

Análisis de sentimientos en mensajes de WhatsApp

Gustavo Amador, Luis Fernando Amey and Javier Hernández

Resumen Esta investigación analiza los sentimientos expresados en mensajes de un grupo de WhatsApp de estudiantes de Ciencias Actuariales mediante técnicas de procesamiento de lenguaje natural y algoritmos de agrupamiento no supervisado. A partir de la vectorización emocional de los mensajes usando un lexicón en español, se compararon diversos métodos de clusterización como Mapas Auto-Organizativos (SOM), DBSCAN, clustering jerárquico y modelos de mezcla, evaluados con métricas internas de cohesión y separación. El SOM fue seleccionado como el modelo más adecuado por su equilibrio en estas métricas y su capacidad para representar la estructura emocional compleja de los datos. El análisis reveló clústeres con emociones predominantes y patrones comunicativos diferenciados entre participantes. Este trabajo también aborda desafíos técnicos en la limpieza, adaptación del diccionario y limitaciones del análisis palabra por palabra, aportando así una visión profunda sobre la dinámica emocional en chats grupales académicos.

Keywords: análisis de sentimientos, procesamiento de lenguaje natural, clusterización no supervisada, mensajes de WhatsApp.

Gustavo Amador, e-mail: gustavo.amador@ucr.ac.cr
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

Luis Fernando Amey, e-mail: luis.amey@ucr.ac.cr
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

Javier Hernández, e-mail: javier.hernandeznavarro@ucr.ac.cr
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

1. Introducción

Durante el I Ciclo del 2024, en el curso Herramientas de Ciencia de Datos II, se desarrolló un proyecto centrado en el análisis de conversaciones grupales de WhatsApp y Telegram, con el objetivo de explorar el uso de herramientas de procesamiento de lenguaje natural para identificar patrones de comunicación y evaluar el sentimiento expresado en los mensajes.

Utilizando Python y bibliotecas como *NLTK*, *TextBlob* y *VADER*, se procesaron los datos de dos chats grupales: uno de estudiantes de Ciencias Actariales y otro del curso oficial. Se aplicaron técnicas de limpieza y preprocesamiento de texto, y se analizaron aspectos como la frecuencia de mensajes, el uso de multimedia y la polaridad emocional (positiva, negativa o neutral). Entre los hallazgos destacan los días con mayor actividad, los usuarios más participativos y las tendencias emocionales predominantes.

En este nuevo proyecto se propone un desafío técnico más ambicioso: lograr comparar modelos de clasificación no supervisada para encontrar una agrupación tanto buena como interpretable. Se investigará sobre los métodos de clasificación tales como DBSCAN y los Mapas Auto-Organizativos (SOM) para poderlos implementar en el análisis de sentimientos.

Este enfoque permitió comprender más a fondo los fundamentos del análisis de sentimientos y comparar los resultados obtenidos con los de herramientas convencionales, documentando todo el proceso, los retos enfrentados y los aprendizajes adquiridos.

2. Marco teórico

La pregunta de investigación que guía este proyecto es: ¿Qué sentimientos predominan en los mensajes de texto del grupo de WhatsApp de la generación de 31 estudiantes de Ciencias Actariales de la Universidad de Costa Rica, identificados mediante algoritmos de clasificación?

Este trabajo se basa en el análisis de sentimientos aplicado a mensajes de WhatsApp, utilizando algoritmos de clasificación para identificar patrones emocionales. Un mensaje de WhatsApp, entendido como una unidad de comunicación escrita dentro de una plataforma digital de mensajería instantánea, ofrece un entorno propicio para estudiar emociones expresadas de manera cotidiana y espontánea (Meta Platforms, Inc., 2025; Church and de Oliveira, 2013). Dado que estos mensajes no están limitados por formato ni longitud, reflejan una amplia gama de expresiones afectivas.

Los algoritmos de clasificación, por su parte, son métodos de aprendizaje automático que asignan clases predefinidas a datos nuevos a partir de ejemplos conocidos (Murel and Kavlakoglu, 2024). En este proyecto, se utilizan para categorizar los mensajes según su polaridad emocional (positiva, negativa o neutra), siguiendo

do el enfoque del análisis de sentimientos, que forma parte del Procesamiento del Lenguaje Natural (PLN) (Sande, 2018).

