

## Representación gráfica del análisis de sentimientos en las redes sociales

Juan Fuentes Saldaña, Dora María Calderón Nepamuceno

<sup>a</sup> Universidad Autónoma del Estado de México, jfuentes.sdy@gmail.com, Valle de Chalco Solidaridad, Estado de México.

<sup>b</sup> Universidad Autónoma del Estado de México, dmcalderonn@uaemex.mx, Ecatepec, Estado de México.

### Resumen

En la era actual, las interacciones en las redes sociales han trascendido su función inicial de los sistemas de comunicación para convertirse en generadores masivos de datos a medida que los usuarios interactúan con su contenido. Frecuentemente, las respuestas de los usuarios a estos contenidos están arraigadas en las sensaciones que estos provocan. La comprensión de estas emociones y la confianza en soluciones técnicas de sincronización de datos son imperativas para obtener un análisis y visualización de datos que capturen plenamente las complejidades de las emociones sociales. Las funciones habilitadas en las redes sociales para rastrear las preferencias de los usuarios mediante herramientas de minería de datos no solo desembocan en análisis, sino que también generan resultados fundamentados en algoritmos intrincados que perfilan las emociones y preferencias de los usuarios, representándolos gráficamente. Esta capacidad no solo permite la creación de imágenes personales, sino también la formulación de representaciones visuales intrigantes tanto para el usuario individual como para aquellos dentro de la red social en su conjunto.

**Palabras clave** — Redes sociales, minería de datos, visualización de datos, algoritmos, soluciones técnicas, emociones.

### Abstract

*In the current era, social media interactions have transcended their initial role as the communication systems to become massive data generators as users interact with their content. Frequently, users' responses to this content are rooted in the feelings it elicits. Understanding these emotions and relying on technical data synchronization solutions are imperative to obtain data analytics and visualization that fully capture the complexities of social emotions. The functions enabled in social networks to track user preferences through data mining tools not only lead to analytics, but also generate results based on intricate algorithms that profile user emotions and preferences, representing them graphically. This capability not only enables the creation of personal images, but also the formulation of intriguing visual representations for both the individual user and those within the social network as a whole.*

**Keywords**— Social networks, data mining, data visualization, algorithms, technical solutions, emotions.

## 1. INTRODUCCIÓN

En el contexto actual, las redes sociales han adquirido una posición central en la vida de las personas. Estas plataformas son utilizadas por los usuarios para compartir sus vivencias cotidianas y diversas situaciones con una audiencia variada [1]. En particular, Twitter se destaca como una red social desafiante debido a su contenido explícito y a los foros abiertos de expresión de opiniones. Las discusiones continuas sobre temas fundamentales perturban la integridad psico-emocional de los usuarios, generando una serie de eventos emocionales que influyen en su comportamiento social, sesgando la comunicación y la libertad de expresión, y generando conflictos internos con su estado de ánimo [2].

Por otro lado, en el ámbito de la informática, los datos desempeñan un papel fundamental en disciplinas como el Deep Learning, la Inteligencia Artificial y la Tecnología de la Información. La Minería de Datos (data mining) se presenta como una herramienta que vincula estas áreas [3]. Este proceso automatizado se encarga de extraer información de grandes bancos de datos, buscando patrones que revelen tendencias o reglas explicativas del comportamiento de los datos en contextos específicos, dependiendo de su variabilidad [4].

La minería de datos ha ganado relevancia a nivel mundial gracias a su eficacia en el manejo de conjuntos de datos enormes y de difícil interpretación, especialmente en el ámbito empresarial [5]. Facilita a las empresas la detección de patrones de conducta para la predicción de comportamientos estadísticos. Este proceso, fundamentado en algoritmos matemáticos, ha simplificado tareas anteriormente complejas, convirtiendo la búsqueda de información en grandes conjuntos de datos en una labor rutinaria para muchas empresas [6].

La gestión de grandes volúmenes de datos, común en empresas internacionales, implica la minería de datos como una actividad cotidiana. Diversas herramientas, entre ellas R [7], se han vuelto indispensables en este proceso. R, un lenguaje de programación ampliamente adoptado en los últimos años, destaca por su utilidad en la manipulación de datos, análisis estadísticos y visualización, convirtiéndose en una herramienta esencial para empresas que buscan optimizar sus procesos [8].

