

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**FACULTAD DE CIENCIAS**

**Inteligencia Artificial**

**PROYECTO 3:**

**Marco Teórico “Analizador de sentimientos para  
redes sociales”**

**PROFESOR: Estefania Prieto Larios**

**AYUDANTES:**

- **Andrea Fernanda Muñiz Patiño**
- **Edwin Antonio Galván Gámez**
- **Rafael Robles Ríos**

**EQUIPO: Comic-OS**

**INTEGRANTES:**

- **González Lara Malinali**
- **Ortiz Medrano Francisco Javier**
- **Serrato Solano Victor Manuel**

## MARCO TEÓRICO

Pensar en la implementación de un analizador de sentimientos para redes sociales como una necesidad es algo que hace algunos años no hubiese ocurrido, esto debido a que las redes sociales que hoy conocemos apenas comenzaban a crecer y a volverse del dominio público, en aquellos momentos no muchos lograron ver en lo que se convertirían a futuro pues redes como Facebook o Twitter tenían apenas unos años de haberse creado. Muchas personas aún se encontraban renuentes a utilizarlas y aseguraban que solo era algo que usaban los “jóvenes” y parecía ser que era un lugar en el que los adultos no deseaban ingresar.

Hoy con el paso de los años, descubrimos que la concepción que se tenía sobre estas redes ha cambiado considerablemente y como prueba tenemos los siguientes datos:

- Hay 3.800 millones de usuarios de redes sociales, número que aumentó más del 9% desde 2019.<sup>[2]</sup>
- En 2019, 90,4% de millennials, 77,5% de la generación X y 48,2% de *baby boomers* eran usuarios activos de las redes sociales.<sup>[2]</sup>
- La mayor audiencia de Twitter está conformada por personas entre los 18 y 49 años de edad.<sup>[2]</sup>
- El 81% de los menores de 24 años en Latinoamérica están conectados a Facebook.<sup>[2]</sup>

Con estas estadísticas es fácil comprender porqué las redes sociales se han convertido en un punto estratégico para implementar planes de todo tipo: marketing, políticos y un largo etcétera. Lo que nos lleva a el problema ¿por qué surge la necesidad de implementar un sistema de este tipo?

En la guerra el bando que tenga más información de su oponente tendrá una clara ventaja contra su enemigo; cada gran estrategia que ha existido ha tenido claro que la información era su recurso máspreciado pues el tenerla les permitía adecuar su plan de batalla. Actualmente las empresas, políticos, figuras públicas viven en una guerra constante contra sus competidores, Pepsi vs Coca-Cola, PRI vs PAN, Team King Kong vs Team Godzilla,

Al conocer las predilecciones de los usuarios que siguen cierta campaña (política o publicitaria) podemos entender que es conveniente ofrecerles, así que ahora tenemos una nueva pregunta ¿de qué forma podemos analizar toda esa cantidad de datos generada por millones de usuarios para convertirla en información que ayude a tomar decisiones de mejora?

Una forma de realizar dicho análisis es mediante un analizador de sentimientos como el que proponemos y desarrollamos en este proyecto. Si logramos detectar el sentimiento que impera en el usuario al publicar sus opiniones mediante posts o tweets (según sea el caso) podemos detectar que tipo de contenido genera descontento en una marca, una figura pública o un evento y con estos datos obtenidos modificar la estrategia para implementar mejoras en cualquier campo. Con la información que consigue la empresa a través de programas como el nuestro puede conocer las emociones de sus usuarios hacia su productos,

## ANTECEDENTES

Este problema ha sido analizado de muchas formas, incluso existen herramientas previas que ya nos brindan la oportunidad de realizar dicho análisis, por ejemplo, empresas como Google e IBM han puesto a disposición del público en general plataformas de inteligencia artificial que procesan el lenguaje natural que permiten realizar las siguientes tareas<sup>[3]</sup>:

- Extraer información sobre la receptividad de los productos (o la experiencia de los usuarios) a partir de las conversaciones que se tengan con los clientes (a través de correos electrónicos, chats o redes sociales) mediante la detección de entidades y el análisis de opiniones.
- Clasificar contenido y crear gráficas a partir de relaciones creadas con esta información por medio de su sistema de análisis sintáctico para filtrar contenido inapropiado y clasificar por temas.
- Extraer entidades y conocer las opiniones de los usuarios.
- Identificar temas principales, crear palabras clave e identificar conceptos.

