

# UNIVERSIDAD SANTO TOMÁS

---

## USO DEL ANÁLISIS DE SENTIMIENTOS PARA MEDIR EL POSICIONAMIENTO DE UNA MARCA A TRAVÉS DE LA INFORMACIÓN EXTRAÍDA DE LA PLATAFORMA TWITTER

---

Realizado por

Nicolás Santiago Cortés Díaz

Anteproyecto enviado en cumplimiento del requisito parcial  
para optar por el grado de Ingeniería Electrónica



Grupo de Investigación GED (Grupo de Estudio y Desarrollo en Robótica)  
Facultad de Ingeniería Electrónica  
División de Ingenierías

Septiembre de 2021

---

**USO DEL ANÁLISIS DE SENTIMIENTOS PARA  
MEDIR EL POSICIONAMIENTO DE UNA MARCA A  
TRAVÉS DE LA INFORMACIÓN EXTRAÍDA DE LA  
PLATAFORMA TWITTER**

---

Realizado por

**Nicolás Santiago Cortés Díaz**

Anteproyecto enviado en cumplimiento del requisito parcial  
para optar por el grado de Ingeniería Electrónica

Dirigido por

**Armando Mateus Rojas**

Co-Dirigido por

**Sindy Paola Amaya**

GED

Facultad de Ingeniería Electrónica  
División de Ingenierías

Septiembre de 2021

# Índice general

<b>Lista de Figuras</b>	<b>III</b>
<b>Lista de Tablas</b>	<b>IV</b>
<b>1. Planteamiento del Problema</b>	<b>2</b>
<b>2. Justificación</b>	<b>4</b>
<b>3. Impacto Social</b>	<b>5</b>
<b>4. Estado del Arte</b>	<b>6</b>
<b>5. Objetivos</b>	<b>9</b>
5.1. Objetivo General . . . . .	9
5.1.1. Objetivos Específicos . . . . .	9
<b>6. Marco Teórico</b>	<b>10</b>
6.1. Aprendizaje Automático . . . . .	10
6.1.1. Aprendizaje Supervisado . . . . .	11
6.1.2. Aprendizaje no Supervisado . . . . .	11
6.2. Procesamiento del Lenguaje Natural . . . . .	12
6.3. Análisis de Sentimientos . . . . .	12
6.3.1. Modelos clasificadores y herramientas de análisis . . . . .	13
6.3.1.1. Naive Bayes . . . . .	13
6.3.1.2. VADER . . . . .	13
6.4. Proceso de Entrenamiento . . . . .	14
6.4.1. Preprocesamiento . . . . .	14
6.4.1.1. Normalización de mayúsculas y minúsculas . . . . .	14
6.4.1.2. Eliminación de stopwords . . . . .	15
6.4.1.3. Tokenización . . . . .	16
6.4.1.4. Stemming . . . . .	16
6.4.1.5. Clasificación . . . . .	17
<b>7. Metodología</b>	<b>20</b>
7.0.1. Acceso de la Información . . . . .	21

---

7.0.2. Manejo de la información . . . . .	23
<b>8. Resultados</b>	<b>28</b>
<b>Bibliografía</b>	<b>37</b>

# Índice de figuras

1.	Clasificación del aprendizaje automático. / Fuente: Autor . . . . .	10
2.	Nivel del análisis de sentimientos. Fuente: Autor / Adaptada de [22] . . . . .	12
3.	Normalización del texto. / Fuente: Autor . . . . .	15
4.	Eliminación de palabras vacías. / Fuente: Autor . . . . .	16
5.	Tokenización del texto. / Fuente: Autor . . . . .	16
6.	Raíz y Morfema de una palabra. / Fuente: Autor . . . . .	17
7.	Stemming o análisis morfológico. / Fuente: Autor . . . . .	17
8.	Clasificación del texto en positivo, negativo o neutro. / Fuente: autor . . . . .	18
9.	Metodología. / Fuente: Autor . . . . .	20
10.	Plataforma de Desarrollo de Twitter. Fuente: Autor . . . . .	21
11.	Menú de control - Portal de Desarrolladores. Fuente: Autor . . . . .	22
12.	Aplicación de proyecto. Fuente: Autor . . . . .	22
13.	Credenciales - Keys and Tokens. Fuente: Autor . . . . .	23
14.	Conjunto de entrenamiento. / Fuente: Autor . . . . .	28
15.	Resultado sentimientos sobre la marca uno (PlayStation). / Fuente: Autor . . . . .	30
16.	Resultado sentimientos sobre la marca dos (Xbox). / Fuente: Autor . . . . .	30
17.	Twitter extraído. / Fuente: Twitter . . . . .	31
18.	Nube de palabras PlayStation. / Fuente: Autor . . . . .	32
19.	Nube de palabras Xbox. / Fuente: Autor . . . . .	33
20.	Puntuación asignada al texto. / Fuente: Autor . . . . .	34
21.	Nube de palabras - Compañías. / Fuente: Autor . . . . .	35

# Índice de tablas

1.	Palabras vacías en español . . . . .	15
2.	Palabras positivas, negativas y neutras. . . . .	18
3.	Estadística de las palabras en el texto . . . . .	18
4.	Credenciales generadas . . . . .	23
5.	Prueba uno realizada . . . . .	29
6.	Prueba dos realizada . . . . .	29
7.	Prueba tres realizada . . . . .	29
8.	Puntuación de la marca . . . . .	31
9.	Probabilidad del tweet por palabra . . . . .	32
10.	Categoría y puntuación de las compañías . . . . .	34

# Introducción

En la actualidad, muchas de las relaciones que una empresa establece con los clientes se hace a través de medios como las redes sociales y son tendencia de marketing para comunicar acciones corporativas. Las redes sociales se ofrecen como canales para transmitir valores de la marca, encontrar a nuevos consumidores, publicar información de marketing y brindar una retroalimentación de clientes insatisfechos.

El constante crecimiento de las redes facilita la comunicación de millones de usuarios que activamente interactúan en estas plataformas, además de ofrecer una mayor visibilidad digital de la marca o negocio junto a los productos y servicios que comercializan. Twitter cuenta con 353 millones usuarios [1] y se caracteriza por generar una gran cantidad de datos y mensajes diariamente. Los conocidos “tweets” son expresiones de un punto de vista u opiniones de los usuarios que generan interés para analizarlos. Este proceso de extracción de datos permite identificar información útil sobre las opiniones y conocer informáticamente los sentimientos, las emociones y las actitudes de las personas hacia ciertos temas.

La inteligencia artificial, por medio del procesamiento del lenguaje natural prepara a las computadoras para entender e interpretar el lenguaje humano, es decir, programa a la máquina para que tenga el poder de leer texto, escuchar sonidos y proporcionar una medición del sentimiento. Considerando la gran cantidad de información basada en el lenguaje, las máquinas son indispensables para un análisis rápido, automático y consistente.

Este proyecto brinda información acerca del proceso que hay detrás del análisis de sentimientos utilizando el clasificador bayesiano de texto y el algoritmo VADER, apoyado de algunas librerías ofrecidas por python para el procesamiento del lenguaje natural. Los resultados presentados son producto del análisis entre algunas marcas competidoras en el mercado.

# Capítulo 1

## Planteamiento del Problema

La rivalidad de las marcas por mostrar superioridad en el mercado y demostrar estar un paso adelante de la otra es un tema cotidiano. Cada año, nuevos productos y nuevos servicios salen al mercado y son presentados. Las redes sociales cambiaron directamente el entorno social, la forma de relacionarnos y el intercambio de información entre las personas. Además, también son un buen sitio para promocionar las nuevas ideas debido a la gran cantidad de usuarios que poseen, asimismo de tener la posibilidad de responder a la situación a partir de los comentarios y reseñas sobre cada uno de los productos según lo expresa [2].

Twitter es una de las plataformas más efectivas para realizar campañas de marketing online. Considerado un servicio de microblogueo, su interfaz permite monitorizar y medir aproximadamente todas las acciones que se realizan, con lo cual dado en [3] siempre se podrá ver el alcance y la respuesta comunicativa e incluso mejorar y corregir si es necesario la estrategia . Sin embargo, la complejidad al momento de obtener un informe sobre sus productos y servicios conlleva a las empresas a buscar alternativas como el análisis de sentimientos para conseguirlo.