El marco teórico se complementa con tres enfoques relevantes. Primero, la teoría del PLN, presentada por Jurafsky y Martin, proporciona herramientas para que las máquinas interpreten el lenguaje humano, incluso en contextos informales o ambiguos (Jurafsky and Martin, 1999). Segundo, la teoría de los actos de habla de Austin permite comprender los mensajes como actos comunicativos que reflejan intenciones emocionales, más allá del contenido literal (Austin, 1962). Tercero, la teoría de la comunicación emocional, según Mohammad, destaca la importancia de detectar emociones implícitas presentes en el lenguaje digital moderno, a menudo expresadas mediante abreviaciones, emojis o puntuación creativa (Mohammad, 2021). Estos marcos conceptuales permiten abordar con mayor precisión el análisis de sentimientos en un entorno tan dinámico como un chat grupal académico.

3. Marco metodológico

El análisis comenzó con un proceso exhaustivo de limpieza y tokenización de los mensajes de WhatsApp, donde se eliminaron elementos irrelevantes como imágenes, enlaces y otros símbolos no textuales utilizando expresiones regulares. Posteriormente, cada mensaje fue transformado en una secuencia ordenada de palabras con la ayuda de la librería `tidytext`. Para representar el contenido emocional de cada mensaje, se utilizó el lexicón NRC en español, evaluando la presencia de palabras asociadas a emociones específicas mediante la función `get_nrc_sentiment()`. Así, cada mensaje m_j fue representado por un vector emocional binario de dimensión 10, que refleja la frecuencia relativa de palabras vinculadas a ocho emociones básicas (como alegría, tristeza, ira, entre otras) y dos polaridades (positiva y negativa):

$$X(m_j) = (x_{e_1}(m_j), x_{e_2}(m_j), \dots, x_{e_{10}}(m_j)) \in \mathbb{R}^{10}$$

donde cada componente $x_{e_k}(m_j)$ indica la frecuencia o presencia relativa de la emoción e_k en el mensaje m_j (Mohammad and Turney, 2013).

1. Mapas Auto-Organizativos (SOM)

El algoritmo SOM fue una pieza fundamental en el análisis por su capacidad para proyectar datos de alta dimensión a un espacio bidimensional manteniendo la topología original, facilitando así la visualización y comprensión de patrones emocionales complejos (Cottrell et al., 2018). La red consiste en una grilla de unidades o nodos, cada uno asociado a un vector de pesos $\mathbf{w}_i \in \mathbb{R}^{10}$. Para un vector de entrada \mathbf{x} , correspondiente a un mensaje, se determina la unidad ganadora o Best Matching Unit (BMU) mediante:

$$i^* = \arg \min_i \|\mathbf{x} - \mathbf{w}_i\|$$

Es decir, la unidad cuyo vector de pesos es más cercano en términos de distancia euclidiana al vector de entrada. La actualización de pesos se realiza tanto para la BMU como para sus vecinos dentro de la grilla, usando la fórmula:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(t) \cdot h_{i,i^*}(t) \cdot (\mathbf{x} - \mathbf{w}_i(t))$$

donde $\alpha(t)$ es una tasa de aprendizaje que decrece con el tiempo, y $h_{i,i^*}(t)$ es una función de vecindad que determina la influencia de la BMU sobre las unidades cercanas, típicamente modelada como una función gaussiana:

$$h_{i,i^*}(t) = \exp\left(-\frac{\text{dist}^2(i, i^*)}{2\sigma(t)^2}\right)$$

con $\text{dist}(i, i^*)$ la distancia entre nodos en la grilla y $\sigma(t)$ un parámetro que también disminuye con el tiempo. Este proceso iterativo permite que la red se auto-organice, agrupando mensajes con patrones emocionales similares en regiones contiguas del mapa.

2. DBSCAN (Density-Based Spatial Clustering)

DBSCAN se seleccionó por su habilidad para detectar grupos de forma arbitraria y manejar ruido sin requerir la definición previa del número de clústeres (Ester et al., 1996). El algoritmo trabaja definiendo un vecindario para cada punto p en el espacio emocional, dado por:

$$N_\epsilon(p) = \{q \in D : \text{dist}(p, q) \leq \epsilon\}$$

donde ϵ es un parámetro que fija el radio de búsqueda de vecinos. Según el número de vecinos $|N_\epsilon(p)|$, los puntos se clasifican en núcleos (si $|N_\epsilon(p)| \geq \text{MinPts}$), fronteras (vecinos de núcleos pero con menos vecinos propios), o ruido (puntos aislados). Este método es especialmente útil para identificar regiones densas de mensajes emocionales similares y diferenciar observaciones atípicas o no clasificables.

