



PROGRAMA UNIVERSITARIO
DE ESTUDIOS SOBRE
DEMOCRACIA, JUSTICIA Y SOCIEDAD



DOCUMENTO DE TRABAJO 9

Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad | FEBRERO 2022

Estrategias metodológicas

Análisis de Sentimiento en Twitter

Aproximaciones con métodos
de Aprendizaje de Máquina

Martín Zumaya
Diego Espitia
Luis Ángel Escobar

Autores

FEBRERO 2022





Documento de trabajo 9

Análisis de Sentimiento en Twitter Aproximaciones con métodos de Aprendizaje de Máquina

El presente documento de trabajo fue elaborado por investigadores adscritos al Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad (PUEJJS) de la Universidad Nacional Autónoma de México

Primera edición, febrero 2022

D.R. © Universidad Nacional Autónoma de México,
Programa Universitario de Estudios Sobre Democracia,
Justicia y Sociedad, Torre UNAM-Tlatelolco, Piso 13
Ricardo Flores Magón número 1,
Colonia Nonoalco Tlatelolco Alcaldía Cuauhtémoc,
Código Postal 06995, Ciudad de México
www.puedjs.unam.mx

Cómo citar:

Zumaya Martín, Diego Espitia, Luis Ángel Escobar (2022), "Análisis de Sentimiento en Twitter Aproximaciones con métodos de Aprendizaje de Máquina", Documento de Trabajo núm. 9. PUEJJS, UNAM, México, 34 páginas.



ÍNDICE

Presentación	2
1. Estrategias de análisis de sentimiento	4
1.1. Introducción	4
1.2. Valence Aware Dictionary and sEntiment Reasoner (VADER)	5
1.3. Redes Neuronales Artificiales	7
1.4. Etapas de un modelo de clasificación basado en RNA	11
2. Modelos de evaluación implementados	14
2.1. Modelo de Red Neuronal Artificial	14
2.2. Modelo basado en léxico	20
3. Aplicación de la metodología	22
3.1. Evaluación basada en léxico VADER	23
3.2. Evaluación con Redes Neuronales Artificiales	26
4. Propuestas para la extensión y mejora de los modelos	34
4.1. Ampliación del conjunto de datos de entrenamiento para la RNA	34
4.2. Estrategias de evaluación colaborativa	35
5. Conclusiones	37



Presentación

Las redes socio-digitales se han convertido en un espacio común en el que se genera y comparte contenido alrededor de una gran cantidad de temas, tan variados como los usuarios de la plataformas que participan en ellas. Un vistazo a las publicaciones y al contenido en las plataformas da cuenta de esto, siendo posible encontrar información y comentarios, por ejemplo, sobre conciertos o eventos deportivos, películas o series de televisión de moda, o en otro rango del espectro, alrededor de discusiones políticas o sociales relevantes para las sociedades, como movilizaciones sociales, cambios en la legislación vigente, declaraciones de actores políticos y sociales, entre otros.

El texto es una parte fundamental de las publicaciones en las redes socio-digitales, ya que, aunque puede ir acompañado de imágenes, audio o videos, permite a los usuarios expresarse y comentar el contenido generado por otros usuarios o sobre acontecimientos relevantes. Lo cual, aunque no pueda parecer sorprendente ya que hemos utilizado el lenguaje escrito como especie durante cientos de años para expresar sentimientos, contar historias o narrar eventos por medio de poemas, novelas y crónicas, si representa un fenómeno distinto en el contexto de la comunicación e interacción en medios digitales.

Una de las principales diferencias es la frecuencia y cantidad del contenido que se genera, ya que es posible que los millones de cuentas en las plataformas puedan realizar y compartir publicaciones a discreción. Otra diferencia importante es que todo este contenido, al ser intrínsecamente digital, permite ser almacenada, y sobre todo analizada, por medio de algoritmos y técnicas computacionales que permiten obtener información acerca de ellos, en particular estimar la carga positiva o negativa de un texto. Este tipo de análisis se denomina **análisis de sentimiento**, y forma parte de un conjunto de técnicas computacionales para el análisis del lenguaje, denominada en su conjunto **Procesamiento de Lenguaje Natural**.

Este tipo de análisis se han hecho cada vez más populares, y han sido aprovechadas en gran medida por empresas o marcas para identificar la recepción y percepción de sus productos y servicios, sin embargo, este tipo de técnicas también puede utilizarse para el estudio de culturas políticas, ya que pueden ser útiles, combinados con otros análisis, para la identificación de grupos de cuentas que generan o amplifican contenido con carga positiva o negativa, o para identificar estrategias de criminalización de movimientos sociales, o analizar la carga emotiva de publicaciones en respuesta a acontecimientos relevantes. Sin embargo, cabe hacer notar que estas técnicas identifican el sentimiento de las publicaciones, basado solamente en la información contenida en los textos, partiendo de un conjunto de datos previamente calificados (como los modelos basados en Redes Neuronales Artificiales (RNA)), o al compararlos al nivel de las palabras que los componen con un conjunto de palabras y expresiones también previamente calificadas en relación al tipo de sentimiento o emoción que transmiten (como los modelos basados en léxicos), sin ser posible que los modelos consideren el contexto del evento al que se refieren, o aspectos literarios del lenguaje como la ironía o el sarcasmo.



En este documento, presentamos un par de modelos de análisis de sentimiento, uno basado en léxicos y otro que se compone de una RNA. Así mismo, discutimos resultados obtenidos de su aplicación a un conjunto de datos y proponemos estrategias para la mejoría, tanto de los modelos como de los datos de partida que utilizan.

► 1. Estrategias de análisis de sentimiento

→ 1.1. Introducción

Definir qué es la comunicación es un asunto complejo que no tiene al día de hoy una definición precisa. Etimológicamente, el verbo **comunicar** viene del latín **communicare** (compartir, impartir, difundir) y este de **communis** (común, mutuo) [5].

En términos generales, comunicar hace referencia a la acción de transferir información, y fue esta noción la que inspiró a Claude E. Shannon a formular en 1948 una teoría matemática de la comunicación [18]. Dicha teoría, mejor conocida como teoría de la información, es la pieza fundamental en la existencia de la comunicación digital, y se desarrolló de manera paralela a los avances en la electrónica moderna, cuyo punto de partida se puede establecer en 1947 con la invención del transistor por parte de John Bardeen, Walter Brattain y William Shockley [2].

Esta comunión entre información y electrónica, dio como resultado la invención de las computadoras, que a partir de 1977 con el desarrollo del concepto de **computadora personal**, empiezan a ser producidas masivamente, impactando la vida de millones de seres humanos. Esta revolución se vio potenciada aún más, una vez que fue posible conectar a las computadoras a través de una red, haciendo posible la transferencia de información entre máquinas y dando origen a la Internet.

Ahora bien, en el centro de la transferencia de información, se encuentra el lenguaje que, en su sentido más amplio para el caso de los seres humanos, se da primordialmente a través de las palabras, aunque los gestos y entonaciones de la voz son también muy importantes.

Las palabras, consideradas como los elementos mínimos para la transferencia de información, se corresponden con secuencias de sonidos en el caso del lenguaje hablado, y con secuencias de símbolos (i.e. las letras de los distintos alfabetos) en el caso del lenguaje escrito. Por otro lado las máquinas requieren de un tipo de lenguaje especial llamado lenguaje de máquina, compuesto esencialmente de secuencias binarias (es decir, secuencias de dos símbolos o estados distintos, por ejemplo ceros y unos, o encendido y apagado), que se traducen a impulsos electromagnéticos que permiten dar instrucciones a la CPU para realizar millones de operaciones lógicas por segundo y realizar cálculos. Simplificándolo mucho, se puede decir que los seres humanos entendemos palabras y las computadoras números.

Ésta interacción entre seres humanos y computadoras plantea una serie de problemas de traducción de lenguaje humano-máquina, cuyo estudio se hace a través del procesamiento de lenguaje Natural (NLP por sus siglas en Inglés), el cual surgió en la década de 1950 como una intersección entre la inteligencia artificial y la lingüística [13], compuesta, de

manera general, por un conjunto de técnicas computacionales desarrolladas para el análisis automático y la representación del lenguaje humano [3]. Con estos estudios ha sido posible desarrollar métodos tales como: la recuperación y extracción de información, la traducción automática, detección de palabras clave, resumen de texto.

Sin embargo, con el auge y popularización de las redes socio-digitales; surge la necesidad de desarrollar procesamiento del lenguaje con el objetivo de analizar e interpretar opiniones, sentimientos, emociones, emojis, acrónimos, etc; es decir aspectos subjetivos del lenguaje usado por los seres humanos. Estos procesos son conocidos como **análisis de sentimiento** y son fundamentales para la comprensión de las relaciones e interacciones de los seres humanos en la Internet.

De manera general, podemos entender al análisis de sentimiento como una metodología computacional para calificar cuantitativamente el sentimiento de un texto, y en algunos casos su intensidad, por medio de la asignación de un valor numérico al mismo. Existen principalmente dos tipos de estrategias para hacer este análisis, las basadas en léxicos y las basadas en aprendizaje de máquina [3].

En este trabajo presentamos dos de estas estrategias, Valence Aware Dictionary and sEntiment Reasoner, o VADER (basada en léxicos) y un modelo de clasificación basado en Redes Neuronales Artificiales, las cuales describimos a continuación.

→1.2. Valence Aware Dictionary and sEntiment Reasoner

Valence Aware Dictionary and sEntiment Reasoner, VADER por sus siglas en Inglés, es un modelo basado en un léxico para el análisis de sentimiento, especialmente diseñado para capturar contextos en entornos de redes socio-digitales [9].

El núcleo del método es un léxico anotado de palabras, acrónimos e iniciales relacionados con sentimientos (LOL y WTF por ejemplo), y jerga de uso común con valor de sentimiento (“nah”, “meh”), así como emojis y signos de puntuación, que en caso del idioma Inglés, se componen de cerca de 7520.

Este conjunto de palabras, acrónimos y emojis, que de ahora en adelante llamaremos **tokens** han sido evaluados previamente por seres humanos de acuerdo a la orientación semántica (positivo, negativo, neutro) y a la intensidad del sentimiento.

En esencia el léxico es un diccionario en donde cada entrada corresponde a un token con puntuaciones que indican tanto la polaridad del sentimiento (positivo, negativo, neutro) como la intensidad del sentimiento en una escala de -4 a +4. Por ejemplo, la palabra “**okay**” tiene una valencia positiva de 0.9, “**good**” de 1.9 y “**great**” de 3.1; mientras que “**horrible**” tiene una valencia negativa de -2.5, el emoticono 😞 de -2.2.

La manera como VADER trabaja es muy sencillo, dada una oración, se hace una búsqueda en el diccionario y se identifican los valores para cada una de las palabras. Luego se calcula la puntuación compuesta para la oración sumando las puntuaciones de cada palabra, y normalizándolas para que el resultado esté entre -1 (extremo más negativo) y +1 (extremo más positivo) [9]. Esta puntuación compuesta permite establecer no solo la el sentido de la polaridad de la oración sino también su intensidad. Por ejemplo, la expresión **la comida aquí**

es excepcional tiene un sentimiento positivo, pero dicho sentimiento es más intenso que en la expresión *la comida aquí es buena*.

Para determinar la intensidad, VADER distingue entre distintos tipos de palabras:

- **Impulsores:** palabras que intensifican el sentimiento de los tokens de la oración. Por ejemplo, palabras como “absolutamente” y “asombrosamente”.
- **Atenuadores:** palabras que suavizan el sentimiento de los tokens de la oración. Por ejemplo, palabras como “apenas” y “algo”.
- **Negaciones:** palabras que niegan el sentimiento de los tokens de la oración. Por ejemplo, palabras como “no” y “no es”.

Por ejemplo, al evaluar las siguientes expresiones:

1. Haha , BBQ on the beach , engage smug mode ! 😊 😏 ❤️ #vacation
2. Say it with me - I NEED A #VACATION !!! 😞
3. Say it with me - I need a #vacation 😊

se obtienen valoraciones de 0.9348, -0.5067 y -0.3400 respectivamente. Es importante notar que para este ejemplo, los valores para 2. y 3. son diferentes ya que en la oración 3. se ha puesto la palabra **vacation** en minúsculas y se han omitido los signos de exclamación, haciendo que el sentimiento de dicha expresión sea un poco menos intenso; esto debido a que el método asigna valores a los signos de puntuación y es sensible a las letras mayúsculas (mayor intensidad) y minúsculas (menor intensidad).

Quizá la mayor ventaja de VADER es que su uso es muy sencillo, pues se basa en realizar una búsqueda de los tokens que componen al texto a calificar, en el diccionario del léxico, lo cual es tremendamente eficiente en comparación con métodos de aprendizaje de máquina.