Como se informa en [4], el análisis de sentimientos permite analizar las opiniones públicas de los usuarios sobre un tema particular y ayuda a clasificarlas como positivas o negativas. Aun así, es cierto que puede resultar fácil el análisis a simple vista, pero que manualmente, por la gran cantidad de datos, no es factible. Es por ello que las marcas responden a estos resultados para evolucionar en la forma de ofrecer sus productos o buscar alternativas para mejorar sus servicios pero de una manera automatizada. El constante seguimiento de los datos brinda la posibilidad de crear nuevas estrategias para influir en las decisión de las personas y lograr posicionarse exitosamente.

A partir de lo planteado se formulará la siguiente pregunta de investigación:

¿Cómo aplicar la técnica de análisis de sentimientos Naive Bayes para medir el posicionamiento de marca en la plataforma Twitter?

## Capítulo 2

# Justificación

El análisis de sentimientos busca extraer opiniones, sobre un tema de interés desde el lenguaje natural de los textos. El proceso de clasificación se realiza automáticamente a partir de algoritmos y según el sentimiento que transmite la opinión pueden ordenarse en positivas, negativas o neutras. La importancia se fundamenta en la percepción de la realidad como también de las decisiones que se toman, sin importar la condición por como otras personas ven y perciben el mundo. Por esta razón, desde el punto de vista útil es considerable conocer los criterios u opiniones de otras personas sobre temas relevantes con el objetivo de buscar algunas aplicaciones como por ejemplo recomendar productos y servicios, determinar a que candidato político se votará en las próximas elecciones o incluso medir la opinión publica ante la medida tomada por una empresa o un gobierno como se puede leer en [5].

Generalmente, el análisis intenta medir la actitud del usuario sobre algún contenido o la polaridad contextual general de un documento. El usuario puede tornar una actitud de juicio o evaluación, estado afectivo (el estado emocional del autor al momento de escribir), o la intención comunicativa emocional (el efecto emocional que el autor intenta causar en el lector). La principal idea del análisis de sentimientos es brindar un monitoreo de las redes sociales y plantear ideas o estrategias de la opinión publica que se expreso naturalmente.

Los resultados según [6] del análisis han probado que los cambios reflejados en el sentimiento de las plataformas corresponden a una variación en la bolsa, es decir, si el carácter del sentimiento hacia la empresa es negativo y existe repercusión, la empresa puede sufrir perdidas debido a que las ventas bajan considerablemente. De igual forma, la empresa puede tomar los datos para realizar un cambio en las estrategias y de ser así brindar una mejor atención al usuario.

## Capítulo 3

# Impacto Social

Las redes sociales se han convertido en una herramienta para millones de usuarios de todo el mundo gracias a los grandes avances tecnológicos ya que posibilita la comunicación desde cualquier lugar disponible con conexión a internet. El constante uso de estas redes como Facebook, Twitter, WhatsApp puede contribuir a reducir impactos psicológicos y/o sociológicos en la sociedad.

Hoy en día, cualquier acción en las redes puede ser medible y monitorizada. En la analítica de datos, el bienestar psicológico, resulta ser una prioridad, y la manera en como se maneja la información puede ayudar a distinguir enfermedades mentales en las personas, tratarlas a tiempo y de manera adecuada. Planteado en [7], el bienestar psicológico es fundamental porque representa el grado de estabilidad mental de las personas, permitiéndoles manejar y desarrollar cognitivamente y socialmente funciones con total plenitud.

## Capítulo 4

# Estado del Arte

En [8] Meylan Wongkar y Apriandy Angdresey definen el análisis de sentimiento como la actividad para distinguir el nivel de sentimiento u opinión pública y presentan un estudio de las opiniones expresadas de los usuarios en Twitter sobre las elecciones presidenciales mediante una aplicación de Python. El análisis consiste en recopilar datos mediante las bibliotecas que ofrece Python, el procesamiento de texto y pruebas de datos de entrenamiento y la clasificación de texto empleando el método de Naive Bayes. Los resultados arrojados por la aplicación mostraron el valor de la polaridad del sentimiento positivo y negativo de los candidatos y mediante un programa informático de análisis y minería de datos se realizó una comparación entre otros algoritmos como Support Vector Machines (SVM) y K-Nearest Neighbor (K-NN) para conocer la precisión de cada uno.

En [9] Lokesh Mandloi y Ruchi Patel determinan el análisis de sentimientos como una clasificación de los sentimientos que se encuentran expresados en texto como positivos, negativos o neutrales. El constante uso de las plataformas sociales y la gran cantidad de usuarios activos que cuentan facilita el trabajo del análisis de sentimientos de la sociedad. De igual manera realizan una comparación entre varios métodos de aprendizaje automático como Naive Bayes, clasificación de máquinas de vectores de soporte y la clasificación de máxima entropía con el objetivo de conocer el funcionamiento, la exactitud y la precisión de cada uno de estos algoritmos de clasificación.

En [10] V. Prakruthi, D. Sindhu y Dr. S. Anupama Kumar tratan el análisis de sentimientos como la aplicación para procesar el lenguaje natural, el análisis de textos, la lingüística computacional y la biometría para reconocer, extraer, cuantificar y aprender estados afectivos e información subjetiva. El análisis lo realizan a un conjunto de mensajes o tweets publicados por usuarios

---

sobre una persona, tendencia, producto o marca en la red social Twitter con el objetivo de evaluar el sentimiento. Los datos se extraen mediante una plataforma de desarrollo que ofrece una API y que permite acceder a los datos publicados con el fin de realizar la implementación de una aplicación, clasificar y representar los sentimientos expresados en los comentarios como negativos, neutrales y positivos utilizando técnicas de visualización como el histograma y el gráfico circular.

En [11], según los autores Chae Won Park y Dae Ryong Seo, el análisis de sentimiento se aplica para comprender la tendencia del usuario. El objetivo de los autores es determinar estadísticamente que asistente de inteligencia artificial es más preciso a partir de un léxico llamado Valence Aware Dictionary y sEntiment Reasoner (VADER). Asimismo, se realiza un análisis de los tweets para mostrar la significancia estadística a partir de diferentes procedimientos, la prueba T de muestras independientes, la prueba de Kruskal-Wallis y la prueba de Mann-Whitney, cuyo fin es presentar los resultados del análisis estadístico para los tres asistentes de inteligencia artificial: Siri de Apple, Google Assistant de Google, y Cortana de Microsoft.

En [12], los autores Sanjeev Dhawan, Kulvinder Singh y Priyanka Chauhan, describen el análisis del sentimiento como una denominación de la minería de suposición, que infiere el sentimiento o el estado de ánimo de un usuario. Incluso, en Python, se propone e implementa un método para realizar el análisis de un conjunto de información extraída de la red social Twitter que se basa en la polaridad de cada uno de los mensajes con el fin de clasificar si el sentimiento fue negativo o positivo, es decir, si el usuario expreso enojo, tristeza, felicidad o alegría.

En [13] utilizan una técnica llamada PLN (procesamiento del lenguaje natural) para analizar el sentimiento y determinar el tipo de opinión o la polaridad de las publicaciones realizadas en Twitter. Esta técnica consiste en la tokenización que significa dividir las cadenas de texto extensas en piezas más pequeñas o tokens. Los enfoques de los autores Kusriani y Mochamad Mashuri se centran en el desarrollo del análisis utilizando léxicos y polaridad de multiplicación, pero los resultados son mas imprecisos después de realizar una comparación con el aprendizaje automático.

En [14] se enfocan en la gran cantidad de datos de la red social Twitter para realizar un análisis de los sentimientos de la información y lo definen como un área de la minería de datos de texto y PLN. Los autores Rasika Wagh y Payal Punde presentan los tipos de análisis de sentimientos, los métodos que se utilizan para clasificar los sentimientos extraídos de los tweets como Naïve-Bayes, SVM y Entropía máxima y un estudio para comparar los resultados de las diferentes técnicas. Los enfoques fundamentales en el análisis de sentimientos son el enfoque de aprendizaje supervisado y el enfoque de aprendizaje no supervisado.

---

En [15] Sahar A. El Rahman, Feddah Alhumaidi AlOtaibi y Wejdan Abdullah AlShehri presentan un modelo para realizar un análisis de sentimientos de datos reales extraídos de Twitter combinando el uso de algoritmos de aprendizaje automático supervisados y no supervisados. Las etapas que aplican para el análisis parten del acceso de la información directamente tomada de la API de Twitter, seguido del preprocesamiento de los datos con el objetivo de entrenar la información resultante en un modelo basado en léxico y clasificar los tweets en sentimiento positivo, negativo o neutral.