3. Métodos Complementarios

Además de los modelos principales, se aplicaron diversas técnicas complementarias para enriquecer el análisis y validar resultados, entre ellas:

- *k-means* y *k-medias*: métodos centroides que buscan minimizar la variabilidad dentro de grupos bajo diferentes medidas de distancia.
- Clustering jerárquico: que construye una dendrograma para identificar estructuras jerárquicas en los datos.
- Modelos de mezcla gaussianos y no gaussianos: para capturar distribuciones complejas mediante componentes probabilísticas.
- Agrupamiento espectral y *mean shift*: métodos basados en análisis espectral del grafo de similitudes y en la estimación de modos de densidad, respectivamente.

Estas técnicas proporcionaron distintos enfoques geométricos, probabilísticos y topológicos para explorar la estructura emocional de los mensajes.

4. Evaluación del Agrupamiento

Para seleccionar el modelo más apropiado y asegurar una segmentación coherente y significativa, se emplearon varias métricas de validación interna que evalúan la calidad de los agrupamientos sin necesidad de contar con etiquetas externas. Estas medidas cuantifican aspectos como la cohesión interna de los clústeres, su separación mutua y la consistencia general de la partición. A continuación, se describen las principales métricas utilizadas:

- **Coficiente de Silhouette** (Rousseeuw, 1987):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Esta métrica calcula para cada punto i qué tan bien se encuentra asignado a su clúster. Aquí, $a(i)$ es la distancia promedio entre el punto i y los demás puntos de su mismo clúster (medida de cohesión interna), mientras que $b(i)$ es la distancia promedio al clúster más cercano distinto al suyo (medida de separación). Un valor cercano a 1 indica una asignación adecuada, valores cercanos a 0 señalan puntos en la frontera entre clústeres, y valores negativos sugieren posibles asignaciones erróneas. El promedio de $s(i)$ sobre todos los puntos ofrece una medida global para comparar modelos y elegir el número óptimo de clústeres.

- **Índice de Calinski-Harabasz** (Caliński and Harabasz, 1974):

$$CH = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \cdot \frac{n - k}{k - 1}$$

Este índice evalúa la calidad de la partición mediante la relación entre la dispersión entre clústeres ($\text{tr}(B_k)$) y la dispersión dentro de los clústeres ($\text{tr}(W_k)$), ajustada por los grados de libertad relacionados con el número de observaciones n y clústeres k . Valores altos indican clústeres bien separados y compactos, favoreciendo particiones con fronteras claras y grupos homogéneos.

- **Índice de Davies-Bouldin** (Davies and Bouldin, 1979):

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{s_i + s_j}{d_{ij}} \right)$$

Esta métrica mide la similitud entre cada clúster y su vecino más parecido, considerando la dispersión interna (s_i, s_j) y la distancia entre centroides (d_{ij}). Un valor bajo del índice indica una buena separación entre clústeres y una baja variabilidad interna, siendo deseable minimizar este índice para obtener agrupamientos claros y diferenciados.

4. Descripción de los datos

La base de datos contiene los mensajes de texto de un grupo de WhatsApp conformado por 31 estudiantes de la generación de Ciencias Actuariales de la Universidad de Costa Rica. Para efectos de este proyecto, se procedió a censurar los nombres de cada autor, asignándoles un ID. Se recopilaron los mensajes enviados desde la creación del grupo, el 1/11/2023, hasta el 2/4/2025. La base está compuesta por 3698 observaciones, donde cada observación corresponde a un mensaje enviado por un estudiante, y se estructura en torno a las siguientes 5 variables:

- **Día** (dia, double(date)): Día en que se envió el mensaje.
- **Hora** (hora, double): Hora en la que se envió el mensaje.
- **Autor** (autor, character): Autor del mensaje.
- **Mensaje** (mensaje, character): Contenido del mensaje enviado. Si el usuario envía una imagen, sticker, audio o documento; en el mensaje saldrá que el respectivo ha sido omitido, porque se está analizando texto. En caso de tratarse de los mensajes del sistema, estos aparecerán de manera regular, solamente que el autor será el nombre del grupo.
- **Editado** (editado, logical): Indica si el mensaje fue editado.