Este trabajo de investigación se centra en la obtención, limpieza y análisis de datos, proporcionando una visión más clara de cómo las redes sociales se han convertido en un componente crucial de la comunicación y la expresión libre de las personas. Los gráficos resultantes ofrecen una perspectiva media de los sentimientos expresados en los tweets, proporcionando una comprensión más profunda de la interacción emocional en estas plataformas. Además, mediante el uso de la Inteligencia Artificial, se logran gráficos personalizados basados en el análisis de una amplia gama de sentimientos, enriqueciendo aún más nuestra comprensión de la complejidad emocional en las redes sociales.

## 2. CONTENIDO

En el desarrollo de esta investigación, se empleó la metodología KDD (Knowledge Discovery in Databases), una estrategia que engloba diversas fases críticas. Estas etapas abarcan desde la selección inicial hasta la interpretación y evaluación de los resultados. El proceso inicia con la selección cuidadosa de datos, seguida de la limpieza y preprocesamiento de los mismos. Posteriormente, se lleva a cabo la transformación de datos en un formato apropiado para el análisis, se eligen técnicas de minería de datos pertinentes, se aplican estas técnicas a los datos y se concluye con la interpretación y evaluación de los resultados obtenidos.

Es imperativo destacar que la aplicación de la metodología KDD en este trabajo se sustentó en una cuidadosa elección y aplicación de las fases específicas de la metodología. Cada fase fue seleccionada y ejecutada con un propósito claro en el contexto del análisis de sentimientos en las redes sociales. Este enfoque garantiza no solo la realización de las acciones necesarias, sino también la comprensión y justificación de la elección de cada fase en función de la aportación única que brinda a la investigación. La meticulosa aplicación de las fases de la metodología KDD se erige como la columna vertebral de la investigación, respaldando y enriqueciendo la contribución esencial que este trabajo presenta al campo del análisis de sentimientos en redes sociales.

### 2.1 Selección de Datos

Para llevar a cabo el análisis de sentimientos mediante la API de Twitter, el usuario debe conectarse utilizando las credenciales de desarrollador y una librería de programación como Tweepy en Python. Posteriormente, se realiza una consulta a la API para obtener los tweets deseados, ya sea por palabras clave o usuarios específicos.

Una vez obtenidos los tweets, se emplean herramientas de procesamiento de lenguaje natural, como NLTK o spaCy, para realizar el análisis de sentimientos. Estas herramientas clasifican los tweets en positivos, negativos o neutrales según el tono de las palabras. La información resultante se almacena en un archivo CSV o en una base de datos para análisis y visualización posteriores.

Es esencial considerar la complejidad y subjetividad del análisis de sentimientos, recomendándose el uso de diversas herramientas y técnicas para obtener resultados más precisos. Además, se debe seguir las políticas de uso de la API de Twitter y respetar la privacidad de los usuarios al manejar sus datos.

### 2.2 Procesamiento de Datos

Como primera instancia en el caso de uso, dentro del script se ingresa el usuario a quien evaluaremos y haremos la recolección de sus datos. (Ilustración 1 Ingreso de usuario en código)

Habiendo pasado esta parte se procede a la limpieza de datos sobre los tweets recolectados, pero ¿Para qué hacemos una limpieza de datos? Principalmente se realiza para identificar los tweets orgánicos, nos referimos a esto de la forma en que estos sean: originales, escritos por el usuario, retweets y respuestas del usuario.

```
#Se extraeran los datos del usuario
tweetsMyAccount = get_timeline(user = "@elonmusk", n = 3500, lang="es",
                               include_rts = T)
saveRDS(tweetsMyAccount, "myTweets.RDS")
tweetsMyAccount %>%
  head()
```

Ilustración 1. Ingreso de usuario en código

De lleno en la API podremos hacer referencia a estos datos, gracias a las columnas que llevan por nombre "is\_retweet" y "reply\_to\_status\_id".

Continuando con la limpieza de los mismos, profundizaremos más y a esto se agregará la eliminación de hipervínculos, menciones y puntuación, esto a través de las funciones "str\_replace\_all()", "removeNumbers()" y "removePunctuation", esto permite analizar el texto y no valores ajenos a este.