Otro gran acierto de VADER es que la evaluación y clasificación de cada una de las palabras del léxico fue hecha en su totalidad por seres humanos usando la estrategia denominada **sabiduría de la multitud**¹. Cuando una estrategia para evaluar sentimientos es hecha enteramente por seres humanos, se le conoce como **estándar dorado**, y generalmente presenta excelentes resultados frente a otras estrategias tales como la evaluación hecha por expertos, o por estrategias basadas en el aprendizaje de máquina [9].

La principal desventaja que tiene VADER es que ha sido construido en base al Inglés americano ya que las personas que hicieron la evaluación del léxico fueron estadounidenses, y para hacer análisis de sentimiento en otros idiomas la estrategia general es usar los servicios de traducción de internet (Google Translate por ejemplo) [17], para primero traducir el texto original al Inglés y así poder comparar con el léxico para evaluar el texto. Esta estrategia posee múltiples problemas, el primero es que los servicios de traducción normalmente tienen cuotas y permiten traducir una cantidad limitada de caracteres (al rededor de 5000) a menos que se acceda a API's (que pueden ser gratuitas

¹*Wisdom of the crowd en Inglés, esto es usando plataforma colaborativa. Por ejemplo, VADER usó Amazon Mechanical Turk, una plataforma colaborativa que recompensa con micro-pagos a seres humanos por realizar ciertas tareas sencillas*

o no), haciendo poco eficiente la evaluación de los textos, al agregar un paso extra de preprocesamiento. Adicionalmente, los valores asignados al léxico por parte de las personas estadounidenses reflejan aspectos inherentes de su lenguaje y cultura, que no son necesariamente equivalentes en otros; y por lo tanto deficientes para el análisis de sentimiento.

Por ejemplo, expresiones tales como: “Los maestros son vagos”, sería traducido como: “The teachers are lazy”, que sí se traduce de nuevo al español, quedaría “Los profesores son flojos”. En el contexto mexicano la palabra “vago” tiene una connotación más negativa que la palabra “flojo”, haciendo que los resultados obtenidos por VADER, aunque identifican el sentimiento general del texto de manera correcta, pueden presentar distorsiones en la intensidad.

Por ello, es necesario desarrollar estrategias que permitan una mejor evaluación del sentimiento en idiomas distintos al Inglés, más allá de la traducción automatizada de los textos por medio de servicios externos. Como por ejemplo, construir léxicos anotados para el idioma español y sus variantes regionales, que consideren algunas de las expresiones (intensificadores, atenuadores, acrónimos o negaciones) que representan de mejor manera, tanto el sentimiento como la intensidad del mismo en cada contexto en particular.

→1.3. Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA), son un conjunto de métodos computacionales inspirados en las observaciones y estudios de la actividad neuronal del cerebro y el sistema nervioso central [21], que consisten en una colección de elementos interconectados (las neuronas artificiales) que, a grandes rasgos, tienen el objetivo de transformar un espacio de datos de entrada, hacia un espacio de datos de salida deseado [14].

Las RNA son capaces de realizar cálculos y operaciones como: el agrupamiento y clasificación de datos, el reconocimiento de patrones, y problemas de ajuste y aproximación de funciones [19], lo que les ha permitido encontrar una gran diversidad de aplicaciones en una amplia variedad de disciplinas. En el contexto particular del análisis de sentimiento de textos en redes sociodigitales, utilizaremos una RNA como algoritmo de clasificación que nos permitirá ordenar los textos en nuestro conjunto de datos en dos categorías o clases: la positiva y la negativa.

En las siguientes secciones, describimos el componente elemental de las RNA, la neurona artificial, y la metodología general a seguir para la implementación de un sistema de clasificación.

Neuronas artificiales

Las neuronas biológicas son células nerviosas interconectadas capaces de comunicarse por medio de señales eléctricas y químicas, que permiten la transmisión y procesamiento de la información sensorial que recibe el cerebro, y que en su conjunto, permiten procesos tan complejos como la visión y la percepción del color y la temperatura, la acción de fibras musculares para generar el movimiento de las extremidades, entre otros. Las neuronas se componen de 3 elementos principales: las dendritas, que les permiten recibir las señales eléctricas y químicas de otras neuronas, el soma, que procesa la información recibida por la neurona en las dendritas, y genera una respuesta en caso de que la información recibida sobrepase un umbral, la cual se propaga a través del axón, para comunicar dicha señal hacia otro conjunto de neuronas (véase la figura 1.1).

En 1943, Warren McCulloch y Walter Pitts, presentaron uno de los primeros modelos matemáticos inspirados en las características de las neuronas biológicas [12], en el sentido en que el modelo recibe un conjunto de datos de entrada (de la misma forma en que las neuronas biológicas reciben información por medio de sus dendritas), los cuales son procesados por una función de activación (como en el soma de las neuronas biológicas), para obtener un resultado o un conjunto de datos de salida, que pueden comunicarse a otras neuronas artificiales (axón). Esta capacidad de activarse (o “disparar”) en función de la información que recibe de sus vecinas de las neuronas biológicas, les permitió a McCulloch y Pitts, idealizar a las neuronas como dispositivos binarios (con dos estados, encendido y apagado), que al relacionarlo con la naturaleza binaria de las proposiciones lógicas (verdadero y falso), hizo posible la construcción de redes neuronales hipotéticas que permitirían traducir proposiciones lógicas a patrones de conexión entre neuronas [1], y poder realizar cálculos.

Si bien el modelo de McCulloch y Pitts fue trascendental para el desarrollo de la Cibernética y la Inteligencia Artificial [1], el modelo del **perceptrón** introducido por Frank Rosenblatt en 1957 [16], sentó las bases de las RNA como las conocemos (y utilizamos) hoy en día, ya que considera un **algoritmo de aprendizaje** que le permite al modelo ajustar algunos de sus parámetros internos para realizar una tarea.

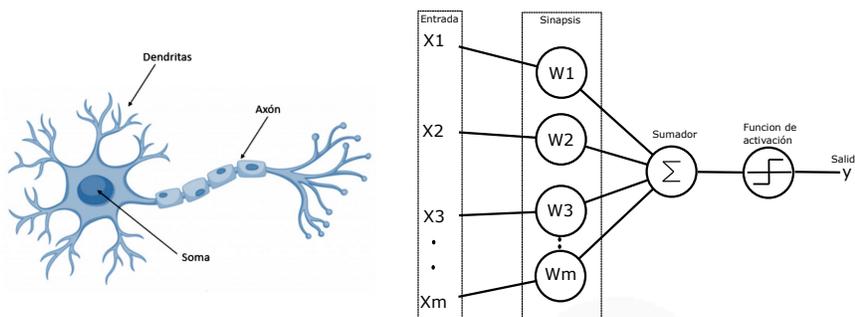


Figura 1.1: Las RNA son un conjunto de herramientas computacionales bioinspiradas, que utilizan las características de las neuronas biológicas (izquierda) como inspiración para la implementación de algoritmos de aprendizaje, como el perceptrón (derecha). Las RNA tienen aplicaciones en una gran variedad de aplicaciones, como el agrupamiento y la clasificación de datos, entre otros.

De la misma forma que el modelo de McCulloch y Pitts, el perceptrón de Rosenblatt se

inspira en la fisiología de las neuronas biológicas, componiéndose de 3 elementos principales: un conjunto de datos de entrada, los cuales, junto con un conjunto de pesos sinápticos, son procesados por un sumador (o integrador) y una función de activación, que resultan en un conjunto de datos de salida (véase la figura 1.1).

Clasificación binaria con un perceptrón

Con el objetivo de ejemplificar el funcionamiento y componentes del perceptrón, consideramos una de las tareas más elementales que puede llevar a cabo: la clasificación binaria. Esta tarea consiste en clasificar un conjunto de elementos $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ en dos clases disjuntas, las cuales nombramos como positiva y negativa y representamos por los valores numéricos 1 y -1 respectivamente.

Los elementos en el conjunto \mathbf{X} , que corresponden a los datos de entrada que recibe el perceptrón, y los pesos sinápticos, que permiten al perceptrón realizar la tarea de clasificación, los representamos por vectores m -dimensionales de números reales, es decir:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}. \quad (1.1)$$

A partir de este par de elementos, definimos la entrada neta, z , para cada elemento del conjunto \mathbf{X} , como

$$z = \sum_{j=1}^m w_j x_j. \quad (1.2)$$

Esta integración de los datos de entrada, es la que será utilizada por el perceptrón para la clasificación de los elementos del conjunto \mathbf{X} , asignando la categoría positiva (con la etiqueta 1) si $z \geq \theta$, y asignando la categoría negativa (con la etiqueta -1) en el caso $z < \theta$, donde θ es el umbral de activación del perceptrón. Este procedimiento de clasificación elemental, lo podemos expresar mediante la variante de la función escalón, $\phi(z)$:

$$\phi(z) = \begin{cases} 1 & \text{si } z \geq \theta \\ -1 & \text{si } z < \theta. \end{cases} \quad (1.3)$$

La función $\phi(z)$ se denomina función de activación, y es la que determina la respuesta del perceptrón en función de la entrada neta z para cada muestra del conjunto \mathbf{X} , que, en este ejemplo sencillo, nos permite clasificar a los elementos del conjunto \mathbf{X} , como miembros de la clase positiva a todos aquellos para los que se cumple la condición $\phi(z) = 1$, y como miembros de la clase negativa, a los elementos para los que se cumple la condición $\phi(z) = -1$.

Hasta el momento hemos considerado que las componentes del vector de pesos sinápticos \mathbf{w} tienen valores fijos, y que son tales que al calcular la entrada neta z , la

función de activación clasificará de manera correcta a los elementos del conjunto X en las categorías que les corresponden. Sin embargo, este no es necesariamente el caso, y una de las características más relevantes del perceptrón, es que es posible definir una **regla de aprendizaje** que permite obtener los valores óptimos de las componentes del vector \mathbf{w} que permiten al perceptrón clasificar de manera correcta a los elementos en sus categorías correspondientes.

Regla de aprendizaje del perceptrón

La regla de aprendizaje de Rosenblatt, es un algoritmo (i.e. una secuencia de operaciones) que permite obtener los valores óptimos para los pesos sinápticos del perceptrón, para que este clasifique de manera correcta los elementos de un conjunto. Cabe hacer notar que, para aplicar la regla de aprendizaje, es necesario conocer de antemano las etiquetas correctas de cada elemento, las cuales denotamos por el conjunto $Y = \{y_1, \dots, y_N\}$. Al conjunto de datos (o elementos), junto con sus etiquetas correspondientes, se denomina como **conjunto de entrenamiento** y, junto con la regla de aprendizaje, constituyen la base de lo que se conoce como **entrenamiento del modelo**, que discutimos en las siguientes secciones.

La regla de aprendizaje del perceptrón, se puede resumir en los siguientes pasos:

1. Inicializar los valores de las componentes del vector \mathbf{w} en 0 o en valores aleatorios pequeños.
2. Para cada elemento del conjunto X :
 - a) Calcular el valor de salida \hat{y}
 - b) Realizar una comparación del valor de salida calculado por el perceptrón \hat{y} , con el valor real de la etiqueta de cada elemento y , y ajustar los valores de los pesos sinápticos del perceptrón (los valores de las componentes del vector \mathbf{w}).

En el ejemplo de la tarea de la clasificación binaria discutida anteriormente, el valor de salida \hat{y} , es el resultado de la función de activación, $\phi(z)$, para la entrada neta asociada a cada elemento del conjunto X , la cual puede tomar los valores, 1 o -1 .

La actualización simultánea de cada peso w_j en el vector \mathbf{w} se puede escribir como:

$$w_j := w_j + \Delta w_j, \quad (1.4)$$

donde Δw_j está dado por la expresión:

$$\Delta w_j = \eta(y - \hat{y})x_j, \quad (1.5)$$

donde $\eta \in [0, 1]$ es la tasa de aprendizaje, y es la etiqueta de real (o presente en el conjunto de datos de entrenamiento), \hat{y} es la etiqueta asignada por el modelo, y x_j es la j -ésima entrada del vector \mathbf{x} .

La aplicación iterativa de esta regla para todos los elementos en el conjunto de entrenamiento, permite obtener los valores óptimos de los pesos sinápticos del perceptrón para realizar una correcta clasificación de los elementos. Sin embargo, es importante mencionar que la convergencia de este procedimiento, solamente está garantizada para conjuntos de datos linealmente separables y para valores bajos de la tasa de aprendizaje η . En el caso de conjuntos de datos que no son linealmente separables, es posible establecer criterios de paro, para obtener un modelo de clasificación óptimo.

