterio 4	Criterio 5	Total
Representacion 1	1	1	1	1	1	5
Representacion 2	3	3	3	3	3	15
Representacion 3.A	5	3	3	7	5	23
Representacion 3.B	5	5	5	7	5	27

Referencias:

<https://sites.google.com/site/cienciasdelacompuacion/algoritmos-y-estructuras-de-datos/con/historia/campos-de-la-ciencia-de-la-computacion/fuandamentos-de-la-matematica/teoria-de-grafos>

<https://audiovisualsocialmedialover.com/twitter-funcionamiento-basico-consejos/>

<https://www.ciudadano2cero.com/twitter-que-es-como-funciona/>

<https://www.thebalancesmb.com/what-is-social-media-2890301>