Es aquí cuando se hace uso del modelo pre-entrenado de UDPipe, que nos permite tokenizar es decir protege los datos, para que no sean expuestos a los demás, taggear, lematizador encuentran las palabras idénticas e identifica cuantas veces se repiten estas palabras y analizador de dependencias. Así mismo a este se le tiene que especificar el idioma en el cual este trabajará, para nuestro caso lo haremos con el modelo en español.

### 2.3 Transformación de Datos

De esta manera y gracias a la limpieza de datos realizada con anterioridad, se procede al análisis de sentimientos que es lo que realmente interesa. Para esta parte se hará uso de una técnica de diccionarios y unigramas que lleva por nombre "NRC Emotion Lexico", el cual es una lista de palabras en inglés y sus asociaciones con 8 emociones básicas, como pueden ser: ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto. Estos se dividen en 2 sentimientos (negativo y positivo). El análisis se realiza a través de la función "get\_nrc\_sentiment()", lo cual se optimizará al paralelizarlo.

Gracias a esto, se demostrará mediante el código y el modelo anteriormente mencionado la creación de un diccionario cuyo objetivo principal es poder seleccionar las palabras en las 10 categorías, todo esto a través de un algoritmo de búsqueda intensiva.

## 2.4 Minería de Datos

De esta manera y con la ayuda de los plots integrados de R se grafican los resultados obtenidos, los cuales quedan de la siguiente manera. (Ilustración 2 Representación gráfica de los sentimientos)

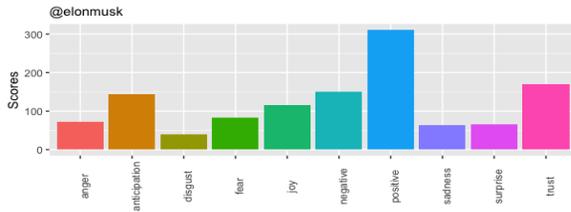


Ilustración 2. Representación gráfica de los sentimientos

De esta manera y habiendo ya recolectado los sentimientos sobre el timeline de tuits, es momento de pasar a diseñar el dibujo con base a los resultados obtenidos acerca de los sentimientos.

Por último, destacando la colección de datos obtenidos previamente al realizar el análisis de sentimientos con los algoritmos correspondientes, nos centraremos esta vez en una librería de R que no permite desarrollar arte con tan solo proporcionándole datos de entrada. El visualizar los resultados de nuestros sentimientos de forma gráfica, puede ser lo ideal para algunos intérpretes, pero el ser humano tiene la capacidad de apreciar y desarrollar arte y que los resultados se muestren por medio de una obra personal, la interpretación se convierte en algo más profundo.

## 3. CONCLUSIONES Y RECOMENDACIONES

### 3.1 Presentación de Resultados

Como se observa, se toma en consideración y se nombra la función con los argumentos previamente mencionados, pasa por una función de R preestablecida para la generación y disminución de ruido en un Data set. Se ingresa el Data frame donde se alberga la cantidad correspondiente de datos guardados en variables, las cuales se pueden modificar para otorgar diferentes valores a los argumentos dependiendo la imagen resultante a buscar. Se realizan los ajustes correspondientes con KNN y este devuelve el objeto grafico que se proyectara en un frame de R. Para finalizar haciendo

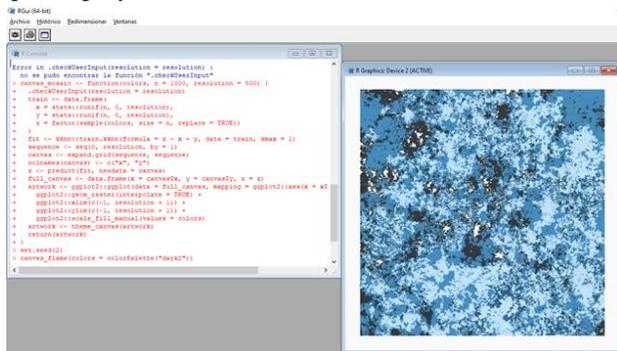


Ilustración 3. Resultados

uso de la imagen generada es necesario en consola transcribir el siguiente comando:

```
canvas_music(colors = colorPalette("dark2"))
```