Así mismo, la clasificación binaria de clases linealmente separables, es un caso extremadamente sencillo que no se observa usualmente en datos reales y que solamente hemos utilizado para ejemplificar las características de este modelo. En general, es necesario utilizar redes neuronales artificiales más sofisticadas, con un mayor número de neuronas artificiales y capas de procesamiento, para poder llevar a cabo tareas de clasificación más complejas, como el que nos ocupa en este trabajo.

Ya que hemos descrito los componentes y características esenciales de las redes neuronales, en las siguientes secciones describimos las etapas y la descripción general de la implementación en un modelo de clasificación, basado en RNA.

→1.4. Etapas de un modelo de clasificación basado en RNA

La implementación de un modelo de clasificación consta de 4 etapas: el preprocesamiento de datos, el entrenamiento y evaluación del modelo y la predicción. Estas etapas están relacionadas entre ellas (véase figura 1.2), y las describimos a continuación.

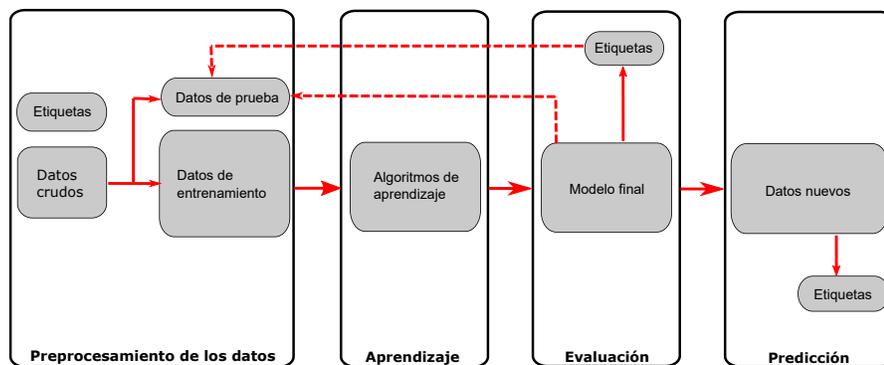


Figura 1.2: Las etapas de la implementación de un modelo de clasificación son: el preprocesamiento, el aprendizaje, la evaluación y la predicción. Imagen adaptada de [15].

Preprocesamiento

Dado que las fuentes de datos son en general muy variadas y comúnmente no se encuentran en el formato adecuado para que puedan ser procesados por una RNA, es necesario realizar un preprocesamiento de los datos con los que contamos para que puedan ser utilizados de manera óptima por nuestro modelo.

El preprocesamiento de datos puede referirse a distintos tipos de procedimientos como: filtrar las características más representativas de los datos de entrada, procedimiento

también conocido como reducción de dimensionalidad, que esencialmente consiste en eliminar las componentes menos relevantes (con menor varianza en el conjunto de datos, por ejemplo) o redundantes, o normalizar las escalas de los valores de entrada. Estos procedimientos permiten un mejor manejo de los datos por el modelo, reducir el tiempo de entrenamiento y obtener mejores resultados.

Otro aspecto importante del preprocesamiento de datos, es el de generar un conjunto de datos de prueba a partir del conjunto de datos de entrenamiento. Este conjunto de datos nos permitirá evaluar el entrenamiento (o aprendizaje) del modelo, para determinar si se ha obtenido un modelo óptimo, que sea capaz de identificar las características de los elementos de cada categoría y clasificar correctamente a nuevos datos que se le presenten. En general, el conjunto de datos de evaluación corresponde a una muestra aleatoria de los datos de entrenamiento.

Finalmente, como parte de esta etapa de la implementación de un modelo de clasificación, es necesario definir la forma de los datos de salida que deseamos obtener de la RNA. Por ejemplo, las etiquetas numéricas que se asignarán a cada una de las categorías presentes en los datos (de manera análoga a las etiquetas numéricas 1 y -1 para las categorías positiva y negativa de la sección 1.3), y que nos permitirán interpretar los resultados obtenidos del modelo. Una vez que hemos definido el conjunto de datos de entrenamiento, de prueba y las etiquetas de salida, podemos proceder al entrenamiento del modelo, el cual describimos a continuación.

Entrenamiento del modelo (Aprendizaje)

Esta etapa de la implementación del modelo de clasificación, corresponde a la generalización de la regla de aprendizaje del perceptrón (véase la sección 1.3), en el sentido en que en esta etapa se realiza el ajuste de los pesos sinápticos de todas las neuronas artificiales (también denominados nodos) que forman parte de la RNA. De la misma manera que en el caso del perceptrón, el procedimiento general es el de comparar el valor de salida asignado por el modelo, con el valor real de cada elemento en el conjunto de datos de entrenamiento, para posteriormente ajustar los valores de los pesos sinápticos de los nodos en la red, lo que resulta en una estimación del **error de clasificación** de la RNA, el cual será utilizado para su evaluación para determinar el fin la etapa de entrenamiento del modelo.

El objetivo de la etapa de entrenamiento es la de obtener un modelo óptimo, que permita clasificar adecuadamente los datos del conjunto de entrenamiento, evitando que este solamente sea capaz de identificar a este conjunto de datos (lo que se conoce como sobreentrenamiento del modelo). Una manera de identificar esta situación, es un crecimiento del error de clasificación al presentar al modelo los datos de prueba, lo que indica un sobreajuste del modelo a los datos de entrenamiento.

El sobreentrenamiento del modelo, provocar que este no sea capaz de clasificar datos nuevos correctamente. Haciendo una analogía con las características del cerebro, es necesario que la RNA tenga cierto grado de **plasticidad**, que le permita procesar datos nuevos que comparten características, pero que no son idénticos, con los datos del conjunto de entrenamiento.

La siguiente etapa en la implementación del modelo de clasificación, es la de evaluación, la cual, nos permite identificar que nuestro modelo funciona adecuadamente.

Evaluación

Una vez que hemos entrenado el modelo, utilizamos el conjunto de datos de prueba para evaluar el desempeño del modelo y estimar el error de generalización. Una forma eficiente de evaluar al modelo, es por medio del cálculo de su **matriz de confusión**.

Las entradas de la matriz de confusión, permiten comparar los resultados obtenidos con el modelo, que corresponden a las filas de la matriz, con los valores reales, presentes en el conjunto de entrenamiento, que corresponden a las columnas. La diagonal principal de la matriz contiene los verdaderos positivos (**VP**) y los verdaderos negativos (**VN**), que corresponden a los datos clasificados correctamente por el modelo para ambas clases, mientras que las entradas restantes corresponden los falsos positivos (**FP**) y los falsos negativos (**FN**), los datos clasificados de manera errónea (véase la figura 1.3).

		Valores reales	
		P	N
Valor predicho	P	VP	FP
	N	FN	VN

Figura 1.3: Matriz de confusión. Los valores de esta matriz, obtenidos de la aplicación del modelo al conjunto de prueba, permite realizar una evaluación del entrenamiento del modelo.

Una vez que, de acuerdo con las entradas de la matriz de confusión, nuestro modelo ha sido entrenado correctamente, podemos presentar al modelo datos nuevos para obtener predicciones de sus categorías correspondientes.

Predicción

Ya que hemos determinado que el entrenamiento del modelo es óptimo mediante su evaluación, podemos proceder a aplicarlo a datos nuevos para que sean clasificados. Es importante mencionar que puede que sea necesario someterlos a un preprocesamiento, para que tengan el mismo formato y características de los datos en el conjunto de entrenamiento, y puedan ser procesados correctamente por el modelo.

Resultado de las distintas etapas que hemos descrito, obtenemos las etiquetas asignadas por el modelo entrenado para cada uno de los datos nuevos que le fueron presentados, que en caso del modelo que implementamos en este trabajo, corresponde a determinar el sentimiento (positivo o negativo) de los textos de publicaciones en la plataforma Twitter, para finalmente proceder a culminar el proceso de análisis de datos, con la interpretación de los resultados obtenidos.

En el siguiente capítulo, describimos los detalles particulares, de los modelos de análisis de sentimiento que fueron implementados.

▶ 2. Modelos de evaluación implementados

Como hemos mencionado anteriormente, el objetivo principal de la implementación de estos modelos, es el de contar con herramientas para estimar la carga emotiva de los textos de publicaciones en redes socio-digitales, como *posts* de Facebook, o *tweets* en la plataforma Twitter, lo cual, en el contexto de su aplicación a las publicaciones realizadas alrededor de eventos focales, i.e. momentos de alta intensidad en la actividad de las plataformas resultado de eventos relevantes en la capa física [10], como por ejemplo, movilizaciones sociales y manifestaciones, jornadas electorales, o acontecimientos políticos y sociales relevantes, nos permite identificar tanto el tipo de conversación que se desarrolla, como la intención del lenguaje utilizado alrededor de eventos y personajes relevantes.

En las siguientes secciones, describimos los modelos de análisis de sentimiento implementados, uno basado en léxico y otro que hace uso de una red neuronal artificial.

▶ 2.1. Modelo de Red Neuronal Artificial

La estrategia utilizada por los modelos de clasificación basados en Redes Neuronales Artificiales, se puede resumir a grandes rasgos como una manera de extrapolar las características de un conjunto de datos anotados, previamente clasificados en distintas categorías (el conjunto de datos de entrenamiento), a nuevos datos que pueden compartir algunas de las características que los identifican.

La implementación de uno de estos modelos se compone de distintas etapas: preprocesamiento de los datos y selección de un conjunto de datos de entrenamiento, entrenamiento y evaluación. Las cuales describimos en las siguientes secciones.

Conjunto de datos de entrenamiento (TASS)

El conjunto de datos de entrenamiento es una parte fundamental de los modelos de clasificación de aprendizaje de máquina, ya que representan las características que definen a cada una de las categorías presentes en los datos, y que precisamente son las que se desea extrapolar a nuevas observaciones para que el modelo pueda clasificarlos en alguna de ellas. Generalmente, el conjunto de datos de entrenamiento se compone de un conjunto de observaciones o elementos representativos de categorías relevantes, los cuales son clasificados o calificados por personas con conocimiento previo o experiencia para cada caso en particular.

Para el tema que tratamos en este trabajo, el análisis de sentimiento de publicaciones en redes socio-digitales alrededor de eventos focales, el corpus TASS [20], desarrollado por el Taller de Análisis Semántico de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN), es especialmente adecuado como conjunto de datos de entrenamiento, ya

que se compone de textos de publicaciones de la plataforma Twitter cuyo sentimiento ha sido previamente evaluado.

Más aún, dado que uno de los objetivos de este trabajo, es el de implementar modelos para analizar el sentimiento de la conversación alrededor de eventos políticos y sociales, principalmente en México y latinoamérica, utilizar el corpus TASS resulta adecuado ya que se compone de textos de tweets en el idioma español sobre temas en un contexto político, lo cual es relevante para la labor de investigación que desarrollamos en el Tlatelolco-LAB.

Hemos decidido utilizar el corpus general TASS de 2012 para el entrenamiento de nuestro modelo de clasificación, ya que es el corpus con un mayor número de datos, el cual cuenta con 7219 textos evaluados y con más de 60 mil textos para evaluación. Todos estos textos fueron generados por cuentas de usuarios de Twitter españoles, sobre temas políticos de su país.

Cada uno de los elementos (textos) del corpus TASS de 2012, cuenta con 6 atributos:

1. **tweetid:** El id o número de identificación de la publicación asignada por la plataforma Twitter.
2. **user:** El nombre del usuario que realizó la publicación.
3. **content:** El texto de la publicación.
4. **date:** La fecha de creación de la publicación.
5. **lang:** El lenguaje del texto de la publicación.
6. **sentiments:** Un diccionario con las evaluaciones de sentimiento del texto de la publicación, con la siguiente información anidada:
 - **polarity:**
 - **value:** El valor que representa cada una de las siguientes categorías:
 - **P+:** Sentimiento muy positivo
 - **P:** Sentimiento positivo
 - **N+:** Sentimiento muy negativo
 - **N:** Sentimiento negativo
 - **NEU:** Sentimiento neutral
 - **NONE:** El sentimiento del tweet se desconoce
 - **type:** El tipo de sentimiento: **AGREEMENT** (de acuerdo) o **DISAGREEMENT** (desacuerdo).
 - 7. **topics:** El tema del que habla el tweet, como política, deporte, entretenimiento, entre otros.

Una vez que hemos descrito a grandes rasgos las características de los elementos en el corpus que utilizaremos, procedemos a describir el procesamiento de los datos para el entrenamiento de nuestro modelo

Procesamiento de los datos

El primer tratamiento que le dimos a los datos fue el de reducir la dimensionalidad de los datos en el corpus de entrenamiento, al solamente considerar el texto y las evaluaciones de sentimiento de cada elemento.