5. Análisis Exploratorio de Datos

El análisis exploratorio permitió caracterizar tanto la estructura como el contenido de los mensajes del grupo de WhatsApp. Se verificó la calidad de los datos y se estandarizó el formato de las variables. Además, se depuró la base eliminando mensajes irrelevantes, como notificaciones automáticas o contenido multimedia omitido, y se descartaron elementos poco informativos para el análisis de sentimientos, como *emojis* y preposiciones frecuentes.

Desde el punto de vista temporal, se observó una actividad concentrada en franjas horarias específicas, con un pico claro alrededor de las 3:00 p.m. y una participación mínima durante la madrugada y los períodos de vacaciones. Estas tendencias reflejan el ritmo del grupo y sugieren que el contexto académico influye en las pautas comunicativas. Por otro lado, en cuanto a los autores, se identificaron diferencias significativas en frecuencia y horarios de participación; algunos escriben de manera constante en franjas acotadas, mientras que otros muestran una distribución más dispersa, evidenciando estilos comunicativos variados.

Finalmente, el análisis léxico permitió construir un mapa de palabras depurado mostrado en la Figura 1, destacando los términos más representativos del grupo. Este insumo fue fundamental para desarrollar un diccionario personalizado de polaridad, que a su vez justificó la implementación de un modelo propio de clasificación. Así, el enfoque adoptado se ajusta específicamente al contexto y al lenguaje característico del grupo, permitiendo capturar con mayor precisión el tono emocional de los mensajes.

Figura 1 Mapa de palabras depurado

“ Fuente: Elaboración propia.

6. Resultados

Para cada método de agrupamiento se definieron los hiperparámetros principales con base en recomendaciones teóricas y pruebas preliminares orientadas a optimizar la calidad del clusterizado según métricas internas de evaluación. En particular, para DBSCAN, el parámetro ϵ , que determina el radio de búsqueda para los vecinos, se fijó en 1, un valor apropiado para los datos escalados que asegura una densidad mínima adecuada sin fragmentar excesivamente los clústeres. Además, MinPts se estableció en 20 para evitar la formación de grupos demasiado pequeños o clasificar como ruido puntos atípicos no representativos.

En el caso de los mapas auto-organizativos (SOM), la cantidad de clústeres se decidió mediante un análisis visual de la topología resultante junto con la evaluación mediante índices internos, seleccionando finalmente 9 clústeres que ofrecen un nivel de detalle suficiente para distinguir patrones emocionales diversos sin comprometer la interpretabilidad.

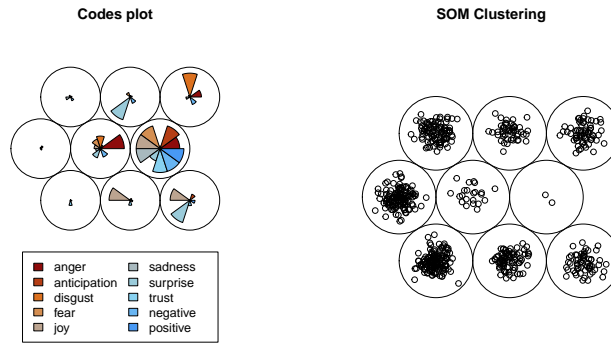
La comparación cuantitativa entre los distintos métodos se presenta en la Tabla 1, mostrando los valores de las métricas de Silhouette, Calinski-Harabasz y Davies-Bouldin. Se destaca que el método *meanshift* alcanzó el valor más alto en Silhouette (0.31), indicando una asignación clara de los puntos a los clústeres, mientras que el clustering jerárquico exhibió el índice Calinski-Harabasz más alto (171.65) y el índice Davies-Bouldin más bajo (1.17), señalando una partición con buena separación y cohesión interna. Sin embargo, el SOM mostró un desempeño robusto con una buena combinación de métricas, lo que, junto con su capacidad para preservar la topología de los datos, lo posicionó como la opción más equilibrada.