Los resultados obtenidos con el data set integrado y previamente analizado de un usuario en específico, con base a los sentimientos proyectados de manera estadística ahora se conforma en una sola imagen, que expresa dicha susceptibilidad y contrastes que se obtiene al realizar alguna interacción un dicha red social en la cual nos centramos principalmente en este proyecto y como se mencionaba con anterioridad para cada usuario existe un resultado diferente aleatorio que ahora a los usuarios les ayudara a visualizar más allá de su estado de ánimo de manera grafica en una imagen, los convertirá en sus artistas de sus propias emociones. (Ilustración 3 Resultados)

De esta manera y gracias a la limpieza de datos realizada con anterioridad, se procederá al análisis de sentimientos que es lo que realmente interesa. Para esta parte se hará uso de una técnica de diccionarios y unigramas que lleva por nombre “NRC Emotion Lexico”, el cual es una lista de palabras en inglés y sus asociaciones con 8 emociones básicas, como pueden ser: ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto. Estos se dividen en 2 sentimientos (negativo y positivo). El análisis se realiza a través de la función “get\_nrc\_sentiment()”, lo cual se optimizará al paralelizarlo.

Gracias a esto, se demuestra mediante el código y el modelo anteriormente mencionado la creación de un diccionario cuyo objetivo principal es poder seleccionar las palabras en las 10 categorías, todo esto a través de un algoritmo de búsqueda intensiva.

Además de los puntos mencionados anteriormente, también se puede agregar que el análisis de sentimientos en las redes sociales tiene un gran potencial para la investigación de opiniones y comportamientos de grupos específicos. Por ejemplo, se puede analizar el sentimiento de un grupo de clientes hacia un producto o servicio, lo que puede ayudar a las empresas a mejorar la satisfacción del cliente y desarrollar estrategias de marketing más efectivas. [9]

Otro beneficio del análisis de sentimientos es que puede ser utilizado en la detección de tendencias y eventos en tiempo real. Por ejemplo, se pueden analizar los sentimientos en las redes sociales durante un evento en vivo, lo que puede proporcionar información valiosa para los organizadores y patrocinadores. [10]

También es importante destacar que el análisis de sentimientos en las redes sociales puede ser utilizado en diversos campos, como la política, la salud mental y el bienestar social. Por ejemplo, se puede analizar el sentimiento de un grupo de personas hacia un tema específico de salud mental, lo que puede ayudar a los investigadores y profesionales a desarrollar tratamientos más efectivos y estrategias de intervención.

### 3.2 Recomendaciones Adicionales

Además de los puntos mencionados anteriormente, se puede agregar que el análisis de sentimientos en las redes sociales puede beneficiarse de diversas herramientas y técnicas. Las siguientes recomendaciones adicionales buscan proporcionar una guía más detallada para mejorar la calidad y la aplicabilidad de los resultados obtenidos:

#### 3.2.1 Variedad de Técnicas

Para obtener resultados más robustos y abordar la complejidad del análisis de sentimientos en redes sociales, se recomienda implementar una variedad de técnicas, como:

- **Análisis Léxico:** Utilizar diccionarios de palabras y su polaridad para clasificar los sentimientos.
- **Aprendizaje Automático:** Aplicar algoritmos de aprendizaje automático, como máquinas de soporte vectorial o redes neuronales, para mejorar la capacidad predictiva.
- **Enfoques Basados en Reglas:** Incorporar reglas específicas para capturar matices y contextos que pueden escapar a otros métodos.

#### 3.2.2 Herramientas de Procesamiento de Lenguaje Natural (PNL)

Implementar herramientas avanzadas de Procesamiento de Lenguaje Natural (PNL), como spaCy o NLTK, para mejorar la precisión en la clasificación de sentimientos y permitir un análisis más profundo del contenido textual.

#### 3.2.3 Comparación de Resultados

Realizar una comparación detallada de los resultados obtenidos mediante diferentes técnicas y herramientas. Esto permitirá identificar cuáles son más efectivas en el contexto específico de la investigación, proporcionando insights sobre la idoneidad de cada enfoque.

#### 3.2.4 Visualización Avanzada

Explorar herramientas de visualización avanzada, como gráficos interactivos o representaciones tridimensionales, para mejorar la representación gráfica de los resultados y facilitar su interpretación.