Así mismo, aunque la evaluación de los elementos en el corpus consta de 6 categorías, solamente consideramos 4 de ellos, en específico: el sentimiento positivo (P y P+), el negativo (N y N+), el neutral (NEU) y la falta de una calificación o el desconocimiento del sentimiento (NONE). A cada una de estas categorías les asignamos los valores numéricos:

- **Positivo:** 1
- **Negativo:** -1
- **Neutral:** 0
- **Indeterminado:** 99

Estos valores numéricos se normalizaron y representaron por medio de un vector, resultando en la matriz de clase binaria mostrada en la Tabla ??.

Categoría	Representación	Vector
Positivo	1	[1,0,0,0]
Negativo	-1	[0,1,0,0]
Neutral	0	[0,0,1,0]
Indeterminado	99	[0,0,0,1]

Tabla 2.1: Matriz de clase binaria. Representamos cada categoría por medio de un valor numérico, y de igual forma a cada categoría le asignamos un vector que corresponde a la la representación binaria de la salida, para facilitar el procesamiento de los datos.

El segundo procesamiento de datos fue a los textos a evaluar, los cuales, en primer lugar fueron vectorizados, i.e. se generó una matriz de tokens a partir de las palabras contenidas en el texto, las cuales fueron utilizadas para generar un vocabulario.

Posteriormente, cada texto se transforma en una secuencia de tokens que coincidan con los del vocabulario, añadiendo ceros a la derecha de cada secuencia, de tal forma que obtengamos una matriz de características donde cada muestra tiene la misma longitud. Decidimos utilizar secuencias 33 tokens, debido a que esta es la longitud de la secuencia más larga presente en el corpus.

Descripción del modelo

El modelo de clasificación que implementamos,¹ consta de una red neuronal compuesta de las capas enumeradas a continuación:

1. La primera capa corresponde a los parámetros de entrada de la red, que incluye el tamaño del lote, el tamaño del vector de características y la dimensión del vocabulario.

¹Se utilizo el lenguaje de programación Python y la paquetería Keras para la implementación del modelo, así como el modelo secuencial.

2. Una capa convolucional cuyo tamaño de características es el mismo que el tamaño de lote, el tamaño de kernel que especifica la longitud de la ventana de convolución 1D es 5 y con una función de activación ReLU.
3. Una capa GlobalMaxPooling1D, la cual es una capa de agrupación que reduce la dimensionalidad, y que permite que la siguiente capa puede trabajar con los datos obtenidos en la segunda capa.
4. Una última capa densa de nodos con función de activación sigmoide, con tantos nodos como el tamaño de la salida, es decir las categorías que se buscan predecir.

De manera general, podemos representar el proceso de clasificación de los datos (tanto de los de entrenamiento, evaluación o predicción) del modelo M , con la expresión:

$$M(o_i) = (p_i^0, \dots, p_i^M), \quad (2.1)$$

donde o_i representa el i -ésimo elemento evaluado por el modelo, y p_i^j representa la probabilidad de pertenencia a la j -ésima categoría, asignada por la RNA. Cabe hacer notar que, a diferencia del modelo basado en léxico, en donde la intensidad del sentimiento también es tomada en cuenta, las entradas del vector $M(o_i)$, representan la probabilidad de pertenencia de la observación o_i a cada una de las distintas categorías o clases consideraras, se cumple que

$$\sum_{j=0}^M p_i^j = 1, \quad (2.2)$$

para cada una de los elementos evaluados.

Entrenamiento

Con los datos preprocesados entrenamos la RNA durante 50 épocas, lo que significa que el corpus de entrenamiento completo fue presentado al modelo 50 veces. La presentación de los datos para el entrenamiento del modelo se realizó en lotes de: 10, 20, 32, 50, 64, 70, 80, 90, 100, 128, 150, 200 y 256 secuencias de 33 tokens, con la finalidad de determinar el modelo con el mejor rendimiento.

Con la intención de identificar el modelo y el tamaño de bloque óptimo, realizamos un registro de los parámetros que maximizan la precisión al evaluar el conjunto de datos de prueba, la cual está dada en términos de las entradas de la matriz de confusión, por la expresión:

$$\text{Precisión} = \frac{VP}{VP + FP}, \quad (2.3)$$

y que simultáneamente minimizan el valor de pérdida, es decir la suma de los errores de clasificación cometidos. Este proceso de optimización de parámetros se realiza con la finalidad de evitar que el modelo se sobre ajuste a los datos del conjunto de entrenamiento,

lo cual no permitiría realizar un procesamiento adecuado de nuevos datos por el modelo. Los valores de precisión y pérdida para el mayor y menor tamaño de lote considerados, pueden observarse en la figura 2.1.

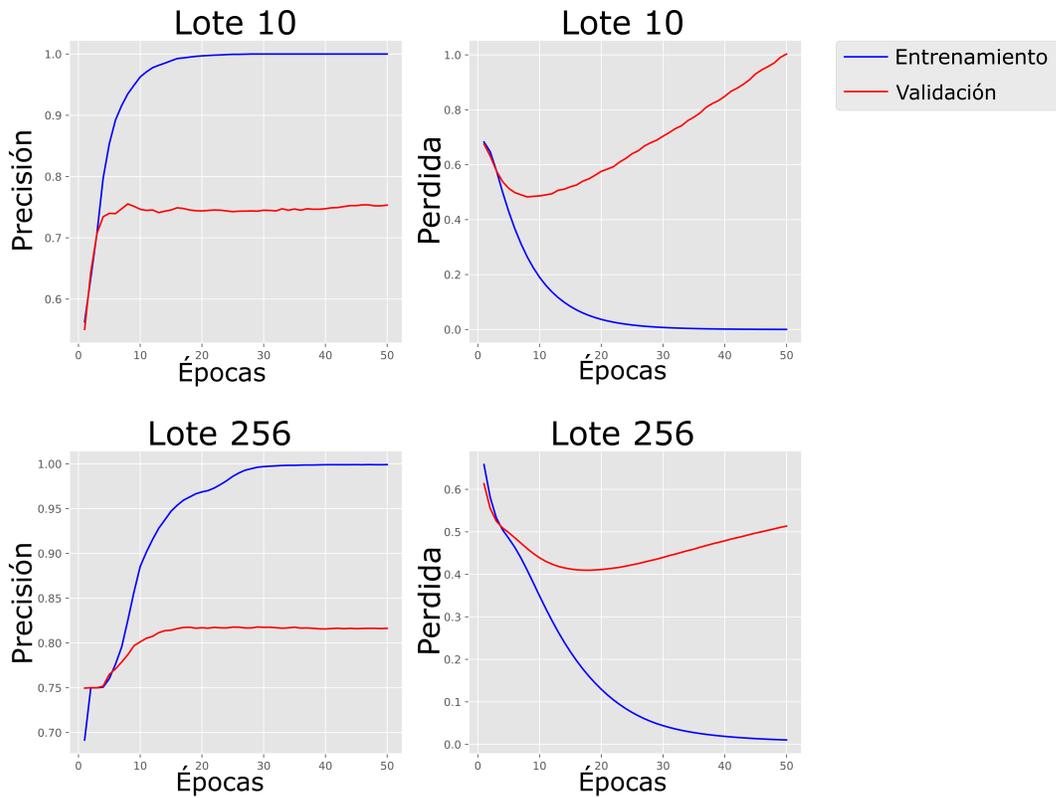


Figura 2.1: Comparación de los valores de precisión y pérdida para el menor tamaño de lote, en este caso 10 y para el mayor que es 256, dicho análisis se aplica al entrenamiento (representado por la línea azul) y la validación de dicho entrenamiento (representado por la línea roja) para 50 épocas

Evaluación

Evaluamos el desempeño del modelo en la clasificación de los datos de prueba mediante su matriz de confusión, la cual se muestra en la figura 2.2. Se puede observar que el modelo muestra un rendimiento variable para cada categoría de sentimiento, lo cual también se ve reflejado en la precisión de clasificación para cada categoría.

		Valores reales			
		P	N	Neu	Non
Valor predicho	P	17300	3273	10	1646
	N	1782	13500	8	554
	Neu	414	868	12	11
	Non	5974	7782	25	7635

Figura 2.2: Matriz de confusión resultante de la evaluación del modelo con mayor precisión para las categorías positivo(P), negativo(N), neutral(Neu) y None(Non), las columnas representan los valores reales, las filas los valores predichos, por lo que la diagonal principal representa los valores que el modelo predice de manera correcta

Como se reporta en la tabla 2.2, el modelo tiene una precisión del 77 % en la clasificación de textos negativos y del 85 % para la clasificación de textos positivos, mientras que para el caso de la categoría neutral tiene una precisión menor al 1 %, así mismo la precisión para la clasificación de los datos en la categoría NONE, es de 35 %.

Clase	Precisión
Positiva	0.778303
Negativa	0.852058
Neutral	0.0091954
None	0.356509

Tabla 2.2: Precisión de los valores obtenidos del modelo para cada una de las clases.

La principal razón de esta diferencia de precisión en la clasificación de los elementos de cada categoría, es el desbalance en el número de elementos. Como se muestra en la tabla 2.3, la categoría NONE está fuertemente subrepresentada, en comparación con el resto de las categorías, lo que provoca que no sea posible entrenar un modelo que clasifique de manera correcta a los elementos de esta categoría.

Clase	Tweets
Positiva	2884
Negativa	2182
Neutral	670
None	1483

Tabla 2.3: Numero de elementos para cada una de las categorías en el conjunto de datos de entrenamiento

Debido a esta observación, decidimos omitir a los textos etiquetados con NEU y NONE en el corpus original, y reentrenar el modelo con este conjunto de datos de entrenamiento filtrado. Con esta modificación del modelo, siguiendo la notación de la ecuación 2.1, podemos expresar el resultado de la clasificación de los elementos por el modelo como:

$$M(o_i) = (p_i^+, p_i^-), \tag{2.4}$$

donde, p_i^+ representa la probabilidad de pertenencia del elemento (texto) o_i a la clase positiva, y p_i^- representa la probabilidad de pertenencia a la clase negativa. La matriz de confusión de esta modificación el modelo se muestra en la figura 2.3, y los valores de precisión se reportan en la tabla 2.4, obteniendo una mejoría de alrededor del 20 % con respecto al modelo anterior.

		Valores reales	
		P	N
Valor predicho	P	19183	2331
	N	3050	13513

Figura 2.3: Matriz de confusión resultante de la evaluación del modelo con mayor precisión para las categorías positivo(P) y negativo(N), las columnas representan los valores reales, las filas los valores predichos, por lo que la diagonal principal representa los valores que el modelo predice de manera correcta

Clase	Precisión
Positiva	0.862817
Negativa	0.852878

Tabla 2.4: Precisión de los valores obtenidos del modelo para cada una de las clases.

➔2.2. Modelo basado en léxico

Cómo se explicó en la Sec. 1.2, la estrategia de traducir los textos al inglés para poder hacer uso de VADER no es muy eficiente dadas las limitaciones de los traductores automáticos. Sin embargo, dado que VADER es una herramienta de software libre [9], fue posible traducir todas las palabras del léxico al español.

Esta estrategia permite usar a VADER en idioma español, con la gran salvedad que todos los tokens siguen teniendo la misma evaluación dada por las personas que participaron en la construcción del léxico (todas ellas estadounidenses), y por lo tanto los valores de la intensidad y el sentimiento que se obtienen al evaluar textos, no necesariamente reflejan las sutilezas propias de nuestra lengua.

Dicho esto, esta estrategia permite hacer una evaluación rápida de los textos, y hasta dónde hemos podido observar, es bastante eficiente para identificar el sentimiento general de los tweets que estamos trabajando, aunque no su intensidad.

VADER permite evaluar el sentimiento de una oración en la escala de -1 a +1, siendo -1 el extremo más negativo y +1 el extremo más positivo. El valor de 0 ocurre cuando ninguna

de las palabras en la oración se encuentra en el léxico (o las palabras se balancean justo para dar una puntuación compuesta de cero; aunque este caso es extremadamente raro). En este caso decimos que los textos son inclasificables. Para fines estadísticos, todos aquellos textos cuyo valor sea 0, serán excluidos de los análisis que se realicen.

Ahora bien, establecer si un conjunto de documentos tiene una valoración positiva, negativa o neutra, puede ser un poco complicado. Dicho valor dependerá de la confianza que se tenga en los valores del léxico (que por el momento es poca) y del conjunto de textos que se esté evaluando.