En [16] los autores M. Trupthi, Suresh Pabboju y G. Narasimha implementan un sistema automático para predecir el sentimiento de los mensajes o tweets publicados por los usuarios en las redes sociales utilizando hadoop. El objetivo es analizar los sentimientos en tiempo real a partir de un clasificador basado en características y en resumen de opiniones para proporcionar analíticas basadas en el tiempo al usuario.

En [17] Dionysis Goularas y Sani Kamis realizan una comparación de varios métodos de aprendizaje profundo utilizados para el análisis de sentimientos en los datos de Twitter. Particularmente utilizan dos categorías derivadas de las redes neuronales; las redes neuronales convolucionales (CNN), y las redes neuronales recurrentes (RNN) que comúnmente se aplican en tareas de procesamiento del lenguaje natural (PLN).

En [18], según los autores Lei Wang, Jianwei Niu y Shui Yu, los patrones de difusión de sentimientos en Twitter se relacionan ligeramente con las polaridades de sentimiento de los mensajes de Twitter. El enfoque que proporcionan se basa en fusionar información textual de los mensajes de Twitter y los patrones de difusión de sentimientos para mejorar el rendimiento del análisis en los datos de Twitter. El proceso para fusionar la información inicialmente consta del análisis de la difusión de la opinión a partir de la investigación de un fenómeno llamado inversión de la opinión cuyo motivo es encontrar algunas propiedades. Posterior se consideran las relaciones entre la información textual de los tweets y los patrones de difusión de sentimientos con el fin de implementar un algoritmo llamado SentiDiff que permite predecir o clasificar las polaridades de los sentimientos expresados en los mensajes de Twitter.

En [19], de acuerdo con los autores Md. Rakibul Hasan, Maisha Maliha y M. Arifuzzaman, el proceso de filtraje de los datos de twitter se realiza mediante un marco de datos preprocesados basados en el procesamiento del lenguaje natural (PLN). Los resultados del análisis de sentimientos se fundamentan en los modelos Bag of Words (BoW) y Término Frecuencia-Frecuencia Inversa de Documentos (TF-IDF) que finalmente clasifican los tweets en positivos y negativos con una alta precisión.

# Capítulo 5

## Objetivos

### 5.1. Objetivo General

Emplear un modelo de clasificación Naive Bayes de análisis de sentimientos para medir el posicionamiento de marca a través de la plataforma Twitter.

#### 5.1.1. Objetivos Específicos

- Descargar un conjunto de datos de tweets positivos y negativos para el entrenamiento del modelo.
- Utilizar una técnica para el análisis de sentimientos de la información extraída para clasificar el nivel de sentimiento.
- Comparar los resultados del análisis de sentimientos de algunas marcas para conocer la opinión de los usuarios sobre cada una.

# Capítulo 6

## Marco Teórico

### 6.1. Aprendizaje Automático

El machine learning o aprendizaje automático consiste en identificar automáticamente distintos patrones o tendencias que se encuentran presentes entre los datos a través de diferentes algoritmos. La implementación de estos algoritmos pueden clasificarse en tres tipos: Aprendizaje supervisado, Aprendizaje no supervisado y Aprendizaje reforzado.

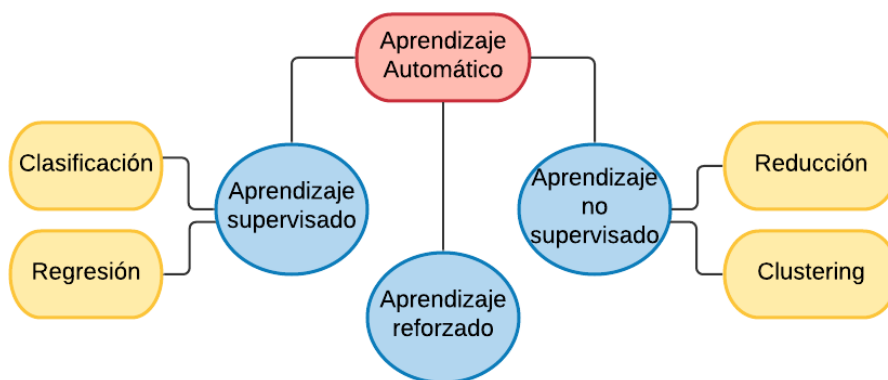


FIGURA 1: Clasificación del aprendizaje automático. / Fuente: Autor

En la figura 1 se encuentran los tipos de aprendizaje que comprenden el aprendizaje automático junto a los modelos de clasificación para cada uno. De [20], se puede definir el aprendizaje supervisado y no supervisado como:

### 6.1.1. Aprendizaje Supervisado

El aprendizaje supervisado permite realizar predicciones futuras que se basan en el comportamiento o características identificadas en datos históricos etiquetados. Este conjunto de técnicas relacionan la predicción como una función de variables de entrada y salida donde los valores que entran corresponden a las características analizadas mientras que los datos de salida corresponden a la variable que se quiere predecir. Sin embargo, la salida de la función puede ser de tipo numérico o de tipo categórico respectivamente del problema de clasificación (regresión o clasificación).

Algunos de los algoritmos mas aplicados en este tipo de aprendizaje son los siguientes:

- Árboles de decisión.
- Clasificación de Naïve Bayes.
- Regresión por mínimos cuadrados.
- Regresión Logística.
- Support Vector Machines (SVM).
- Métodos “Ensemble” (Conjuntos de clasificadores).

### 6.1.2. Aprendizaje no Supervisado

El aprendizaje no supervisado utiliza un conjunto de técnicas para inferir patrones y extraer información de un conjunto de datos no referenciados o desconocidos. A comparación del aprendizaje supervisado, la única información conocida son los datos de entrada en vista de que no existen datos de salida que correspondan con cada una.

Algunos de los algoritmos mas populares en este tipo de aprendizaje se listan a continuación:

- Algoritmos de clustering
- Análisis de componentes principales
- Descomposición en valores singulares

## 6.2. Procesamiento del Lenguaje Natural

El procesamiento del lenguaje natural (PLN) como se evidencia en [21] es un campo de la inteligencia artificial y la lingüística que se centra y estudia la comunicación entre los seres humanos y las máquinas mediante el uso del lenguaje natural, que procura identificar las partes más relevantes del mensaje que puede estar expresado por escrito, oralmente o a través de signos. No obstante, independientemente del objetivo de la aplicación, el estudio se puede descomponer en análisis morfológico o léxico, sintáctico, semántico o pragmático.

## 6.3. Análisis de Sentimientos

El análisis de sentimiento lo definen en [22] como el proceso evaluativo de los comentarios escritos por los usuarios para extraer información subjetiva y determinar la polaridad a partir de una herramienta automática. Cuando se habla de polaridad, las oraciones pueden dividirse en objetivas y subjetivas. Las oraciones objetivas no informan opinión mientras que las oraciones subjetivas, al depender de un punto de vista, si contienen y son fundamentales para identificar y estimar la polaridad positiva, neutral o negativa. De igual manera, es necesario tener en cuenta el cómo se quiere trabajar al momento de realizar la extracción de la polaridad.

Observemos la siguiente figura:

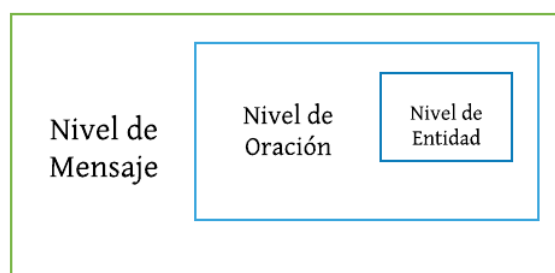


FIGURA 2: Nivel del análisis de sentimientos. Fuente: Autor / Adaptada de [22]

Cada nivel representado en la figura 2 se puede definir como la forma a realizar el análisis. A nivel de mensaje, la clasificación se realiza tomando el mensaje completo, a nivel de oración, la clasificación de la polaridad se realiza independientemente para cada división del mensaje, y a nivel de entidad cualquier opinión se forma por un sentimiento dirigido a un sujeto.

Sin embargo, el análisis puede resultar complejo debido a ciertas características gramaticales como la morfología o el modo de expresar las opiniones. También, las opiniones pueden ser explícitas o implícitas, donde se expresa directamente o indirectamente la opinión respectivamente.

### 6.3.1. Modelos clasificadores y herramientas de análisis

Los clasificadores son algoritmos, que a partir de un entrenamiento asigna los datos proporcionados a diferentes categorías para establecer la polaridad de los mensajes de Twitter.