#### 3.2.5 Incorporación de Datos Contextuales

Considerar la inclusión de datos contextuales, como la hora del día, eventos externos o tendencias en redes sociales, para enriquecer el análisis de sentimientos y proporcionar insights más profundos sobre las circunstancias que influyen en las interacciones en línea.

### 3.3 Conclusión General

En el presente trabajo, se abordó la relevancia de las redes sociales en la sociedad contemporánea, centrándonos particularmente en el desafío que plantea Twitter debido a su contenido explícito y foros abiertos de expresión. La integridad psico-emocional de los usuarios se ve afectada por las discusiones incesantes sobre temas fundamentales, generando eventos emocionales que impactan en su comportamiento social y su libertad de expresión.

En el ámbito de la informática, la minería de datos se erige como una herramienta fundamental, vinculando disciplinas como el Deep Learning, la Inteligencia Artificial y la Tecnología de la Información. La gestión eficaz de grandes volúmenes de datos se ha vuelto esencial, y herramientas como R desempeñan un papel indispensable en la extracción de patrones y la simplificación de tareas complejas.

Una de las contribuciones clave de este trabajo radica en la capacidad para procesar grandes cantidades de datos de manera eficiente. Esto posibilita la realización de cálculos y representaciones más limpias, permitiendo la visualización de la comunicación en redes sociales y el establecimiento de estándares de interacción. De manera específica, se detalló la metodología KDD, que abarca desde la selección y limpieza de datos hasta la interpretación y evaluación de los resultados.

En el proceso de selección de datos, se utilizó la API de Twitter para recolectar tweets, seguido de un exhaustivo proceso de limpieza que incluyó la identificación de tweets orgánicos y la eliminación de elementos no deseados. El procesamiento de datos involucró el uso de herramientas como NLTK, spaCy y modelos pre-entrenados de UDPipe, allanando el camino para un análisis de sentimientos preciso y contextual.

La transformación de datos reveló el uso del "NRC Emotion Lexico" para clasificar palabras según ocho emociones básicas, proporcionando una base sólida para el análisis de sentimientos. La minería de datos culminó en representaciones gráficas que reflejan la variabilidad emocional en el timeline de tweets de un usuario específico.

Adicionalmente, se ofrecen recomendaciones para mejorar y ampliar el alcance del análisis de sentimientos, proponiendo la implementación de diversas técnicas, herramientas avanzadas de PNL y la incorporación de datos contextuales.

En síntesis, este trabajo no solo proporciona una visión detallada del análisis de sentimientos en Twitter, sino que también sienta las bases para futuras investigaciones, destacando la importancia de la versatilidad y la profundidad en las herramientas y técnicas utilizadas para comprender de manera integral las complejidades de las interacciones en redes sociales.

#### 4. REFERENCIAS

- [1] Aguilar, D., Sidorov, G., & Batyrshin, I. (25 de 04 de 2018). Caso de estudio de análisis de sentimientos en Twitter: Tratado de libre comercio de América del Norte.
- [2] Hernández Esquivias, P. (05 de 2021). Análisis de sentimientos en Twitter en castellano con redes neuronales recurrentes LSTM.
- [3] Sande, J. C. (06 de 2018). Análisis de sentimientos en twitter.
- [4] Saura, J. R., Reyes Mendez, A., & Palos Sanches, P. (01 de 06 de 2018). Un Análisis de Sentimiento en Twitter con Machine Learning: Identificando el sentimiento sobre las ofertas de #BlackFrida.
- [5] Fayyad, U., & Stolorz, P. (1997). Data mining and KDD: Promise and challenges. *Future generation computer systems*, 13(2-3), 99-115.
- [6] Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Using data mining to detect health care fraud and abuse: a review of literature. *Global journal of health science*, 7(1), 194.
- [7] Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2), 1-41.
- [8] Sarlan, A., Nadam, C., & Basri, S. (2014, November). Twitter sentiment analysis. In *Proceedings of the 6th International conference on Information Technology and Multimedia* (pp. 212-216). IEEE.
- [9] Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics* (pp. 1041-4347). New York: Springer.
- [10] Bach, M. (2018). Data mining and knowledge discovery for big data: Methodologies, challenge and opportunities. In *Proceedings of the 3rd International Conference on Big Data Analysis and Data Mining* (pp. 43-47).