En esta metodología proponemos una manera sistemática para determinar cuál es sentimiento general de un corpus, esto es, determinar qué porcentaje de los documentos son positivos, negativos o neutros; calculando la distribución de probabilidad acumulativa de los valores de sentimiento obtenidos con VADER.

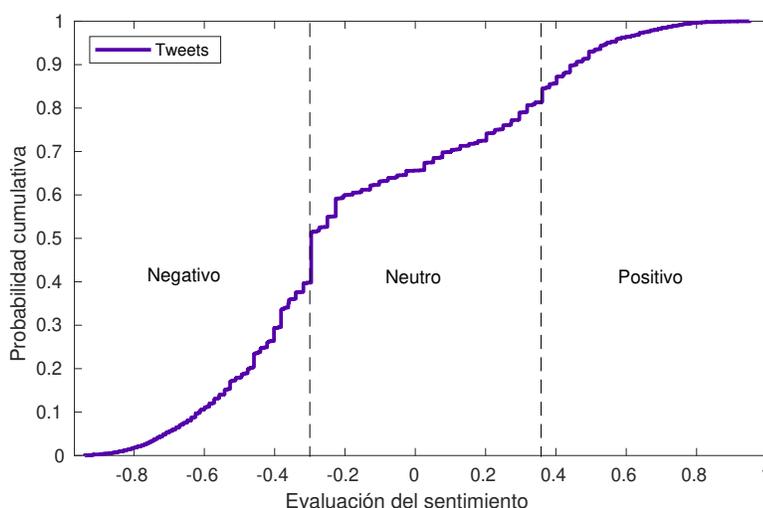


Figura 2.4: Distribución acumulativa para los valores del sentimiento obtenidas por medio de VADER, para un conjunto de Tweets. Para este conjunto de textos, la distribución acumulativa permite distinguir tres distintas secciones: negativa (valores menores que -0.29), neutra (valores mayores que -0.29 y menores que 0.36) y positiva (valores mayores que 0.36).

La figura 2.4 es la distribución acumulativa para los valores del sentimiento obtenidas por medio de VADER, utilizando el léxico original traducido al español, para un conjunto de Tweets. La propuesta es la de usar los cambios en la distribución acumulativa para determinar las regiones para los cuales podemos considerar a los textos como positivos, negativos y neutros. En este caso vemos que textos cuyos valores del sentimiento sean menores que -0.29 pueden ser considerados como negativos. De igual manera, textos cuyos valores del sentimiento sean mayores que 0.36 se pueden considerar como positivos; y para valores entre -0.29 y 0.36 , los textos se pueden clasificar como neutros.

Es importante recalcar que estas distinciones dependerán esencialmente del conjunto de textos estudiados y podrá evaluarse su pertinencia una vez se cuente con un léxico anotado para el español.

▶ 3. Aplicación de la metodología

En este capítulo presentamos la aplicación de las dos metodologías descritas en el capítulo anterior (Red neuronal artificial y VADER) para un conjunto de tweets alrededor de las manifestaciones llevadas a cabo por la Coordinadora Nacional de Trabajadores de la Educación (CNTE) durante el periodo de agosto a octubre de 2013 en la Ciudad de México y en varios estados del país, con el fin de protestar en contra de la Reforma Educativa durante el gobierno de Enrique Peña Nieto.

Las manifestaciones se concentraron principalmente en el Zócalo de la Ciudad de México, dando lugar a una serie de enfrentamientos entre distintos sectores de la sociedad. Dichos enfrentamientos tuvieron eco en las redes socio-digitales, en especial el Twitter, en donde se llevó a cabo una discusión muy interesante en términos sociales, que ha sido estudiada ampliamente por el Programa Universitario de Estudios Sobre Democracia, Justicia y Sociedad, de la UNAM (PUEDJS-UNAM).

Durante el periodo comprendido entre agosto y octubre de 2013, se recopilieron 293,237 Tweets, que incluían el hashtag #CNTE (entre otros). Estos tweets fueron procesados para determinar las redes de retweets entre cuentas y posteriormente analizarlas a través del algoritmo de Girvan-Newman [7] para determinar a las comunidades involucradas.

De éstas comunidades se conservaron para el presente estudio a las 4 más grandes en terminos del número de cuentas involucradas, que denominamos como comunidades 0, 1, 2, 3. Así mismo, la Tabla ?? muestra el número de publicaciones por tipo en la recopilación de los datos, cabe hacer notar que el número de **Quotes** (o citas) recuperados, a diferencia del resto de tipos de publicaciones, no es suficiente para realizar una evaluación estadística de los datos, por lo cual, este tipo de publicación será omitida en el análisis.

Comunidad	Tweet	Retweet	Reply	Quote
0	27,459	1,4508	6,661	22
1	45,533	2,7147	5,456	42
2	8,527	6,995	1,562	4
3	10,666	10,711	1,477	3

Tabla 3.1: Tipos de publicación para cada una de las comunidades.

A continuación se muestran los resultados del análisis del sentimiento obtenidos a través de Machine Learning y VADER para cada una de las comunidades identificadas en las redes de retweets entre cuentas, presentes en la recopilación de datos.

→3.1. Evaluación basada en léxico VADER

Presentamos ahora los resultados para el análisis del sentimiento de las cuatro comunidades usando el léxico traducido de VADER. Como se mencionó anteriormente, para este análisis se removieron todos los textos cuya evaluación del sentimiento dio un resultado de cero. Además, no se consideró que las publicaciones puedan clasificarse como neutras (solo se hizo la distinción entre positivos y negativos), esto debido a que los valores de la intensidad no son del todo confiables dado que la traducción del léxico no puede reflejar de manera correcta la intensidad del sentimiento en español.

Como se dijo en la sección 1.2, al presentar el i -ésimo elemento del conjunto de documentos al modelo, se obtiene una evaluación $e_i \in [-1, 1]$. El conjunto de todas las evaluaciones, permite hacer análisis estadísticos que ayudan a determinar las propiedades generales de los conjuntos de textos asociados a cada comunidad; en otras palabras para cada comunidad definimos al conjunto de todas las evaluaciones con sentimiento positivo como: $x_i^+ = \{e_i | e_i > 0\}$ y de manera análoga las evaluaciones con sentimiento negativo como $x_i^- = \{e_i | e_i < 0\}$. Un vistazo rápido a estos valores de x_i^\pm , da una idea del sentimiento general para cada comunidad.

En la figura 3.1 se muestra la proporción del sentimiento de las publicaciones para las cuatro comunidades estudiadas. Al comparar esta figura, con la figura 3.4, es posible ver que los dos métodos identifican de manera similar el sentimiento de las publicaciones, asignando una evaluación negativa a un mayor número de publicaciones.

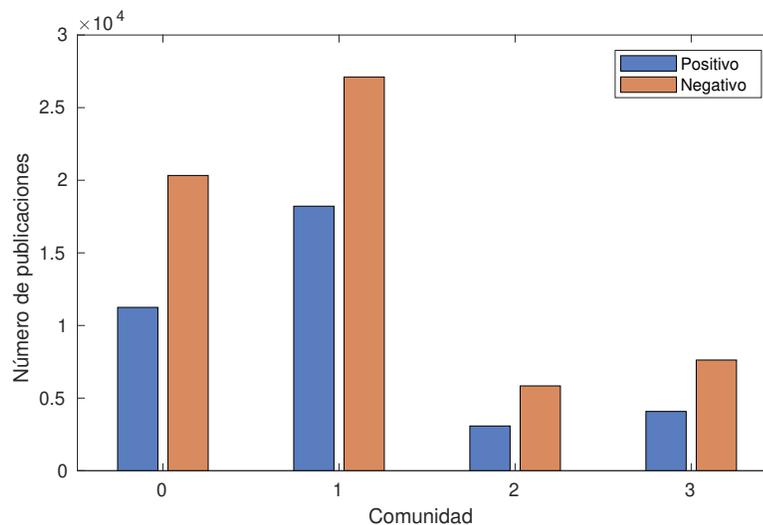


Figura 3.1: Proporción de publicaciones positivas y negativas identificadas por el léxico VADER, para los textos de los tweets de las cuatro comunidades estudiadas.

Se puede hacer un análisis más detallado al separar los datos por tipo de publicación. Por ejemplo, si consideramos solamente los tweets, podemos calcular la distribución de probabilidad de las evaluaciones de sentimiento (positivo y negativo) de los textos obtenidos con VADER, las cuales se muestran en la figura 3.2 para cada una de las comunidades

estudiadas.

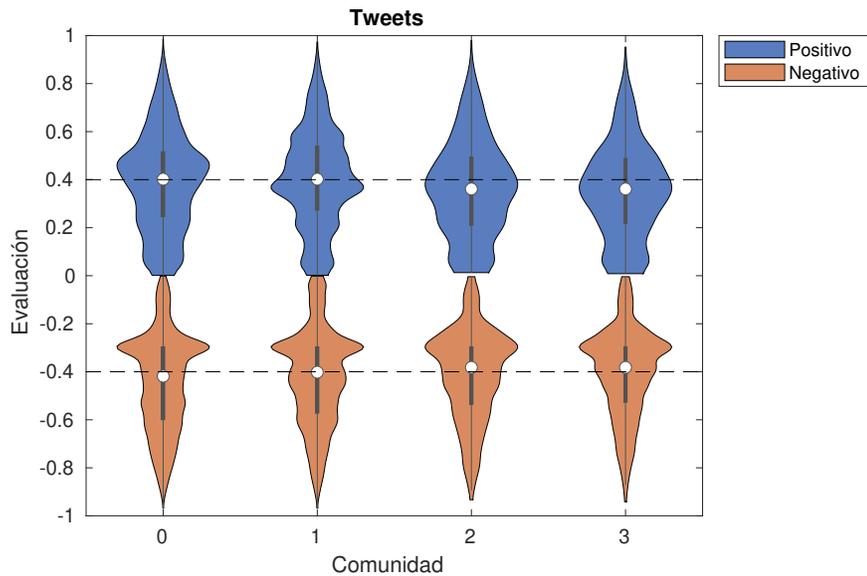


Figura 3.2: Distribuciones de probabilidad de las evaluaciones de los tweets para cada una de las comunidades. La evaluación positiva del sentimiento (Azul) no muestra grandes diferencias entre las comunidades. De igual manera, la evaluación negativa del sentimiento (Naranja) tampoco evidencia diferencias entre los sentimientos de las distintas comunidades. Las líneas punteadas indican valores alrededor de la media (puntos blancos) para cada una de las comunidades; esto es, 0.4 y -0.4 para evaluaciones positivas y negativas respectivamente. La caja gris indica el rango intercuartil.

Varias cuestiones interesantes pueden notarse en figura 3.2. La gráfica de violín permite identificar cuál es el valor de la media para los datos. En este caso se observa que para las evaluaciones positivas, el valor del sentimiento estimado por VADER se encuentra alrededor de 0.4 para todas las comunidades, indicando así una intensidad moderada.

De manera análoga, VADER hace una estimación del sentimiento negativo de aproximadamente -0.4 para las cuatro comunidades; indicando también una intensidad moderada. Estos valores de las medias, permiten establecer un valor para la polarización del sentimiento en el conjunto de tweets asociado a cada comunidad, el cual denotamos por P_i , simplemente calculando el valor absoluto de la diferencia de las medias, i.e.

$$P_i = \|\tilde{x}_i^+ - |\tilde{x}_i^-|\|, \quad (3.1)$$

donde, \tilde{x}_i^+ , \tilde{x}_i^- , son las medias de las distribuciones de las evaluaciones positivas y negativas respectivamente, de la i -ésima comunidad. Para el caso aquí presentado, se tiene que $P_0 = 0.0165$, $P_1 = 0$, $P_2 = 0.0206$ y $P_3 = 0.0206$; para cada una de las comunidades. Valores cercanos a cero indicarían poca diferencia entre la evaluación del sentimiento positivo y negativo de los textos, y por lo tanto, poca polarización en las comunidades estudiadas. El caso opuesto, cuando el valor de la polarización P_i es cercano a uno, sería un indicador de una polarización extrema en los sentimientos evaluados en las comunidades.

Cabe mencionar que, la polarización del sentimiento de las publicaciones definida en la 3.1, puede calcularse para todos los tipos de publicación (Retweets, Quotes y Replies).

La figura 3.3 muestra las distribuciones de probabilidad de la evaluación de sentimiento asignada por el modelo, para los Retweets y Replies de las cuatro comunidades. No se aprecian diferencias notables en estas figuras. Esto puede indicar que no se está detectando correctamente la intensidad en los sentimientos de las publicaciones; y para hacerlo es imperativo desarrollar un léxico para el español.