Ahora conozcamos un poco acerca de ambas plataformas:

### **Google Cloud Natural Language.**<sup>[3]</sup>

Utiliza una API que permite conocer fácilmente la estructura y el significado del texto mediante modelos basados en machine learning. Es posible utilizarla para recopilar información acerca de personas, lugares, eventos y otros elementos en los que se mencione documentos de texto, noticias, entradas a blogs, foros, etc.

También puede usarse para conocer las opiniones de un producto, situación o persona en redes sociales, así como para analizar las intenciones de los clientes a partir de conversaciones, también es posible integrar esta función en el almacenamiento de documentos de Google Storage.

La API funciona de manera automática y basta con copiar y pegar el texto que se desea analizar, y pulsar el botón “Analyze”. Una vez que el texto es procesado la API es capaz de extraer Entidades, analizar sentimientos e identificar categorías.

Pese a que esta API cuenta con un periodo de prueba gratuito al hacer el registro fue necesario ingresar una TC y, al no contar con ella no pudimos ejecutar nuestro propio ejemplo.

### **IBM Watson Natural Language Understanding**<sup>[4]</sup>

Se trata de una plataforma de IA que esta basada en el procesamiento del lenguaje natural al introducir texto o una URL (según sea el caso) se produce un resultado que muestra 7 opciones con resultados:

TextURL

In the rugged Colorado Desert of California, there lies buried a treasure ship sailed there hundreds of years ago by either Viking or Spanish explorers. Some say this is legend; others insist it is fact. A few have even claimed to have seen the ship, its wooden remains poking through the sand like the skeleton of a prehistoric beast. Among those who say they've come close to the ship is small-town librarian Myrtle Botts. In 1933, she was hiking with her husband in the Anza-Borrego Desert, not far from the border with Mexico. It was early March, so the desert would have been in bloom, its washed-out yellows and grays beaten back by the riotous invasion of wildflowers. Those wildflowers were what brought the Bottses to the desert, and they ended up near a tiny settlement called Agua Caliente. Surrounding place names reflected the

English

For results unique to your business needs consider building a [custom model](#).

\* This system is for demonstration purposes only and is not intended to process Personal Data. No Personal Data is to be entered into this system as it may not have the necessary controls in place to meet the requirements of the General Data Protection Regulation (EU) 2016/679.

By using this application, you agree to the [Terms of Use](#)

Analyze

SentimentEmotionKeywordsEntitiesCategoriesConceptSyntax

Semantic Roles

Start for free in IBM Cloud

Imagen1. Texto a ser procesado.

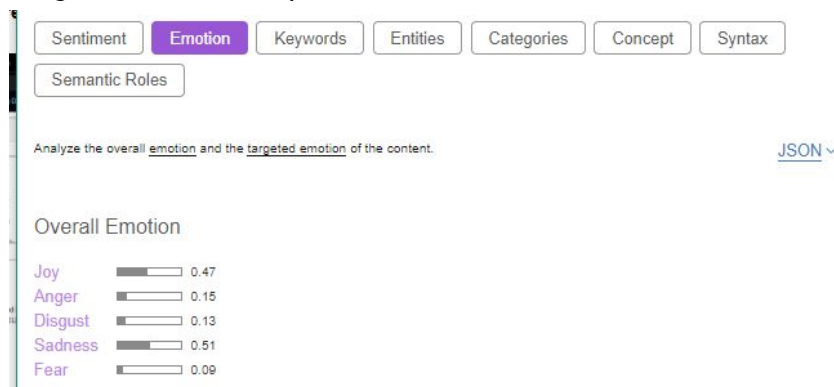


Imagen2. Opciones mostradas tras efectuar el análisis.

De lo anterior podemos deducir que un analizador de sentimientos puede aplicarse a básicamente cualquier texto, pero en nuestro caso hemos decidido centrarnos en un tema en particular: analizar emociones en tweets, por lo que nuestro analizador está pensado para la red social Twitter.

## INVESTIGACIONES PREVIAS

Ya hemos revisado algunas de las ventajas de contar con un analizador de sentimientos para texto o para redes sociales pero... ¿por qué hacerlo para Twitter y no para otra red social?