#### 6.3.1.1. Naive Bayes

Para [23], Naive Bayes es un clasificador probabilístico que aplica el Teorema de Bayes para estimar la probabilidad  $P(y|x)$  de la clase  $y$  dada la variable  $x$

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)} \quad (6.1)$$

De la ecuación 6.1 podemos observar que el denominador  $P(x)$  es constante para algún valor de  $y$ , por lo que no es necesario calcularlo en orden de tomar una predicción. Por lo tanto podemos usar la siguiente aproximación:

$$P(y|x) = P(x|y)P(y) \quad (6.2)$$

Donde  $P(y|x)$  se define como la probabilidad posterior,  $P(x|y)$  como el valor de probabilidad Likelihood y  $P(y)$  como el valor de probabilidad Prior.

A partir de esos valores, la predicción en los tweets se puede expresar como:

$$p = \logprior + \sum_i^N (\loglikelihood_i) \quad (6.3)$$

#### 6.3.1.2. VADER

Valence Aware Dictionary y Sentiment Reasoner (VADER) es un algoritmo para el análisis de sentimientos que utiliza un listado de palabras anotadas (léxico de sentimiento) cada una con una puntuación de sentimiento correspondiente y una lista de palabras que modifican las puntuaciones de las palabras en el texto:

- Impulsores: son palabras o n-gramas que intensifican el sentimiento de las palabras tokenizadas. Por ejemplo, palabras como "absolutamente" y "asombrosamente".
- Amortiguadores: son palabras o n-gramas que amortiguan el sentimiento de las palabras tokenizadas. Por ejemplo, palabras como "apenas" y "algo".
- Negaciones: son palabras que niegan el sentimiento de las palabras tokenizadas. Por ejemplo, palabras como "no" y "no es".

La puntuación compuesta se calcula sumando las puntuaciones de valencia de cada palabra en el léxico con el fin de normalizar los valores entre -1 y +1 y poder realizar la clasificación negativa o positiva respectivamente como se informa en [24]:

$$x = \frac{x}{\sqrt{x^2 + \alpha}} \quad (6.4)$$

donde  $x$  representa la suma de las puntuaciones de valencia de las palabras constituyentes y  $\alpha$  una constante de normalización.

## 6.4. Proceso de Entrenamiento

### 6.4.1. Preprocesamiento

Antes de empezar a extraer las características es necesario realizar un tratamiento a los datos que van a ser entrenados para evitar posibles afectaciones en el resultado. Es muy común encontrar mensajes en las redes escritos de maneras diferentes, con errores de ortografía, con algún carácter repetido, con mezclas entre mayúsculas y minúsculas o con algunos emoticones.

Tomando en cuenta estos factores, el objetivo es normalizar los datos, para evitar la pérdida de la polaridad de sentimiento. El preprocesamiento sigue las siguientes reglas:

#### 6.4.1.1. Normalización de mayúsculas y minúsculas

Las palabras expresadas en mayúsculas o minúsculas tienen diferentes significados para los algoritmos de aprendizaje, es decir, una palabra en mayúscula es completamente diferente a la misma palabra en minúscula. En la figura 3, la palabra "excelente" contenida en uno de los mensajes no representa el mismo significado de la misma palabra "Excelente" del otro mensaje para el algoritmo aun si contextualmente signifiquen lo mismo.

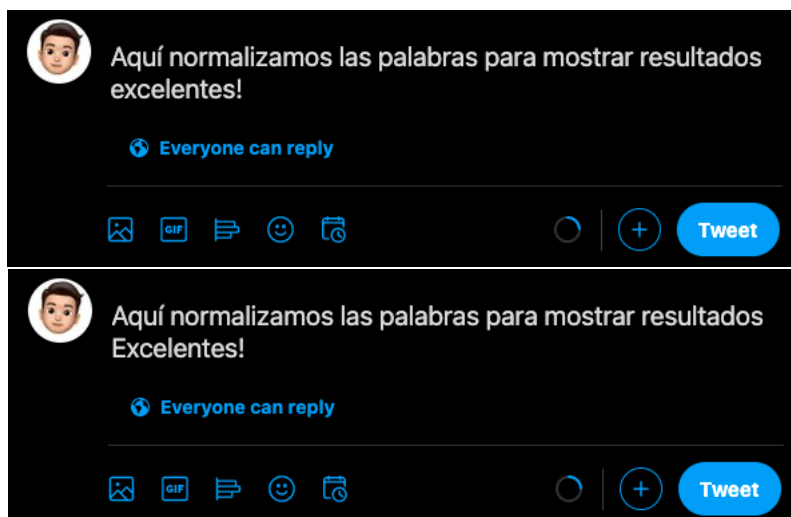


FIGURA 3: Normalización del texto. / Fuente: Autor

#### 6.4.1.2. Eliminación de stopwords

Las preposiciones, los pronombres, las conjunciones hacen parte de un conjunto de palabras que son necesarias para construir oraciones con sentido pero carecen de información para determinar la polaridad de los textos.

En la tabla 1 se muestran algunas palabras vacías o sin significado que estructuran los textos y que deberán ser filtradas antes del procesamiento.

Palabras Vacías						
a	actualmente	cuales	cualquier	para	que	un
ahí	ahora	da	dado	ningunos	si	una
algunas	alguno	aquí	las	pues	siendo	tenga

TABLA 1: Palabras vacías en español

Los mensajes de entrada pueden contener muchas de estas palabras pero serán eliminadas porque no se encuentran dentro de la lista de palabras con significado y no aportan en la clasificación. En la figura 4, el mensaje normalizado anteriormente muestra una menor longitud y contiene solo las palabras que abundan de información.

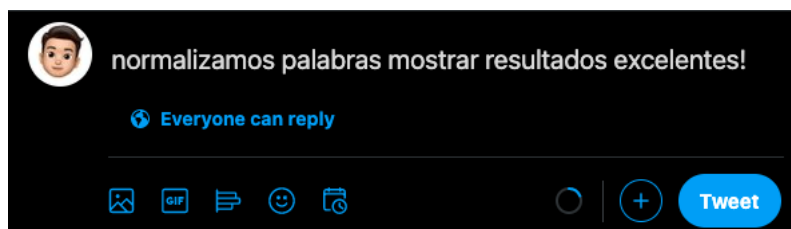


FIGURA 4: Eliminación de palabras vacías. / Fuente: Autor

### 6.4.1.3. Tokenizacion

Este proceso denominado tokenizacion divide el texto en oraciones o palabras para facilitar el análisis léxico. Las palabras se segmentan como se observa en la figura 5 y pueden entenderse como una secuencia de caracteres separados por espacios en blanco.

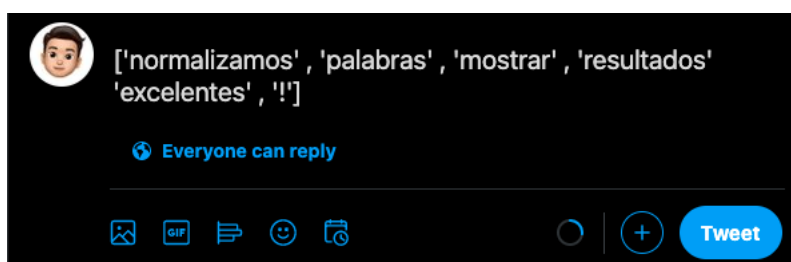


FIGURA 5: Tokenizacion del texto. / Fuente: Autor

### 6.4.1.4. Stemming

Definido en [25] el stemming es un método de análisis morfológico que asocia variantes de la palabra a su raíz por medio de la supresión de sus sufijos e inflexiones. Por ejemplo, como se observa en la figura 6 las palabras «Teléfono», «Telégrafo», «Telescopio» y «Teleférico» están asociadas a la raíz «Tele».

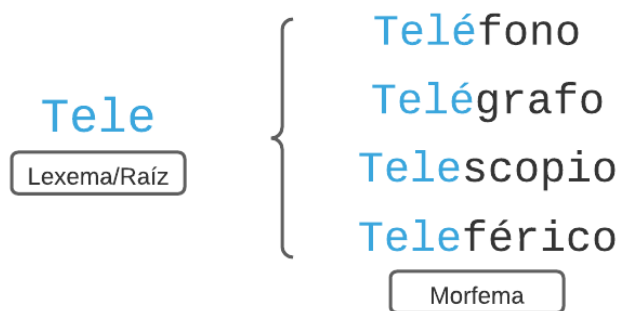


FIGURA 6: Raíz y Morfema de una palabra. / Fuente: Autor

En la figura 7 se realiza el proceso de morfología de las palabras para el mensaje entrante.