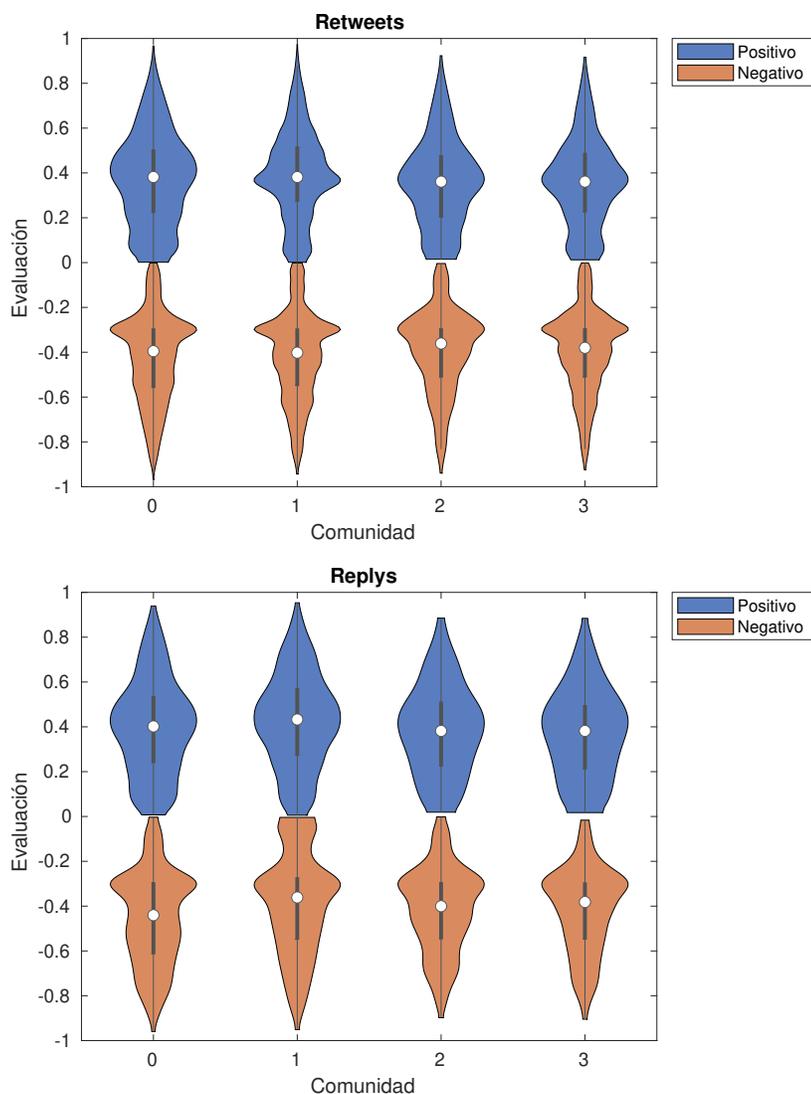


Figura 3.3: Distribuciones de probabilidad de las evaluaciones de sentimiento para los Retweets (arriba) y los Replies (abajo). Los valores de la media se encuentran nuevamente al rededor de 0.4 y -0.4 para las evaluaciones positivas y negativas.

→3.2. Evaluación con Redes Neuronales Artificiales

En esta sección describimos los resultados obtenidos, de la evaluación de sentimiento de los textos de las publicaciones de cada una de las comunidades realizada mediante el modelo de RNA.

Como se define en la ecuación 2.4, el resultado de la evaluación del modelo de los textos de cada comunidad es el vector $M(o_i) = (p_i^+, p_i^-)$, donde cada entrada corresponde a la probabilidad de pertenencia del elemento o_i (en este caso el texto de una publicación en la plataforma Twitter), a la clase positiva o negativa.

Ya que obtenemos un par de valores de la RNA para cada elemento, la manera más sencilla de asignar la categorías de la observación o_i , la cual denotamos por $c(o_i)$, es por medio de la probabilidad con el valor más alto, i.e. el texto de un tweet se considera como negativo cuando $p_i^- > p_i^+$ y positivo cuando $p_i^+ > p_i^-$. Este criterio puede expresarse haciendo uso de la diferencia entre las probabilidades de pertenencia, $\delta_i^\pm = p_i^+ - p_i^-$, de la siguiente manera:

$$c(o_i) = \begin{cases} \text{Positivo,} & \text{si } \delta_i^\pm > 0 \\ \text{Negativo,} & \text{si } \delta_i^\pm < 0 \end{cases} \quad (3.2)$$

En la figura 3.4 mostramos el número de publicaciones positivas o negativas identificadas para cada comunidad.

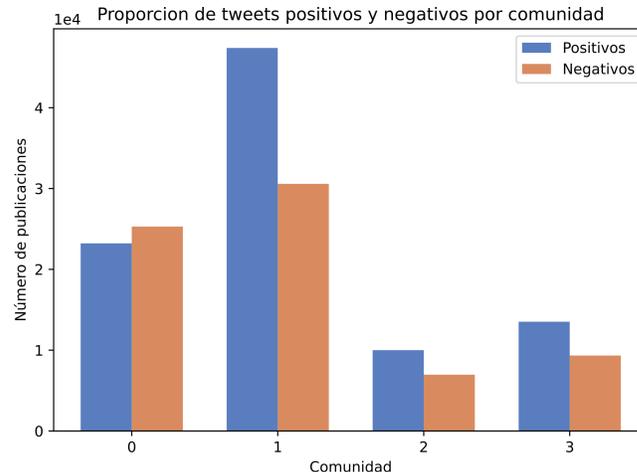


Figura 3.4: Número de publicaciones identificadas como positivas o negativas, de acuerdo con el criterio de la ecuación 3.2, para cada una de las comunidades en la recolección de datos.

Como es evidente de la definición del criterio de asignación de categoría, aunque útil, presenta dificultades cuando $\delta_i^\pm \approx 0$, i.e. en los casos en los que el modelo asigna valores muy similares para las probabilidades de pertenencia. En la figura 3.5 mostramos la distribución de probabilidad de los valores de δ_i^\pm , de las evaluaciones de los textos de todas

las comunidades. Como se puede observar, la distribución de δ_i^\pm es esencialmente uniforme, hasta valores ≈ 0.7 , en donde se observa un decaimiento.

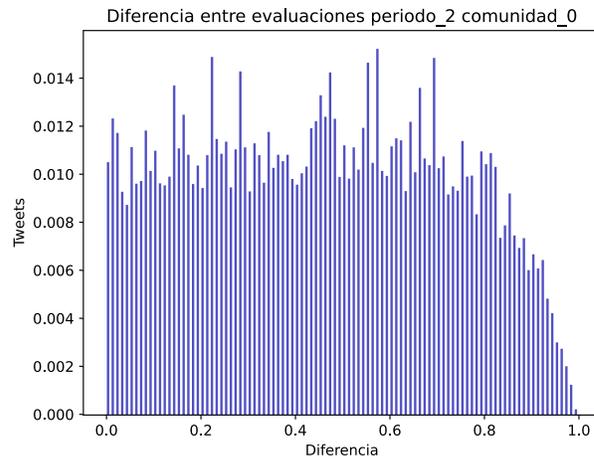


Figura 3.5: Diferencia entre los valores positivos y negativos del vector predicho para la clase 0 de las comunidades que hablaron de la CNTE en twitter entre agosto y septiembre del 2013.

Esta observación, nos permite definir el umbral para las probabilidades de pertenencia, $\theta^\pm = 0.7$, lo cual nos permite refinar el criterio de asignación de categoría a los textos, considerando que los elementos con una $\delta_i^\pm \geq \theta^\pm$ son los que el modelo clasifica con menor incertidumbre. En la figura 3.6 mostramos el número de textos positivos y negativos que satisfacen este criterio de asignación de categoría.

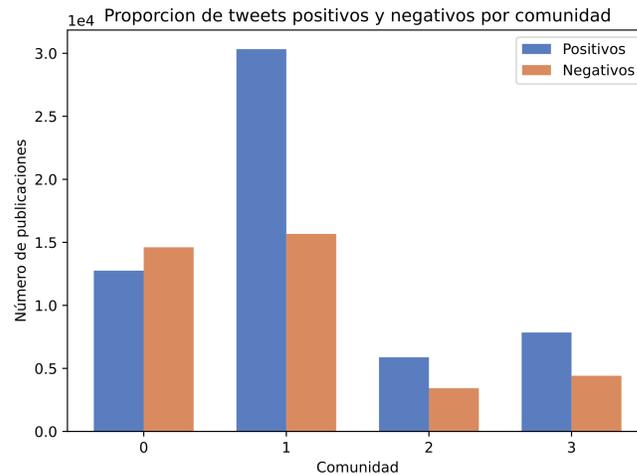


Figura 3.6: Número de publicaciones identificadas como positivas o negativas, con $\delta_i^\pm \geq \theta^\pm$, para cada una de las comunidades en la recolección de datos. Se puede observar que, aunque el número de elementos disminuye, la tendencia del sentimiento de los textos de cada comunidad se mantiene, al compararla con la figura 3.4.

De manera análoga al modelo basado en léxico (véase la sección 3.1), podemos calcular

la distribución de probabilidad de las probabilidades de pertenencia, p_i^+ y p_i^- , asignadas por el modelo, las cuales se muestran en la figura 3.7 y dónde es posible observar que para el caso de la comunidad 0, la distribución de p_i^- tiene un sesgo hacia valores ≈ 1 , lo cual puede interpretarse como una mayor presencia de textos con carga negativa. Cabe hacer énfasis en que a diferencia del modelo basado en léxico, la evaluación del modelo a cada texto, no representa la intensidad del sentimiento, sino la probabilidad de que pertenezca a la categoría positiva o negativa.

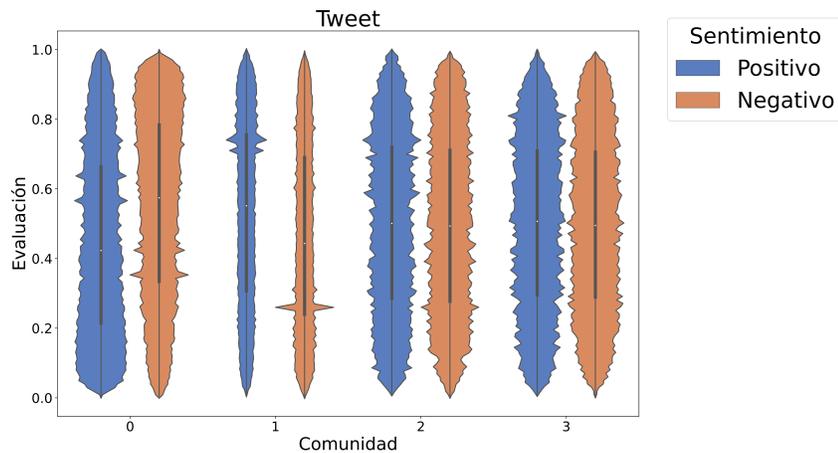


Figura 3.7: Distribuciones de probabilidad de las probabilidades de pertenencia, p_i^+ y p_i^- , asignadas por el modelo de RNA a los textos de tweets originales, de cada comunidad. Es posible observar que la comunidad 0 presenta una tendencia negativa, representada por un sesgo de la distribución de p_i^- hacia valores cercanos a 1.

En la figura 3.8 se muestran las distribuciones de probabilidad de p_i^+ y p_i^- , para los **retweets** y **replies** de cada comunidad. Es posible observar que los textos de los **retweets** de todas las comunidades tienen en su mayoría una carga positiva, mientras que en caso de los **replies**, solamente en el caso de la comunidad 0 hay una mayor densidad de textos con carga negativa.

Con el mismo propósito de estimar la polarización de los textos de las publicaciones de cada comunidad (véase la ecuación 3.1, y la discusión alrededor de ella), calculamos la polarización de cada comunidad, para cada tipo de publicación, partiendo de las medias de las distribuciones de probabilidad de p_i^+ y p_i^- , las cuales se muestran en la figura 3.9, en donde se pueden observar las comunidades y tipos de publicación para las cuales el modelo identificó una mayor densidad de textos con carga positiva o negativa, así como en donde se presentó una mayor polarización de los textos, entendido como una mayor diferencia entre el número de textos positivos y negativos.