Revisamos algunas estadísticas que nos ayudaron a tomar esta decisión y así llegar a la conclusión de que es importante conocer la influencia de las publicaciones que se realizan en esta red:

- Twitter tiene más de 340 millones de usuarios actualmente.<sup>[2]</sup>
- De los cuales, cerca de 139 millones son usuarios activos diarios que pueden ser monetizables.<sup>[2]</sup>
- El 83% de las personas que enviaron un Tweet a una empresa y recibieron una respuesta se sintieron mejor acerca de ella y tenían más probabilidades de hacer negocios con ellos.<sup>[2]</sup>
- Las personas tienen un 31% más de probabilidades de recordar lo que vieron en Twitter que en la navegación en línea en general.<sup>[2]</sup>
- Al 79% de los usuarios de Twitter les gusta descubrir las novedades, por lo que es la plataforma número uno de descubrimiento.<sup>[2]</sup>

Twitter es el lugar perfecto donde muchas personas dan su opinión sobre algún tema, por ejemplo, revisemos el caso “Pepsi vs Coca”, si alguno de ellos saca un producto nuevo, la empresa necesitará saber cómo fue recibido su nuevo producto por el público, si les gusta o si no les gusta, todas estas opiniones pueden ser captadas desde un tweet en Twitter, donde los consumidores de este producto comparten su opinión sobre el producto.

Para cualquier empresa, persona u organización interesada en hacer uso de esta plataforma como un punto de venta y/o promoción de sus servicios debe de interesarle conocer como sus potenciales clientes reaccionan a temas o productos similares, pues esto podría favorecer a sus intereses.

Incluso, si pensamos en cuestiones políticas (ej. El cambio de presidente en EE.UU. o AMLO enfermo de COVID-19) es importante para determinado grupo de personas conocer las opiniones tanto de quienes les apoyan como de quienes son sus detractores. Lo mismo se aplica para temas de cuestiones de salud, inseguridad, etc

Durante nuestra búsqueda encontramos que ya existen algunos analizadores de sentimientos desarrollados para Twitter, por ejemplo:

- *Sentiment140*.<sup>[5]</sup>
- *Peoplebrowsr*. Proporciona un análisis acerca de marcas que la gente consume.<sup>[6]</sup>
- *Pulse of the nation*. Mide el estado de ánimo a lo largo del tiempo de todo EE.UU y lo muestra mediante una animación.<sup>[7]</sup>

No debemos de olvidar el punto de la privacidad pues recordemos que en Diciembre de 2018 Twitter se vio envuelto en un problema en el que “accidentalmente” filtró datos personales de sus usuarios.<sup>[8]</sup> Al tratar con información publicada originalmente por otras personas desde “la seguridad” de sus computadoras debemos de tener cuidado con lo que mostramos, es por esto que solo analizaremos tweets que tengan un carácter público, es decir, tweets que hayan sido publicados por aquellos usuarios que en sus “Configuraciones de privacidad” tengan establecido el valor como “público”, lo cual implica que todo el mundo (con un cuenta de twitter) pueden visualizar estos sin ninguna restricción. De esta forma evitamos complicaciones legales de cualquier índole que pudieran generarse al tratar con información privada.

## CONSIDERACIONES TEÓRICAS

- Nuestro trabajo consiste en crear un analizador de sentimientos para tweets públicos utilizando la API de Twitter, nuestra principal herramienta de trabajo es justo esta API. El lenguaje utilizado será Python, además de la biblioteca “spacy”<sup>[10]</sup>.
- Con la biblioteca spacy podemos resolver problemas que de otra forma tendríamos que resolver más tarde al tener que realizar todo el trabajo desde cero.
- El análisis está centrado en algunos de los tweets publicados en esta plataforma, en particular buscamos analizar aquellos tweets que tengan algún contenido relacionado con COVID pues además de ser un tema en tendencia que nos afecta a todos, Twitter ha sido una plataforma que muchos usuarios han utilizado para mantenerse informados, e incluso compartir su estado de salud en caso de tener sospecha o un caso confirmado.
- Para poder efectuar dicho análisis necesitamos contar con los tweets, es por esto que se requiere de acceso a la API de Twitter.
- Tenemos acceso a 500 mil tweets de forma gratuita.
- Debido a la premura del tiempo para completar este proyecto no nos será posible crear un dataset en español para el entrenamiento, por lo que utilizaremos uno ya existente en inglés.
- Al utilizar un dataset en inglés solo podrá realizar el análisis de emociones en textos (tweets) en inglés.
- De forma adicional trataremos de desarrollar la implementación también en español.
- Los analizadores existentes pueden analizar tweets para obtener cosas como el humor del país o búsquedas realizadas, nuestra propuesta consiste en reconocer “lo que las personas sienten” cuando hablan de COVID.
- La salida del programa se presentará solo en terminal.