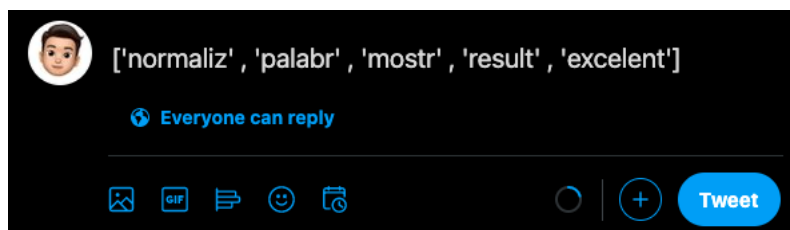


FIGURA 7: Stemming o análisis morfológico. / Fuente: Autor

#### 6.4.1.5. Clasificación

Finalizando, los datos normalizados se ingresan en un analizador de Naive Bayes y clasifica el texto en positivo, negativo o neutro. Las tres categorías de mensajes se muestran en la figura 8:



FIGURA 8: Clasificación del texto en positivo, negativo o neutro. / Fuente: autor

Después del procesamiento previo y la extracción de características, de los mensaje clasificados anteriormente, se pueden enlistar en la tabla 2 algunas palabras en las tres categorías:

Positivas	Neutras	Negativas
alegría	docente	odio
excelente	estudiante	egoísmo
felicidad	minero	miedo

TABLA 2: Palabras positivas, negativas y neutras.

El clasificador, según [26] se basa en un conjunto de funciones que generan estadísticamente que palabras y cuantas veces aparecen en cada texto. En la tabla 3 se supone un modelo de las palabras mencionadas anteriormente con los siguientes datos:

Positivas		Neutras		Negativas	
alegría	21	día	1	odio	5
excelente	3	estudiante	17	egoísmo	4
felicidad	12	proyecto	2	miedo	35
total	36	total	20	total	44
100 palabras en total					

TABLA 3: Estadística de las palabras en el texto

Los datos corresponden hipotéticamente a la cantidad de palabras que se mencionan o se repiten por cada mensaje de entrenamiento independientemente de la categoría a la que pertenece. Cada una de las palabras del texto preprocesado denotan una probabilidad y se calcula a partir de la ecuación (6.2).

## Capítulo 7

# Metodología

Desde la plataforma de desarrolladores de Twitter se accede y se conecta a la API para extraer una cantidad limitada de datos relacionados con los comentarios e interacciones de los usuarios sobre algunas marcas. El algoritmo está desarrollado en Python y utiliza la librería TextBlob, apoyado de nltk, la cual aplica el análisis de sentimientos y clasifica los tweets en positivos, negativos o neutros. Cada una de las etapas se muestran en la figura 9 y se irán explicando brevemente:



FIGURA 9: Metodología. / Fuente: Autor

### 7.0.1. Acceso de la Información

Inicialmente para acceder a la plataforma de desarrollador se debe crear o iniciar sesión en una cuenta de Twitter. Después de tener una sesión activa, como se observa en la figura 10, en la barra de navegación podemos observar y acceder a una opción que dice "Developer Portal".

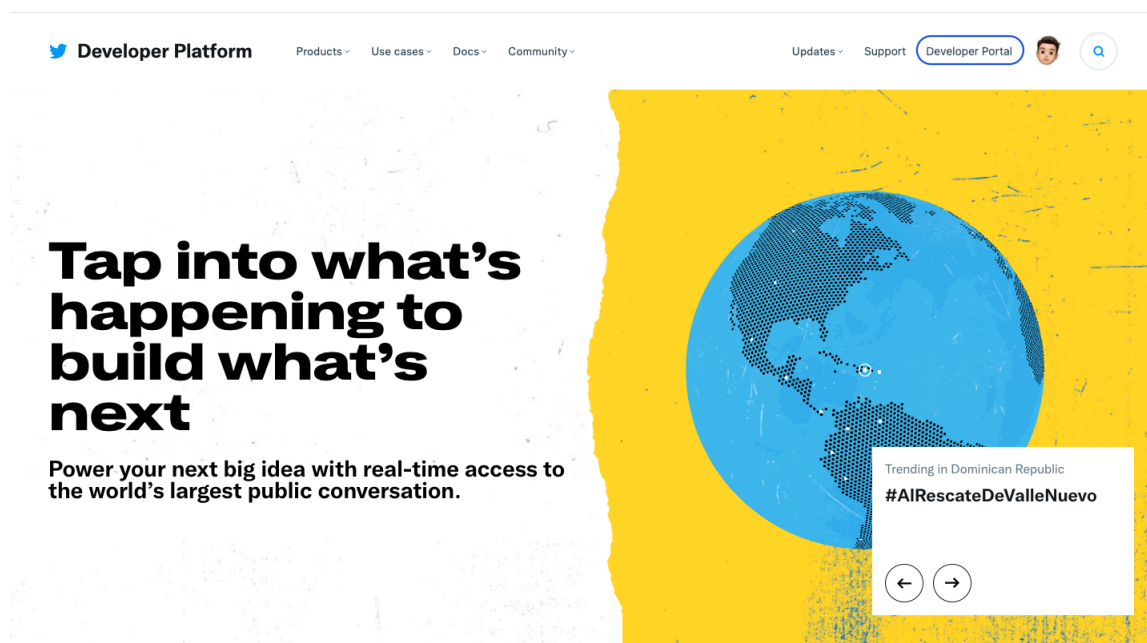


FIGURA 10: Plataforma de Desarrollo de Twitter. Fuente: Autor

Estando dentro del portal de desarrollador podemos observar un menú de control (ver figura 11) que permite crear e indicar detalladamente el nombre del proyecto y su objetivo para acceder a la información.

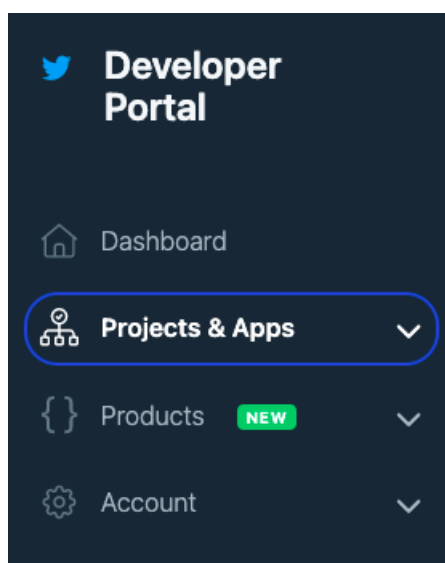


FIGURA 11: Menú de control - Portal de Desarrolladores. Fuente: Autor

Una vez creado el proyecto, se puede ingresar a otro espacio que permite generar las llaves y tokens para acceder a la información de las publicaciones de los usuarios. En la figura 12 se observa la opción correspondiente enmarcada por un cuadro azul.

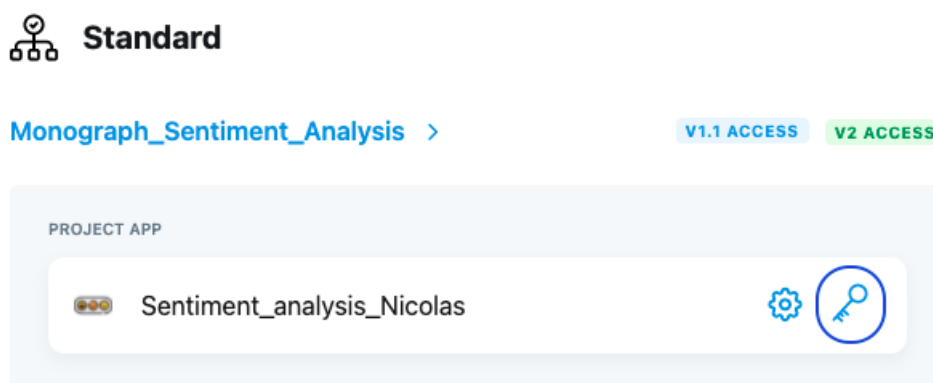


FIGURA 12: Aplicación de proyecto. Fuente: Autor

En la figura 13 se pueden observar las credenciales que se pueden generar para ingresar a la información mediante Python. La API Key le permite realizar una solicitud a partir de un nombre de usuario, la API Key Secret indica la contraseña de ese usuario, el Access Token y el Access Token Secret representan la cuenta propietaria de la aplicación y permite realizar una solicitud en nombre de esta.