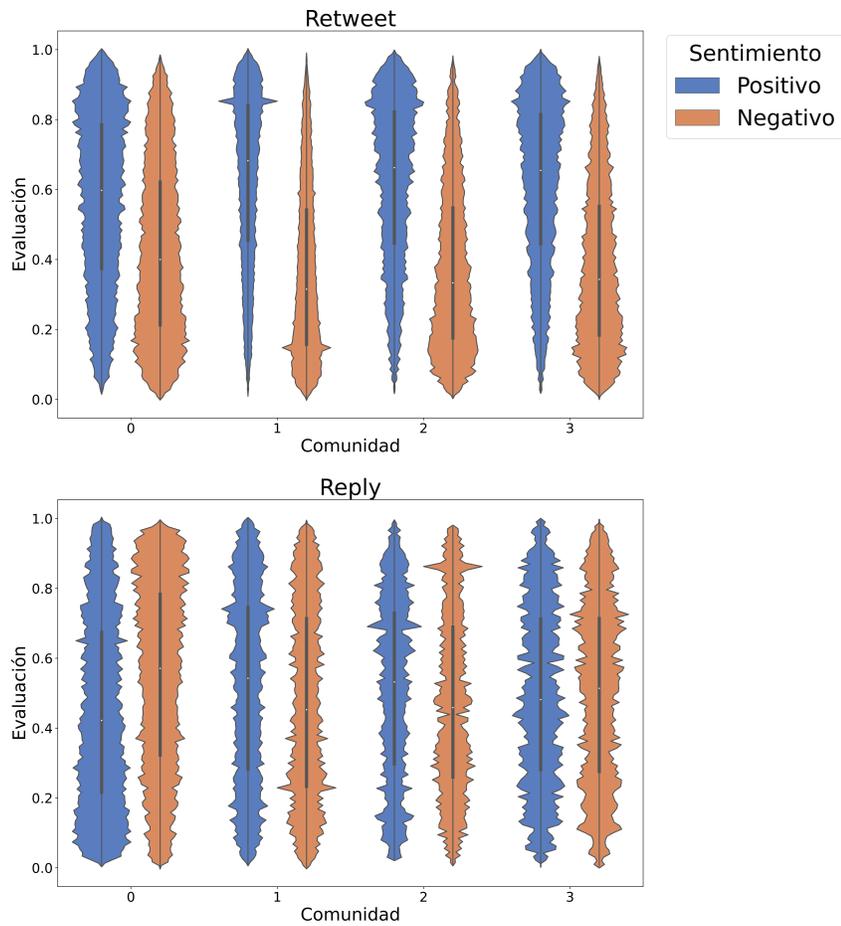


Figura 3.8: Gráfica de violín de las distribuciones de probabilidad de las evaluaciones de los replies y retweets para las comunidades que hablaron de la CNTE en twitter entre agosto y septiembre del 2013.

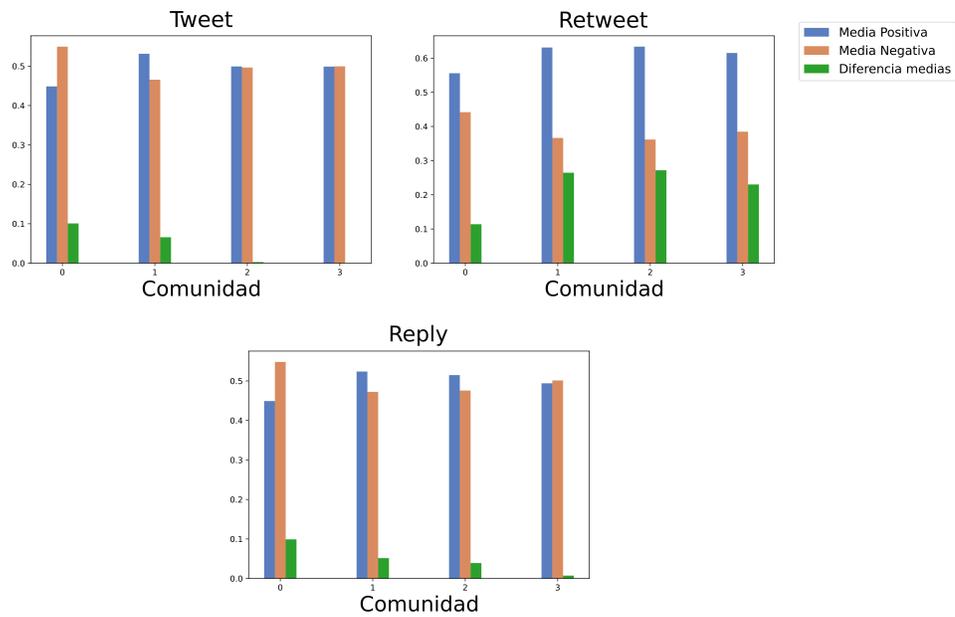


Figura 3.9: Comparación de las medias y su diferencia de los valores obtenidos de la predicción para las comunidades que hablaron de la CNTE en twitter entre agosto y septiembre del 2013.

Comparación de resultados

En las secciones anteriores mostramos los resultados obtenidos de cada modelo para el análisis de sentimiento de un conjunto de textos de publicaciones en la plataforma Twitter, referentes a las movilizaciones de la CNTE en la Ciudad de México en 2013.

A grandes rasgos, al examinar algunos de los ejemplos que obtuvieron las mayores calificaciones tanto positivas como negativas por ambos modelos, mostrados en la tabla 3.3 para el modelo basado en léxico y en la tabla 3.4 para el modelo de RNA, podemos concluir en que las evaluaciones hechas por los dos métodos identifican correctamente el sentimiento de los textos que les fueron presentados. Sin embargo, como también se puede constatar al examinar los ejemplos listados en la tablas, hay casos en los que los modelos clasifican como positivos textos que utilizan el sarcasmo o la ironía para criticar las movilizaciones de la CNTE. Esta es una de las principales limitaciones de estas estrategias, ya que, como hemos mencionado anteriormente, estos métodos no pueden hacer uso del contexto o identificar este uso del lenguaje.

Así mismo, ambos modelos identificaron a los textos de los tweets de la comunidad 0, como los que presentan una mayor carga positiva, ya sea como un mayor número de textos con alta probabilidad de pertenecer a la categoría negativa, como en el caso del modelo basado en RNA, o con una mayor intensidad de sentimiento negativo, como en el caso del modelo basado en léxico (VADER).

Sin embargo, al examinar con mayor detalle los valores de polarización, definida por la ecuación 3.1, obtenidos de las evaluaciones de cada modelo (véase la tabla 3.2), podemos observar que no hay una coincidencia en la comunidad que identifican con una mayor polarización, siendo la comunidad 0 para el modelo de RNA, y las comunidades 2 y 3 para el modelo de VADER. Mas aún, cabe hacer notar que los valores de polarización obtenidos corresponden a valores muy bajos, lo cual, aunado a la ausencia de un punto de referencia para estos valores, dificulta comparar ambos resultados e identificar con certeza si en efecto hay una polarización significativa en los textos de las comunidades. Es necesario aplicar ambos métodos de clasificación a un conjunto mayor de datos, alrededor de distintos tipos de eventos y temas, para poder tener un mejor punto de referencia.

Comunidad	Polarización Red Neuronal	Polarización VADER
Comunidad 0	0.1008	0.0165
Comunidad 1	0.0657	0
Comunidad 2	0.0027	0.0206
Comunidad 3	0	0.0206

Tabla 3.2: Comparación de la polarización en las comunidades de Twitter, usando los métodos de Red Neuronal y VADER. Se puede observar que no hay un consenso entre los modelos respecto a la comunidad con mayor polarización.

En general vemos que la diferencia de las medias, y por lo tanto de la polarización, es muy pequeña y podría indicar que la discusión llevada a cabo por los distintos usuarios está poco polarizada. Sin embargo, dadas las limitaciones de cada uno de los métodos aquí presentados, es necesario desarrollar y entrenar un conjunto de datos, tipo TASS, además de un léxico calificado por hablantes del español, que permita hacer un estudio detallado del sentir de las expresiones en las redes socio-digitales en español.

Texto	Sentimiento
@Washingtongue: Jaaaaaaaaaaaaaaaaa !!! #PinchesMaestros #CNTE... se deben estar empapando las basuras en su plantonzucho xD xD xD	+0.9228
México, país de libertad. Libertad de expresión #CNTE Libertad de acción #EPN Libertad de pensamientos #AMLO ¿Qué hacemos con la libertad?	+0.9517
@negrateran =D jajaja #UnPocoDeBuenHumor jajaja Milagro!! logrará #ElPeje camino en el Zócalo #CNTE?? jajaja	+0.953
Con la salida de #CNTE del Zócalo, ya @epn esta listo para gritar: VIVA LA PROLE K ME DIO PODER! Viva, VIVA SALINAS! VIVA. ONTIEL, VIVA YO	+0.9841
Odio odio odio a los del #CNTE - no puedo llegar a mi oficina por su causa !!!!! La ciudad es un asco!!!!	-0.9677
Duro Duro Duro Duro Duro #CNTE fuera del #zocalo #DF #PRIMERINFORMEMANCERA incapacidad incompetencia desinterés	-0.9612
Hay racistas que critican a la #CNTE pero criticar sus abusos NO es un acto racista.	-0.9512
#CNTE: La estupidez es una enfermedad extraordinaria, no es el enfermo el que sufre por ella, sino los demás. Ustedes están enfermos...	-0.9485
Los trogloditas de la #CNTE ya están cometiendo el delito de ATAQUES A LAS VIAS DE COMUNICACION, es DELITO FEDERAL Y SE CASTIGA CON TAMBO	-0.9423

Tabla 3.3: Ejemplos de los textos de algunos tweets de la comunidad 0, evaluados por el método VADER con los valores más altos de intensidad de sentimiento positivo y negativo.



Texto	Negativo	Positivo
@GinaPaola80 feliz inicio de semana y q esta nos dejen ser más productivos los del #CNTE, besos Lic	0.0040292447	0.9963329
Me encanta como la #CNTE es impune en esta ciudad... Gracias por ser un inútil @ManceraMiguelMX	0.004425528	0.9962792
Acuerda #CNTE #SEGOB y gobierno #Oaxaca replantear calendario escolar, reponer días perdidos e iniciar clases próximo lunes. #EnHoraBuena	0.004429606	0.9959487
Gracias Dios! Corrieron a vándalos #CNTE, empinaron a güilas e inicia un laaargo fin d semana! Y recuerden, el domingo a celebrar a México!	0.0057449467	0.99522495
que bueno que ya quitaron a los porros del #CNTE FELICIDADES a las autoridades del #DF Y #PF por tan grandiosa labor, #yabasta	0.0048826784	0.9948513
Buenas noches capitalinos que sueñen con el #CNTE y de pilon con @lopezobrador_ jajajaja	0.006268619	0.99454004
@ManceraMiguelMX si los porros de la #CNTE tienen derecho a manifestarse, yo tengo derecho a no pagar impuestos?	0.99523073	0.0041380376
Es represión. Siempre y cuando sea del lado de los que .ªsegún esto"tienen la razón como #CNTE, y cuando es al contrario como se le llama??	0.9911398	0.008039658
Se quejan de que no quitan a la #CNTE y cuando la cosa se pone fea dicen "no a la represión". No hay consolador que les entre.	0.99041325	0.008467846
#CNTE se le olvida que sus sueldos los pagan las personas que pagana impuestos y que #jodeaDiario @TapiaFernanda @ehalcazar	0.9891734	0.008910449
Ya cuando @ManceraMiguelMX no le hace caso a @LAURAZAPATAM, quiere decir que los hilos se mueven por otros lados #CNTE	0.98859125	0.010701731
No se trata d si sus demandas son justas o no. Se trata de q están afectando a quien no deben ¡Fuera #CNTE! NO #MeHagoPendejoComoMancera	0.98844916	0.012726655

Tabla 3.4: Ejemplos de los textos de algunos tweets de la comunidad 0, evaluados por el modelo de RNA con las probabilidades de pertenencia más altas, tanto para la clase positiva como la negativa.

►4. Propuestas para la extensión y mejora de los modelos

Si bien los resultados obtenidos de los modelos de análisis de sentimiento implementados, han sido satisfactorios en el sentido de que han sido capaces de identificar correctamente el sentimiento general de los textos que les fueron presentados (véanse las secciones 3.1 y 3.2 para observar los resultados obtenidos del modelo basado en léxico y en RNA respectivamente y las tablas 3.3 y 3.4 para observar ejemplos específicos de textos evaluados por los modelos), consideramos que es posible implementar estrategias que permitan obtener mejores resultados.

►4.1. Ampliación del conjunto de datos de entrenamiento para la RNA

La existencia de un conjunto de datos previamente anotados o evaluados, ya sea en la forma de un corpus de textos, como en el caso del corpus TASS, o en la forma de un léxico de palabras, emojis o acrónimos para el caso del modelo basado en léxico, es fundamental para la implementación de modelos de análisis de sentimiento como los que presentamos en este documento.