El análisis detallado del SOM reveló una estructura compleja pero coherente de los datos emocionales. En la Figura 2 se observa la concentración de mensajes en

Cuadro 1 Comparación de los métodos de clusterización.

método	clusters	silhouette	cal.har	dav.bou
kmeans	5	0.16273	85.97041	1.98684
kmedians	5	0.09895	81.14052	1.98691
jerarquico	5	0.19617	171.65244	1.16791
dbscan	5	0.11451	49.25579	1.66379
gmm	7	0.15318	68.59909	2.48287
som	9	0.23832	123.10330	1.47087
spectral	5	0.22614	108.12655	1.60120
teigen	3	0.26243	93.88190	1.31970
meanshift	5	0.31323	76.44206	1.74662

^a Fuente: Elaboración propia.

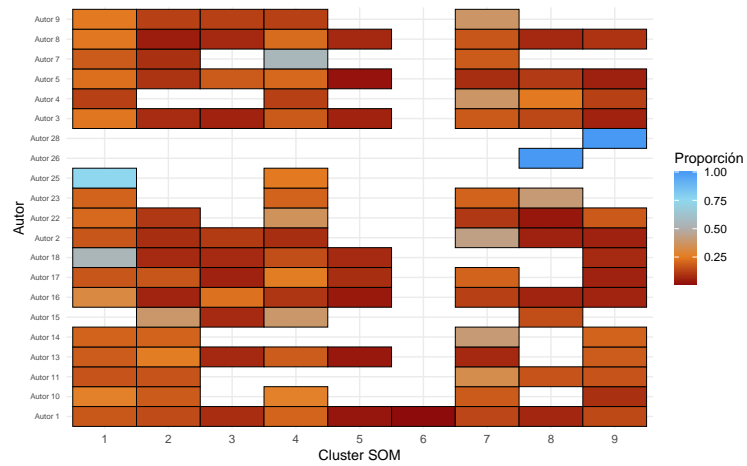
Figura 2 Visualización de la distribución de observaciones por clúster en el mapa auto-organizativo (SOM).

^a Fuente: Elaboración propia.

diferentes nodos y clústeres, mostrando cómo los patrones emocionales se agrupan y separan en el espacio reducido de dos dimensiones. Esta representación permitió identificar regiones del mapa asociadas a emociones predominantes, facilitando la interpretación cualitativa de los resultados.

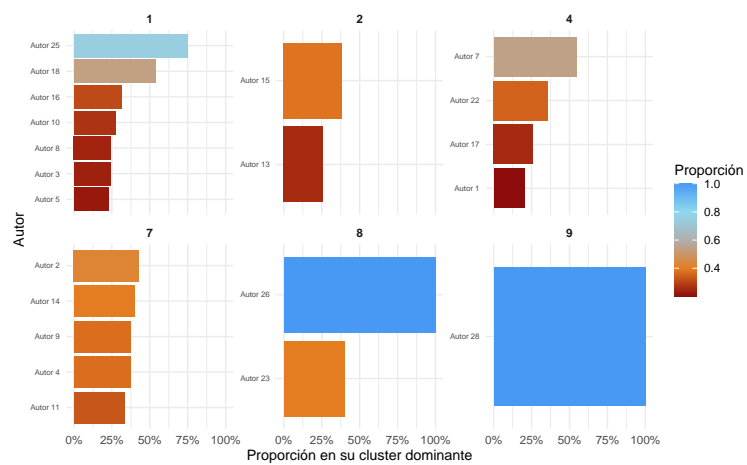
Adicionalmente, la proporción de mensajes por autor en cada clúster, visualizada en un mapa de calor (Figura 3), evidenció diferencias significativas en los estilos comunicativos individuales. Algunos usuarios mostraron una alta concentración en un solo clúster, sugiriendo un estilo emocional más homogéneo, mientras que otros participaron en múltiples clústeres, indicando una variedad emocional más amplia en sus mensajes.

Por último, el clúster dominante por autor (Figura 4) reforzó estas observaciones, señalando que la mayoría de los participantes tienden a expresar predominantemente una o dos emociones, aunque existen casos de dispersión más amplia. Estos patrones podrían estar relacionados con diferencias individuales en la forma de comunicarse o en el contexto de los mensajes, reflejando la riqueza y complejidad emocional del grupo. La combinación de métricas cuantitativas y análisis visual proporcionó

Figura 3 Proporción de mensajes de cada autor por clúster.

^a Fuente: Elaboración propia.

una validación sólida de la clusterización realizada, confirmando que el SOM es una herramienta efectiva para segmentar y comprender las emociones expresadas en mensajes de texto en entornos informales como WhatsApp.