## Consumer Keys ⓘ

API Key and Secret

Regenerate

## Authentication Tokens ⓘ

Bearer Token

Generate

Access Token and Secret

Generated May 2, 2021

For @Nicolas04040221

Created with **Read Only** permissions

Regenerate

Revoke

FIGURA 13: Credenciales - Keys and Tokens. Fuente: Autor

Las llaves generadas dentro de la aplicación permiten realizar solicitudes que requieren autenticación de contexto de usuario OAuth 1.0a y se muestran en la tabla 4:

<b>API Key</b>	<a href="#">rXcjFkZjGG0FZuzAamEQtmfX3</a>
<b>API Key Secret</b>	<a href="#">cpyj6pAtwF10IuY439H6iwiZqbALSSvIgv0TivRue2pl8u9z6</a>
<b>Access Token</b>	<a href="#">1370042835253268483-k4B2r1Hf0000AsVN7Qy3EuoztI859e</a>
<b>Access Token Secret</b>	<a href="#">ISMnqP5Y3ADPHoEKmHnUpFyDTEXx4dom9WdDV4cB3zH2r</a>

TABLA 4: Credenciales generadas

### 7.0.2. Manejo de la información

Ahora, teniendo acceso a la información es necesario colocar las claves como variables de entrada para realizar la autenticación de usuario mediante Python utilizando la librería tweepy. Los mensajes por defecto contienen símbolos y algunos emoticones construidos con los mismos que no aportan valor a la clasificación. La búsqueda de los datos se puede hacer mediante una palabra de un tema de interés, proporcionado la cantidad de datos a extraer e indicando el idioma de cada uno. Al ser autenticado y sabiendo que los datos no se encuentran preprocesados

es necesario definir una función para limpiar cada uno de esos mensajes que se van a extraer implementando los pasos mencionados en el proceso de entrenamiento. Para cada uno se aplicarán la tokenización, la eliminación de palabras vacías, emoticones, signos de puntuación y por último el proceso de stemming.

Continuando, los datos limpiados se analizan mediante un clasificador de Naive Bayes para determinar el nivel de sentimiento del número de tweets extraídos previamente. Terminando, los datos son graficados y como anexo se genera una nube de palabras que indica las cantidades de palabras más usadas en el entrenamiento. Los siguientes pseudocódigos muestran un bosquejo de los pasos realizados para el análisis de sentimientos de la información extraída de Twitter utilizando la técnica de Naive Bayes y el algoritmo de VADER:

---

**Algorithm 1** Análisis de sentimientos expresados en Twitter - Naive Bayes / Tomado de: <https://blog.hacemoscontactos.com/2018/07/28/analisis-de-sentimientos-sobre-twitter-usando-la-libreria-textblob-de-python/>

---

```

1: Credentials API Twitter
2: auth(ConsumerKey, ConsumerSecrets, AccessToken, AccessTokenSecret)
3: Input: wordSearch, lenguajeWord, numTweets
4: function CLEAN_TWEETS(tweet)
5:     emoticons = happy, sad
6:     tweetTokens = sent_tokenize(tweet, "spanish")
7:     tweetsClean = []
8:     for word in tweetTokens do
9:         remove.stopwordsSpanish
10:        remove.emoticons
11:        remove.punctuation
12:        stemWord = stemmer.stem(word)
13:        tweetsClean.append(stemWord)
14:     end for
15:     return tweetsClean
16: end function
17: function GET_DATA(word, times, lenguaje)
18:     positive = 0
19:     negative = 0
20:     listNumber = []
21:     number = 1
22:     for Number of Tweets do
23:         clean_tweets()
24:         NaiveBayesAnalyzer()
25:         tweet.sentiment
26:         listNumber.append(number)
27:         number = number + 1
28:     end for
29:     return analysisClear, listNumber, number
30: end function
31: function PLOT_DATA(listNumber, values)
32:     plt.scatter(listNumber, values)
33: end function

```

---

Por otro lado, VADER se fundamenta en 5 reglas o heurísticas para percibir la intensidad del sentimiento en el texto a nivel de oración: la puntuación, las mayúsculas, los modificadores de grados, el cambio de polaridad debido a las conjunciones y la detección de negación de la polaridad.

---

**Algorithm 2** Análisis de sentimientos expresados en Twitter - VADER / Tomado de: <https://github.com/Amey23/Twitter-Sentimental-Analysis>

---

```

1: Credentials API Twitter
2: auth(ConsumerKey, ConsumerSecrets, AccessToken, AccessTokenSecret)
3: Input: companies
4: function CLEANTWEETS(tweets)
5:   removePattern(tweets, "@, https?:/")
6:   return tweets
7: end function
8: function GETDATA(companies)
9:   scores = []
10:  compoundList = []
11:  positiveList = []
12:  negativeList = []
13:  neutralList = []
14:  for Number of Tweets do
15:    compound = analyzer.polarityScores
16:    pos = analyzer.polarityScores
17:    neg = analyzer.polarityScores
18:    neu = analyzer.polarityScores
19:    scores.append(compound, pos, neg, neu)
20:  end for
21: end function
22: function PLOTDATA(compound, pos, neg, neu)
23:   scoreTable.plot(compound, pos, neg, neu)
24: end function

```

---

Sin embargo, según [27], se puede resaltar que la librería Valence Aware Dictionary and Sentiment Reasoner (VADER) utiliza una combinación de métodos cualitativos y cuantitativos con el objetivo de generar y validar un léxico de sentimiento estándar.

Por otra parte, como se expresa en [28], el gran avance en el mundo de los videojuegos y la idea de sumar popularidad ha llevado a las marcas más representativas y con más presencia en los mercados a posicionarse exitosamente. Un análisis realizado en el año 2020 reveló, que, para un grupo de diferentes edades, las marcas más preferidas son PlayStation, Xbox y Nintendo.

De acuerdo con la información, la marca más importante para los usuarios es PlayStation pues se trata de una serie de productos de gaming que han acompañado a estos grupos de edad desde hace más de 20 años. Por otra parte, Xbox se ubicó en la novena posición, aunque cabe señalar que para la generación Z se trata de la marca más importante de videojuegos, por lo que el relevo generacional podría dar grandes resultados para la división de juegos de Microsoft en

un futuro cercano. Finalmente, Nintendo se ubicó en la décima posición, gracias a las nuevas experiencias de juego que ofrece Switch, su consola actual.

## Capítulo 8

# Resultados

Se realizaron tres pruebas de laboratorio. La primera utiliza un conjunto (ver figura 14) de frases propias de prueba y entrenamiento para verificar el rendimiento del clasificador empleando la librería TextBlob:

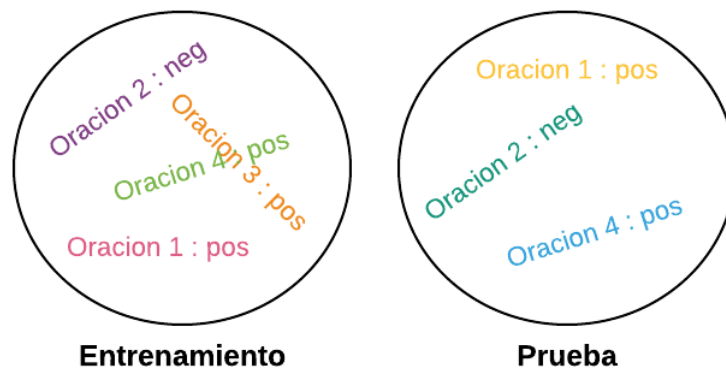


FIGURA 14: Conjunto de entrenamiento. / Fuente: Autor

Para un texto de entrada aplicando el clasificador bayesiano y tokenizando a través de oraciones se obtienen los siguientes resultados:

Palabra de Entrada			
La cerveza se disfruta. Pero la resaca es horrible.			
Tokenizador			
La cerveza se disfruta		Pero la resaca es horrible.	
<b>sentimiento:</b>	+	<b>sentimiento:</b>	-
Precisión: 0.833			

TABLA 5: Prueba uno realizada

Palabra de Entrada			
Amo las hamburguesas. Sin embargo, odio la cebolla			
Tokenizador			
Amo las hamburguesas		Sin embargo, odio la cebolla	
<b>sentimiento:</b>	+	<b>sentimiento:</b>	-
Precisión: 0.833			

TABLA 6: Prueba dos realizada

Palabra de Entrada			
Estoy estresado. Mi jefe me cae mal			
Tokenizador			
Estoy estresado		Mi jefe me cae mal	
<b>sentimiento:</b>	-	<b>sentimiento:</b>	-
Precisión: 0.833			

TABLA 7: Prueba tres realizada

La prueba dos se realiza con un conjunto de datos sobre dos marcas extraídos de la cuenta oficial de PlayStation y Xbox. Los resultados mostrados en la figura 15 y 16 arrojaron que, existe una gran cantidad de comentarios positivos para ambas marcas, pero también cuentan con comentarios negativos.