En el caso del modelo basado en RNA, el conjunto de datos de entrenamiento fue el corpus general TASS de 2012, debido a que es el que contiene más elementos. Sin embargo, los textos que contiene corresponden a español de España y, aunque como hemos podido comprobar, nos han permitido identificar la tendencia general del sentimiento en los textos asociados a cada comunidad, es probable que el uso de este corpus no permita clasificar correctamente textos con expresiones o contextos locales.

Como una primer estrategia de mejoría de nuestros modelos, proponemos complementar el conjunto de datos de entrenamiento de la RNA, considerando las actualizaciones más recientes disponibles del corpus TASS en las que se incluyen ejemplos con las variantes del español de Perú, Costa Rica, Chile, Uruguay y México.

Entrenar nuestro modelo con este conjunto de datos extendido, puede permitirle tener una mayor precisión en la evaluación de los textos de publicaciones en redes socio-digitales, en el contexto mexicano y latinoamericano. Lo cual será evaluado al comparar los resultados obtenidos de la aplicación del modelo que presentamos en este trabajo, y de un modelo entrenado con el corpus extendido, a distintos conjuntos de datos a alrededor de eventos relevantes recientes.

Sin embargo, aunque existan textos con variantes locales del español en el corpus TASS, es probable que no sean suficientes para obtener una mejoría significativa en nuestro modelo o para capturar expresiones utilizadas en la conversación cotidiana en las redes socio-digitales, ya que el lenguaje utilizado en este tipo de contextos muta rápidamente.

Por lo que, con el objetivo de atender estas situaciones, proponemos la creación de una **plataforma de evaluación colaborativa**, en el espíritu de la ciencia ciudadana [6], que permita realizar una evaluación continua y plural de datos.

→4.2. Estrategias de evaluación colaborativa

Los conjuntos de datos utilizados durante los procesos de entrenamiento y aprendizaje de los modelos, para posteriormente aplicarlos a conjuntos de datos para ser evaluados, representan una manera de extender o extrapolar las evaluaciones ya presentes a nuevas observaciones. En este sentido, los modelos de clasificación son susceptibles de reforzar o amplificar sesgos presentes en los datos de entrenamiento, los cuales han sido tratados en su mayoría en el ámbito técnico y de desempeño de los algoritmos, dejando de lado la discusión ética acerca de ellos [8], a pesar de que pueden resultar en resultados y prácticas discriminatorias hacia sectores vulnerables de la sociedad [4, 11].

Con el objetivo de generar **corpus y léxicos con contenido más cercano a las expresiones y contexto local**, y con información precisa y transparente del procedimiento de evaluación de datos, proponemos generar una **plataforma de evaluación colaborativa** de publicaciones en redes socio-digitales, alrededor de eventos relevantes en la conversación en México.

Esta plataforma nos permitirá tener acceso a una evaluación diversa desde distintos grupos socio demográficos, e idealmente, reducir el sesgo en la evaluación del sentimiento de los textos.

Evaluación del sentimiento de textos en redes sociales digitales

El corpus TASS puede considerarse como un estándar para la implementación y evaluación de modelos de aprendizaje de máquina e inteligencia artificial para el análisis de sentimiento de textos de redes socio-digitales. Sin embargo, los elementos del corpus, están relacionados con el contexto español y es posible que se pierdan sutilezas y expresiones en el momento de realizar la extrapolación de la evaluación del corpus a datos del contexto local.

Nos encontramos en el proceso de la recopilación y sistematización de textos de publicaciones en Twitter alrededor de eventos políticos y sociales recientes en el contexto mexicano, los cuales estarán disponibles en una plataforma abierta de evaluación colaborativa, en la que, posteriormente a un registro sencillo en el cual recopilamos información demográfica básica, el usuario procede a evaluar un conjunto de textos.

Una vez que contemos con el corpus de datos anotados, procederemos a entrenar un modelo de RNA con dicho corpus para comparar con los resultados obtenidos con el corpus TASS.

Daniela o Daniel como alternativa a VADER

Cómo se mencionó anteriormente, en el caso del método de análisis de sentimiento basado en léxico VADER, la estrategia de traducir, ya sea el texto o el léxico anotado, para evaluar el sentimiento de textos en español, no es óptima. Por lo que es necesario desarrollar un **léxico anotado colaborativo**, que considere palabras y expresiones acorde



a las peculiaridades de las variantes locales del lenguaje, para la evaluación directa de los textos en español sin tener que recurrir al uso de traducciones automáticas.

En específico, la evaluación de las palabras, expresiones y acrónimos representativos de las variantes locales del idioma español que compondrán los léxicos, se realizará a través de nuestra plataforma de evaluación colaborativa, basada en la metodología de la sabiduría de la multitud.

Proponemos que este léxico (o léxicos) lleve como nombre, alguna de las siguientes opciones:

- **DANIEELA:**
Diccionario para el ANálisis de la Intensidad y el sEntimiento del Español LAtinoamericano
- **DANIEELHA:**
Diccionario para el ANálisis de la Intensidad y el sEntimiento del Español HispanoAmericano
- **DANIEEL_MX:**
Diccionario para el ANálisis de la Intensidad y el sEntimiento del Español Mexicano

Como se puede observar de las opciones listadas, la elección del nombre, y por lo tanto de las estrategias para determinar el contenido de cada léxico, depende del alcance y complejidad que se desee considerar. En principio, por razones prácticas, se comenzará con el desarrollo del proyecto piloto para el léxico DANIEEL_MX, para posteriormente, en base a la experiencia adquirida y a posibles colaboraciones que puedan establecerse en el transcurso del desarrollo del proyecto, extender el léxico al resto de sus variantes.

► 5. Conclusiones

Los modelos de análisis de sentimiento que hemos implementado, nos han permitido identificar el sentimiento general de los textos de publicaciones en Twitter alrededor de eventos focales, lo cual se refleja en las tendencias observadas en las distribuciones de la evaluación de los textos asignadas por los modelos (véase el capítulo 3). En particular, pudimos identificar un grupo de cuentas, cuyas publicaciones tienen una carga negativa mayor en comparación con el resto.

El uso de este tipo de técnicas en el contexto del estudio de las culturas políticas y la identificación de narrativas alrededor de eventos focales es relevante, ya que permite analizar de manera más eficiente y sistemática, el contenido que se genera y comparte en las redes socio-digitales (en este caso en particular el texto de las publicaciones), lo cual, en diálogo con una interpretación multidisciplinaria considerando aspectos cualitativos y de contexto del evento focal, nos permite tener una visión más completa del tipo de conversación que se desarrolla en las redes socio-digitales.

Por ejemplo, en el contexto del conjunto de datos que utilizamos en este trabajo para poner a prueba los modelos, nos muestra el descontento y la estigmatización de las movilizaciones de la CNTE realizada por un conjunto de cuentas, o un tratamiento más neutral (en el sentido de no observar una tendencia clara en la distribución de las evaluaciones de los textos) por otros grupos de cuentas. En este sentido, es importante mencionar que, la aplicación y los resultados obtenidos de los modelos de análisis de sentimiento son de carácter estadístico, es decir, es necesario aplicarlos e interpretarlos para un conjunto de elementos (una colección de textos en este caso en particular), y no a elementos individuales, ya que esto puede llevar a conclusiones sesgadas o erróneas.

Los resultados de la aplicación de este tipo de técnicas, son un insumo adicional para la interpretación y entendimiento de las narrativas alrededor de los eventos, y no un fin último.

Así mismo, aunque hemos obtenido resultados satisfactorios de los modelos que implementamos, consideramos que éstos pueden ser mejorados y extendidos (véase el capítulo 4), en particular, consideramos que las estrategias de evaluación colaborativa que proponemos serán útiles, no sólo para la investigación que desarrollamos en Tlatelolco Lab, sino para el estudio del uso del lenguaje en redes socio-digitales y la construcción colectiva de corpus y léxicos, que reflejen la diversidad de las variantes locales del español y nos permitan entender de mejor manera nuestra relación con el contenido y el uso de estas plataformas digitales.

Bibliografía

- [1] Tara H. Abraham. «(Physio)logical circuits: The intellectual origins of the McCulloch-Pitts neural networks». En: *Journal of the History of the Behavioral Sciences* 38.1 (2002), págs. 3-25. DOI: 10.1002/jhbs.1094.
- [2] W.F. Brinkman, D.E. Haggan y W.W. Troutman. «A history of the invention of the transistor and where it will lead us». En: *IEEE Journal of Solid-State Circuits* 32.12 (1997), págs. 1858-1865. DOI: 10.1109/4.643644.
- [3] Erik Cambria y Bebo White. «Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]». En: 9.2 (2014), págs. 48-57. DOI: 10.1109/mci.2014.2307227.
- [4] Jeffrey Dastin. *Amazon scraps secret AI recruiting tool that showed bias against women*. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>. 2018.
- [5] *Diccionario Etimológico Castellano En Línea*. Available at [http://etimologias.dechile.net/?comunicar,\(2021/10/23\)](http://etimologias.dechile.net/?comunicar,(2021/10/23)).
- [6] Susana Finkelievich y Celina Fischnaller. «Citizen science in the Information Society. New world trends». En: 2014.
- [7] M. Girvan y M. E. J. Newman. «Community structure in social and biological networks». En: *Proceedings of the National Academy of Sciences* 99.12 (2002), págs. 7821-7826. ISSN: 0027-8424. DOI: 10.1073/pnas.122653799. eprint: <https://www.pnas.org/content/99/12/7821.full.pdf>. URL: <https://www.pnas.org/content/99/12/7821>.
- [8] Thilo Hagendorff. «Ethical behavior in humans and machines - Evaluating training data quality for beneficial machine learning». En: *CoRR* abs/2008.11463 (2020). arXiv: 2008.11463. URL: <https://arxiv.org/abs/2008.11463>.
- [9] C. Hutto y Eric Gilbert. «VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text». En: *Proceedings of the International AAAI Conference on Web and Social Media* 8.1 (2014), págs. 216-225. URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.
- [10] Ackerman John y col. *Fundamentos teóricos de Tlatelolco Lab. Nueve dimensiones para el estudio de las luchas por la democracia*. Inf. téc. UNAM, Mexico: Programa Universitario de Estudios sobre Democracia, Justicia y Sociedad, 2021.
- [11] Angwin Julia y col. *Machine Bias*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Mayo de 2016.
- [12] Warren S. McCulloch y Walter Pitts. «A logical calculus of the ideas immanent in nervous activity». En: *The Bulletin of Mathematical Biophysics* 5.4 (1943), págs. 115-133. DOI: 10.1007/bf02478259.



- [13] Prakash M Nadkarni, Lucila Ohno-Machado y Wendy W Chapman. «Natural language processing: an introduction». En: 18.5 (2011), págs. 544-551. DOI: [10 . 1136 / amiajn1-2011-000464](https://doi.org/10.1136/amiajn1-2011-000464).
- [14] Kevin Priddy. *Artificial neural networks: an introduction*. Bellingham, Wash: SPIE, 2005. ISBN: 0819459879.
- [15] Sebastian Raschka y Vahid Mirjalili. *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow, 2nd Edition*. 2nd. Packt Publishing, 2017. ISBN: 1787125939.
- [16] Frank Rosenblatt. «The perceptron—a perceiving and recognizing automation». En: *Cornell Aeronautical Laboratory* (1957).
- [17] Antonio Fernández Anta Philippe Morere Luis F. Chiroque Agustín Santos. «Techniques for Sentiment Analysis and Topic Detection of Spanish Tweets: Preliminary Report». En: *Spanish Society for Natural Language Processing Conference (SEPLN 2012)* (2012).
- [18] Claude E. Shannon. «A Mathematical Theory of Communication». En: *The Bell System Technical Journal* 27 (1948), págs. 379–423, 623–656.
- [19] Kent C. B. Steer, Andrew Wirth y Saman K. Halgamuge. «The Rationale Behind Seeking Inspiration from Nature». En: (2009), págs. 51-76. DOI: [10 . 1007 / 978 - 3 - 642 - 00267-0_2](https://doi.org/10.1007/978-3-642-00267-0_2).
- [20] Julio Villena Román y col. *TASS - Workshop on Sentiment Analysis at SEPLN*. 2013-03.
- [21] Steven Walczak y Narciso Cerpa. «Artificial Neural Networks». En: *Encyclopedia of Physical Science and Technology (Third Edition)*. Ed. por Robert A. Meyers. Third Edition. New York: Academic Press, 2003, págs. 631-645. ISBN: 978-0-12-227410-7. DOI: <https://doi.org/10.1016/B0-12-227410-5/00837-1>. URL: <https://www.sciencedirect.com/science/article/pii/B0122274105008371>.