## METODOLOGÍA

El primer paso (y el más importante de todo este desarrollo) fue conseguir acceso a la API de Twitter pues toda nuestra propuesta esta basada en poder trabajar con dicha API y justamente por esa razón fue que los tres integrantes del equipo nos dispusimos a solicitar una cuenta de desarrollador a través de la página para desarrolladores de Twitter<sup>[9]</sup>. El primer intento fue rechazado:



Por fortuna para todos el siguiente intento resultó ser exitoso, a continuación se muestra el proceso que tuvimos que seguir en un primer filtro, desde la página para el registro respondimos las siguientes preguntas:

**Twitter:** How will you use the Twitter API or Twitter Data? In your words.

In english, please describe how you plan to use Twitter data and/or API's. The more detailed the response the easier it is to review and approve.

**Comic-OS:** I'm in my last semester at the University and I study the degree on Computer Sciences at UNAM. I'm taking the AI class and for my final project I proposed to make an implementation of "Sentiment Analysis for conversations on Social Networks". Obviously this analyzer will be pretty basic. I'm interested in covering just one topic (something like COVID) and try to know how people feel while they are writing their posts about this particular topic.

*NOTA: La respuesta se realizó en singular pues solicitamos una cuenta de desarrollador para una sola persona, no para un equipo y si respondíamos en plural corríamos el riesgo de "levantar sospechas" y que una vez más nos rechazaran.*

**Twitter:** Are you planning to analyze Twitter data? Yes

**Comic-OS:** Only will make an analysis over the tweets from the people who has their privacy settings like "public" (all the world can see these tweets).

I will create a database where I'll put some words associated with negative or positive emotions and I give them a score. After with those tweets I'll make a comparison between the tweet and the database to determine which emotion the person had at the time of he/she wrote the post.

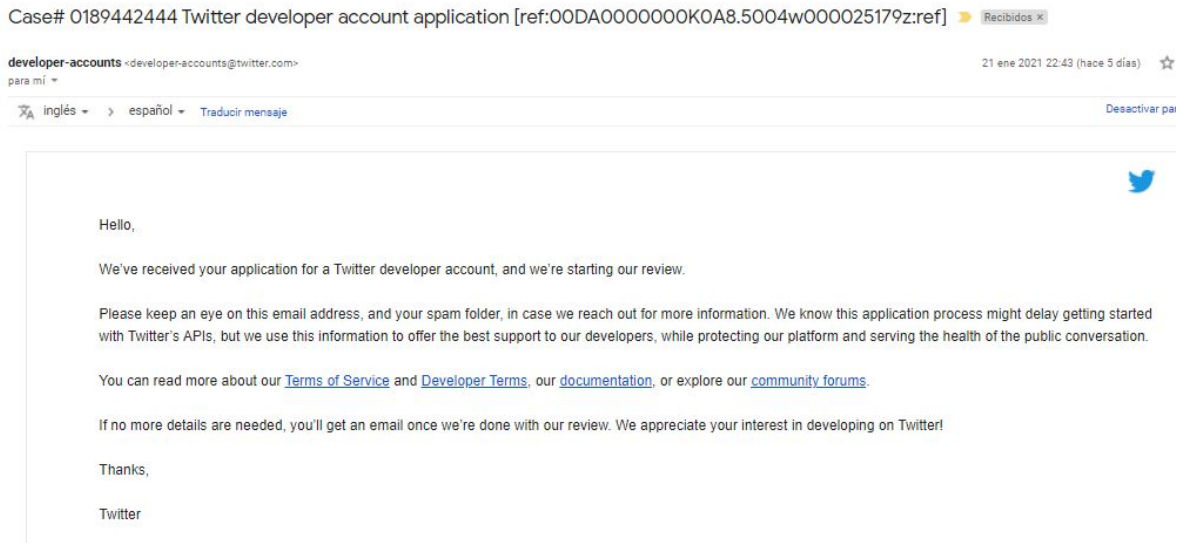
**Twitter:** Will your app use Tweet, Retweet, Like, Follow, or Direct Message functionality? No