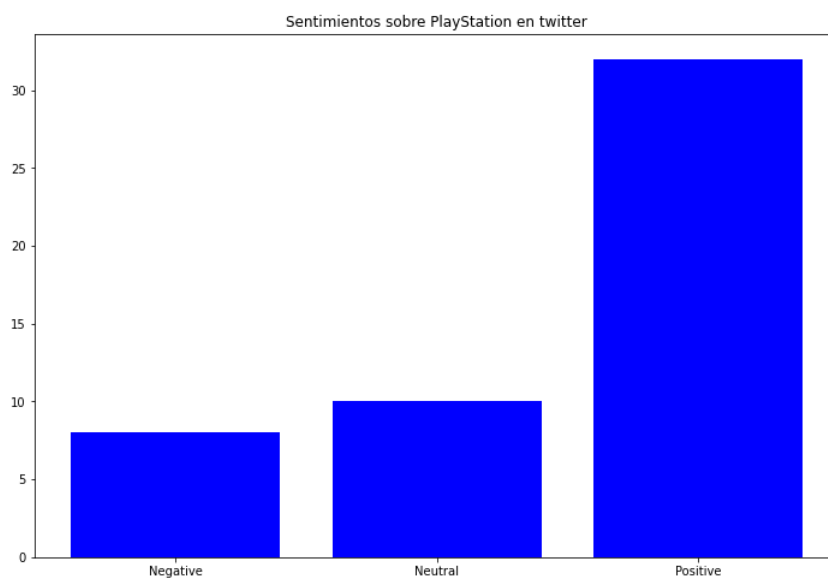


FIGURA 15: Resultado sentimientos sobre la marca uno (PlayStation). / Fuente: Autor

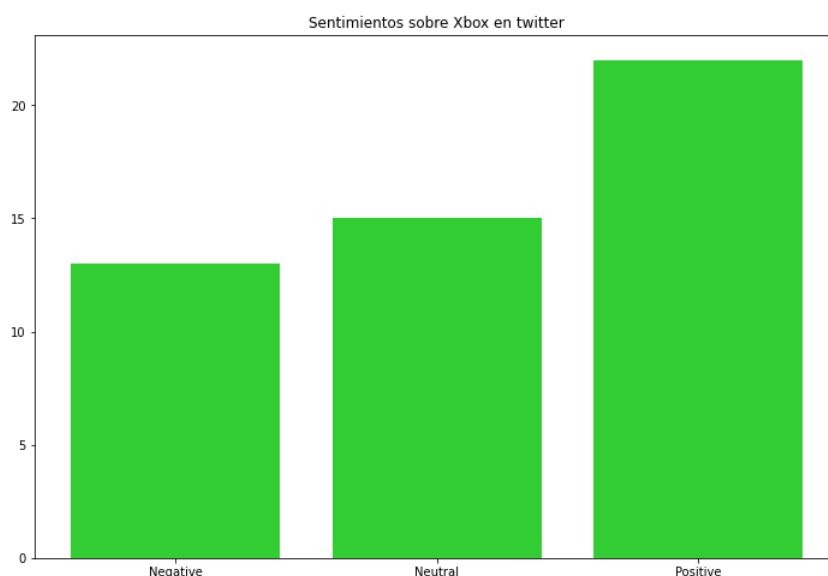


FIGURA 16: Resultado sentimientos sobre la marca dos (Xbox). / Fuente: Autor

Observando la figura 15 y 16 podemos determinar que tan bien se habla de estas marcas con respecto a la otra. Con estos resultados y conociendo la cantidad de tweets para cada uno se plantea la siguiente metodología:

Los comentarios positivos generan una puntuación de (+3) puntos, los neutrales (+1) punto y los comentarios negativos (-1). De esta forma podemos conocer que tan buena imagen representa la marca hacia los usuarios:

Marca	Cantidad de tweets	tweets positivos	tweets negativos	tweets neutros	Puntuación
PlayStation	50	33	7	10	102
Xbox	50	21	13	16	66

TABLA 8: Puntuación de la marca

Los resultados de la tabla 8 permiten identificar la puntuación para ambas marcas. La marca uno contiene una puntuación mayor a la segunda lo que indica que se encuentra mejor calificada por el público y por consecuencia mejor posicionada en el mercado.

Pero observemos de cerca, como se polarizan estos mensajes: La figura 17 muestra un tweet extraído desde la API y encontrado en la plataforma. El mensaje fue procesado y probado en una función para realizar una predicción la cual devuelve la probabilidad del tweet perteneciente a la clase positiva o negativa.



FIGURA 17: Twitter extraído. / Fuente: Twitter

Estos datos son producto del entrenamiento de un conjunto de datos que contienen tweets positivos y negativos y de los valores probabilísticos encontrados de logprior y loglikelihood. Sin embargo, para realizar la predicción, los datos de test son los mensajes extraídos desde la API de Twitter y arrojan los siguientes resultados:

Palabras	Probabilidad	Polaridad
gauu	1,10	positivo
asombroso	1,18	positivo
divertido	0,79	positivo
Twitch	1,61	positivo
playstation	-1.09	negativo
wii	-1.09	negativo
amor	0.97	positivo
jugando	-0,05	neutral
juegos	-0.07	neutral
oh	-0,56	negativo

TABLA 9: Probabilidad del tweet por palabra

De esta manera, a partir de los resultados de la tabla 9 podemos conocer porque el tweet fue un mensaje positivo.

La nube de palabras representa la cantidad de palabras que están presentes en el mensaje y la cantidad de veces que se repite. Entre mas grande sea la palabras mas apariciones tiene. En la figura 18 y 19 se muestra la cantidad de aparición que tuvo esa palabra en el mensaje publicado en Twitter respectivamente de la marca:



FIGURA 18: Nube de palabras PlayStation. / Fuente: Autor



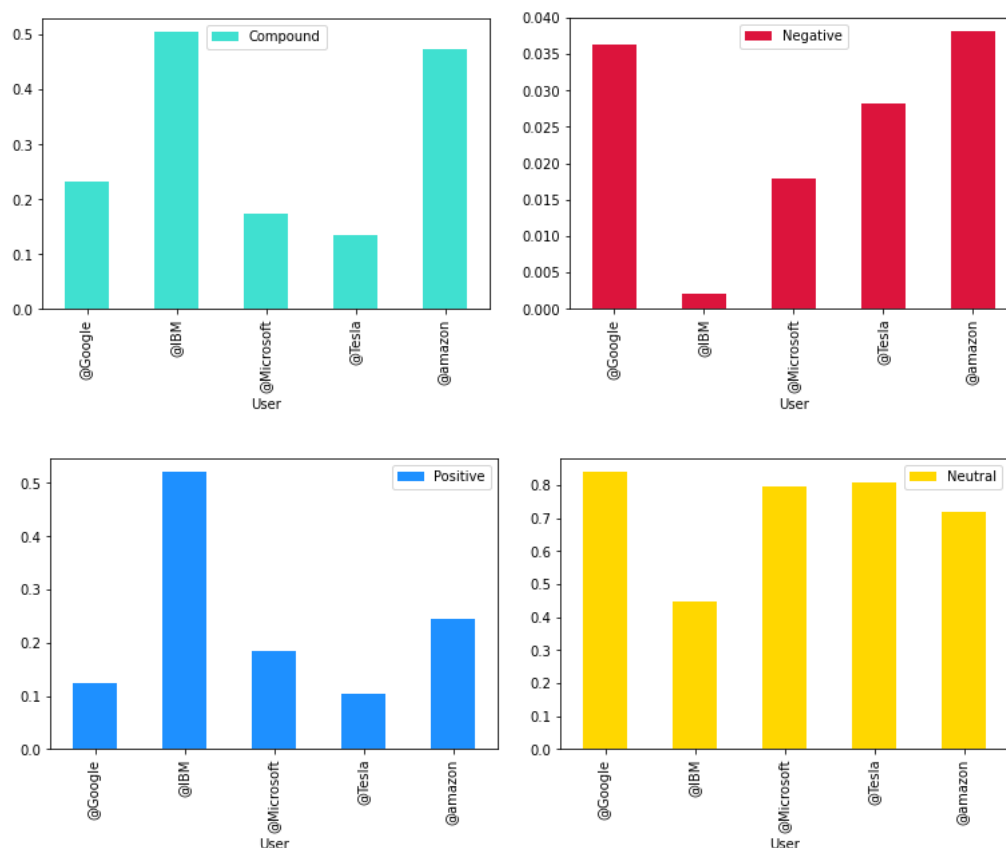


FIGURA 20: Puntuación asignada al texto. / Fuente: Autor

Con base a los resultados de la figura 20 podemos calificar a las empresas según los comentarios. En la tabla 10 podemos observar que ante el público todas son compañías muy neutras pero una resulta ser más positiva para la mayoría. De esta forma podemos conocer, que entre su rivalidad, no existe gran diferencia para con su público.