Figura 4 Clúster dominante de cada autor.

^a Fuente: Elaboración propia.

7. Conclusiones

Este proyecto permitió demostrar que los algoritmos de clasificación no supervisada pueden aplicarse con éxito al análisis de sentimientos en mensajes de texto, específicamente en conversaciones de WhatsApp de un grupo de estudiantes. A través del preprocesamiento textual, la vectorización de los mensajes y la implementación de distintos métodos de clusterización, fue posible identificar agrupamientos que reflejan patrones de contenido emocional en la comunicación escrita del grupo analizado.

Para evaluar de manera objetiva el rendimiento de cada algoritmo, se llevó a cabo una comparación basada en métricas internas. Mediante un proceso de eliminación progresiva de los métodos con menor desempeño, se seleccionó el modelo de Mapas Auto-Organizativos como el más adecuado. Este algoritmo no solo obtuvo resultados competitivos en todas las métricas, sino que además ofreció una visualización estructurada que facilitó la interpretación de los clústeres y permitió una mejor comprensión de la estructura emocional subyacente.

El análisis detallado de los clústeres generados por el SOM mostró que cada grupo tiende a presentar una emoción predominante, lo cual sugiere que el modelo logró capturar consistencias afectivas presentes en los mensajes. A nivel individual, se observó que algunos autores concentran su participación en un único clúster, mientras que otros, especialmente quienes envían más mensajes, presentan una distribución más dispersa. Este comportamiento podría reflejar diferencias en el estilo comunicativo o en la variedad emocional expresada en sus intervenciones, aportando una perspectiva valiosa sobre la dinámica grupal.

No obstante, a pesar de los buenos resultados obtenidos, el estudio enfrentó algunas limitaciones importantes que deben considerarse. En primer lugar, el formato variable del archivo exportado desde WhatsApp requirió un proceso manual de limpieza y estandarización, lo cual, aunque efectivo, pudo haber generado pérdidas en la estructura o el contexto original de los mensajes.

Además, la necesidad de traducir diccionarios de análisis emocional del inglés al español introdujo posibles ambigüedades lingüísticas y redujo la precisión en la detección de sentimientos, dado que no todas las palabras tienen equivalentes exactos entre ambos idiomas.

Por otro lado, el análisis realizado a nivel de palabra por palabra, si bien facilitó la implementación técnica y redujo la complejidad computacional, limitó la capacidad para capturar matices más complejos como la ironía o los cambios de tono dentro de un mismo mensaje.

Para superar estas limitaciones en futuros trabajos, se recomienda desarrollar o utilizar diccionarios emocionales específicos para el español con validación rigurosa y explorar técnicas de análisis que consideren unidades textuales mayores, como oraciones o párrafos completos, a fin de captar mejor la riqueza emocional del lenguaje en contextos digitales.

8. Referencias

- Austin, J. L. (1962). *How to Do Things with Words*. Oxford University Press, Oxford.
- Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1):1–27.
- Church, K. and de Oliveira, R. (2013). What’s up with whatsapp? comparing mobile instant messaging behaviors with traditional sms. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 352–361. ACM.
- Cottrell, M., Olteanu, M., Rossi, F., and Villa-Vialaneix, N. (2018). Self-organizing maps, theory and applications. Technical report, Université Paris I, SAMM.
- Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.
- Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*.
- Jurafsky, D. and Martin, J. H. (1999). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Upper Saddle River, NJ.
- Meta Platforms, Inc. (2025). Whatsapp. <https://www.whatsapp.com/>.
- Mohammad, S. M. (2021). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Meiselman, H. L., editor, *Emotion Measurement, 2nd Edition*, pages 213–238. Woodhead Publishing.
- Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.
- Murel, J. and Kavlakoglu, E. (2024). ¿qué son los modelos de clasificación? <https://www.ibm.com/es-es/think/topics/classification-models#:~:text=Los%20algoritmos%20de%20clasificacin%20se,y%20el%20filtrado%20de%20spam>.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.
- Sande, J. C. S. (2018). Análisis de sentimientos en twitter. Trabajo final de máster, Universitat Oberta de Catalunya.

9. Anexos

9.1. Control de Versiones en GitHub

El repositorio se encuentra en el siguiente *enlace*.