**Twitter:** Do you plan to display Tweets or aggregate data about Twitter content outside Twitter? Yes

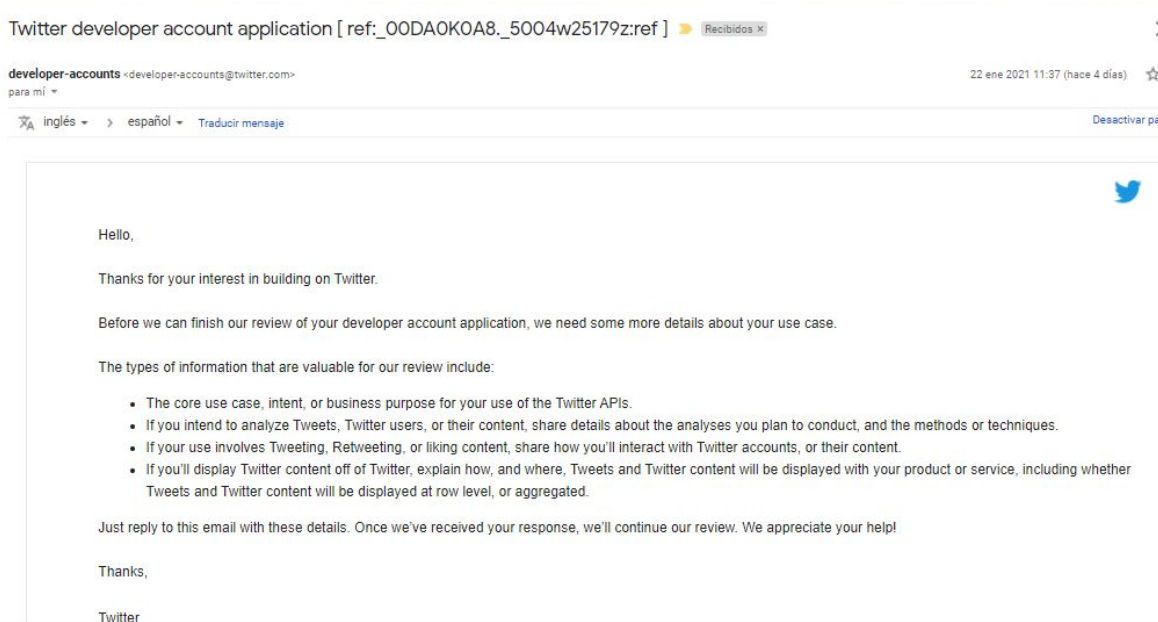
**Comic-OS:** I'll only need to show the tweet (after making the comparison) because I plan that the output of my program takes the tweet that was analyzed and the emotion that was detected and show them on a screen.

**Twitter:** Will your product, service, or analysis make Twitter content or derived information available to a government entity? No.

Después de realizar este registro recibimos el siguiente correo, en el que nos notificaron que nuestra solicitud ya estaba en revisión:



Al día siguiente, recibimos un nuevo correo, esta vez nos solicitaban más detalles de lo que sería nuestro trabajo:



Y aquí nuestra respuesta:



Twitter developer account application [ ref:\_00DA0K0A8.\_5004w25179z:ref ] Recibidos x

**developer-accounts** 22 ene 2021 11:37 (hace 4 días)

Hello, Thanks for your interest in building on Twitter. Before we can finish our review of your developer account application, we need some more details about y

---

**Malinali González Lara** <malinali\_glezl@ciencias.unam.mx>  
para developer-accounts

22 ene 2021 23:39 (hace 4 días)

Hello!