Compañía	Puntaje positivo	Puntaje neutral	Puntaje negativo	Puntuación
<b>Google</b>	10.5 %	84.8 %	3.8 %	Neutral
<b>IBM</b>	52.5 %	47.3 %	0.191 %	Positiva
<b>Microsoft</b>	20 %	78.3 %	1.70 %	Neutral
<b>Tesla</b>	13 %	80 %	2.82 %	Neutral
<b>Amazon</b>	24 %	72 %	4 %	Neutral

TABLA 10: Categoría y puntuación de las compañías

De igual manera, también se genera una nube de palabras como se muestra en la figura 21 para observar aquellas palabras más texteadas que se encuentran en los mensajes publicados por los



# Conclusiones

Como se mostró en la figura 3 a la figura 8 se pudieron conocer algunas bases teóricas que están detrás del proceso del análisis de sentimientos además de maneras diferentes para evaluar y clasificar, mediante un sistema automatizado, los textos cortos según su polaridad. El análisis de sentimiento trata la información con el objetivo de observar la comunicación que se transmite en las redes.

El análisis de sentimientos depende de la construcción sintáctica del mensaje y su contenido semántico como el que se observa en la figura 17. El preprocesamiento de la información es fundamental para captar el nivel de sentimiento de la oración o palabra contenida en un texto. Las extrañas formas de escribir, la gran cantidad de símbolos y la falta de ortografía pueden representar que los resultados sean poco precisos.

Las palabras pueden tener diferentes significados respectivamente del contexto. Una palabra puede tener un sentimiento positivos o negativo dependientemente de la forma en como se diga o se escriba. Esto resulta un problema para los algoritmos de automatización porque dificultan la identificación del sentimiento que conlleva a proponer nuevas soluciones. En la tabla 9, la palabra "Twitch" tiene un sentimiento muy positivo pero debería ser una palabra neutra, esto se debe a que el clasificador no tiene claro que se trata de una plataforma y confunde su significado.

Las estrategias o metodologías pueden ser de gran ayuda para calificar, a partir de las opiniones del público, la actualidad de una marca. En la tabla 8 se puede conocer que la marca de videojuegos Playstation es la favorita con respecto a la otra marca que es Xbox.

Estos valores de sentimientos pueden ser de gran ayuda para conocer y porque no tratar esas acciones o motivos por los cuales se publico el mensaje y quien lo hizo. Sin embargo, queda claro que el sentimiento puede ser manipulado por como las personas ven un producto, servicio o cualquier otro tema a partir del spam.

# Bibliografía

- [1] Yi Min Shum. <https://yiminshum.com/twitter-2021-digital/>. 10 de Febrero de 2021.
- [2] Edgar Castro Monge. «Las estrategias competitivas y su importancia en la buena gestión de las empresas». En: *Revista de Ciencias Económicas* 28.1 (2010).
- [3] Alexandre Fonseca. *Marketing Digital en Redes Sociales: Lo imprescindible en Marketing Online para tu empresa en las Redes Sociales*. Alexandre Fonseca Lacomba, 2014.
- [4] Lucas Montesinos Garcíea. «Análisis de sentimientos y predicción de eventos en Twitter». En: (2014).
- [5] Martián Becerra. «Análisis de sentimientos en Twitter: El bueno, el malo y el>». En: *XX Concurso de Trabajos Estudiantiles-JAIIO 46 (Córdoba, 2017)*. 2017.
- [6] Rafael Sánchez del Hoyo. «Análisis de sentimientos con Twitter: turismo y política electoral». En: (2019).
- [7] Gabriel Valerio Ureña y River Serna Valdivi. «Redes sociales y bienestar psicológico del estudiante universitario». En: *Revista electrónica de investigación educativa* 20.3 (2018), págs. 19-28.
- [8] Meylan Wongkar y Apriandy Angdresy. «Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter». En: *2019 Fourth International Conference on Informatics and Computing (ICIC)*. 2019, págs. 1-5.
- [9] Lokesh Mandloi y Ruchi Patel. «Twitter Sentiments Analysis Using Machine Learning Methods». En: *2020 International Conference for Emerging Technology (INCET)*. 2020, págs. 1-5.
- [10] V. Prakruthi, D. Sindhu y Dr. S. Anupama Kumar. «Real Time Sentiment Analysis Of Twitter Posts». En: *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*. 2018, págs. 29-34.
- [11] Chae Won Park y Dae Ryong Seo. «Sentiment analysis of Twitter corpus related to artificial intelligence assistants». En: *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*. 2018, págs. 495-498.

- 
- [12] Sanjeev Dhawan, Kulvinder Singh y Priyanka Chauhan. «Sentiment Analysis of Twitter Data in Online Social Network». En: *2019 5th International Conference on Signal Processing, Computing and Control (ISPCC)*. 2019, págs. 255-259.
- [13] Kusurini y Mochamad Mashuri. «Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication». En: *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*. 2019, págs. 365-368.
- [14] Rasika Wagh y Payal Punde. «Survey on Sentiment Analysis using Twitter Dataset». En: *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2018, págs. 208-211.
- [15] Sahar A. El Rahman, Feddah Alhumaidi AlOtaibi y Wejdan Abdullah AlShehri. «Sentiment Analysis of Twitter Data». En: *2019 International Conference on Computer and Information Sciences (ICCIS)*. 2019, págs. 1-4.
- [16] M. Trupthi, Suresh Pabboju y G. Narasimha. «Sentiment Analysis on Twitter Using Streaming API». En: *2017 IEEE 7th International Advance Computing Conference (IACC)*. 2017, págs. 915-919.
- [17] Dionysis Goularas y Sani Kamis. «Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data». En: *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*. 2019, págs. 12-17.
- [18] Lei Wang, Jianwei Niu y Shui Yu. «SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis». En: *IEEE Transactions on Knowledge and Data Engineering* 32.10 (2020), págs. 2026-2039.
- [19] Md. Rakibul Hasan, Maisha Maliha y M. Arifuzzaman. «Sentiment Analysis with NLP on Twitter Data». En: *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. 2019, págs. 1-4.
- [20] Osvaldo Simeone. «A Very Brief Introduction to Machine Learning With Applications to Communication Systems». En: *IEEE Transactions on Cognitive Communications and Networking* 4.4 (2018), págs. 648-664.
- [21] Roberto A Manjarrés-Betancur y Monica Mariéa Echeverri-Torres. «Asistente virtual académico utilizando tecnologías cognitivas de procesamiento de lenguaje natural». En: *Revista Politécnica* 16.31 (2020), págs. 85-96.
- [22] Diego Leonardo Pedro. «Análisis de sentimientos: aplicación sobre textos en redes sociales». En: (2019).
- [23] Monserrat Ramirez Garcia, Maya Carrillo Ruiz y Abraham Sánchez López. «Combinación de clasificadores para el análisis de sentimientos». En: ()

- 
- [24] Natalia Niño Martínez y col. «CAPÍTULO V MINERÍA DE TEXTOS Y ANÁLISIS DE REDES SOCIALES EN TWITTER». En: ().
- [25] Ane Arburua Sagastibeltza. «Extracción de conocimiento operativo a partir de descripciones en lenguaje natural». En: (2020).
- [26] Lifeng Li y Wenxing Li. «Naive Bayesian automatic classification of railway service complaint text based on eigenvalue extraction». En: *Tehnički vjesnik* 26.3 (2019), págs. 778-785.
- [27] Cinthya Vergara y col. «Análisis de texto para Gestión de Marca con el uso de Instagram aplicado a una comunidad de mascotas en Chile». En: *Revista Ingenieria de Sistemas Volumen XXXIII* (2019).
- [28] Víctor Rosas. <https://www.levelup.com/noticias/585679/PlayStation-es-la-marca-mas-reconocida-por-los-millennials-y-la-generacion-Z>. 4 de Agosto de 2020.