Here are the details requested:

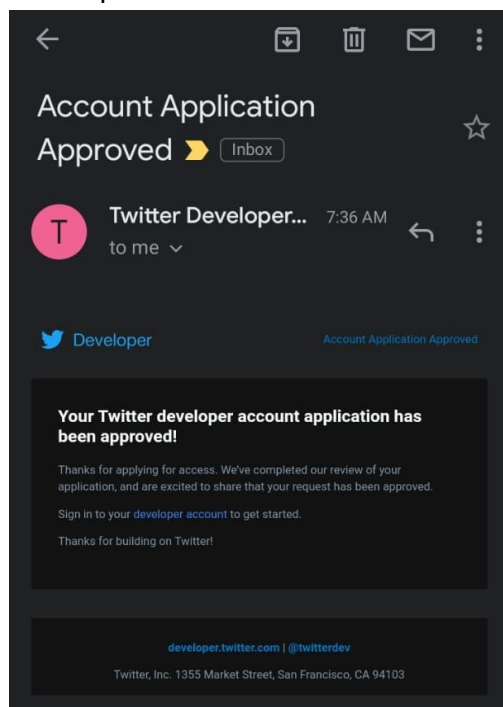
The project that I'm thinking to make will be just for academic use (it'll be my final project for the AI class, I hope); in theory only two people will view the project when I'll finish: my professor and I. The principal idea it's make an emotion analysis over some tweets (public tweets).

The only thing that I need to analyze from the tweets are the emotions this analysis will be made with NPL techniques (Natural Language Processing) normalizing, vectorizing and removing "stop words" from the text (in this case the text are the tweets). To do that implementation I'm going to use the functional language Python and a library called "spacy".

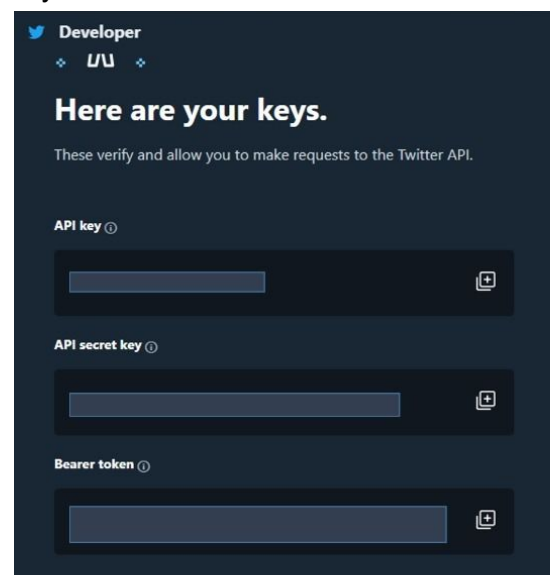
All the information that I'll obtain with the analysis will be locally stored, the way that I'm going to show the results is on a computer terminal because it's just a school project and I don't have enough time to make an interface (actually I only have 2 weeks) and the only important thing is that the program (the analyzer) works.

Thanks.

Finalmente obtuvimos la respuesta que mas esperabamos:



Como último paso para comenzar de verdad con nuestro trabajo fue necesario loguearse en la cuenta de desarrollador que nos había aceptado y así obtener las keys:



Ahora con el acceso a la API es posible comenzar con nuestro desarrollo.

Para comenzar es necesario limpiar el texto lo más posible para hacerlo "legible" para la computadora, esto implica quitar palabras sin contexto, signos de puntuación que no sirven, palabras como banderas o las personas arrobadas, etc.

Con el primer proceso de "limpieza" realizado tenemos que tokenizar para volverlo aún más manejable para la computadora, sin embargo, este proceso no es suficiente pues podemos encontrarnos con muchas palabras que tengan el mismo significado, por ejemplo: "computadora" y "computadoras", para esto se debe de aplicar una normalización que nos llevaría a su raíz "computador".

Tras haber realizado estos procesos ya tenemos algo legible para la computadora pero aún falta realizar la parte fundamental de todo: tomar el texto y predecir su sentimiento, esto lo haremos con clasificación de aprendizaje automático.

Existen varias herramientas de aprendizaje automático, los más utilizados son TensorFlow, PyTorch y Scikit-learn, aunque son muy famosos y muy potentes tienen una curva un poco pronunciada pues que son herramientas muy complejas, por esto mismo utilizaremos la que nos proporciona spaCy pues es más sencillo de utilizar y con menos dependencias.

Para crear nuestro analizador de sentimientos de lenguaje natural seguimos 4 pasos:

1. Carga de datos.
2. Preprocesamiento de datos.
3. Entrenamiento del clasificador.
4. Clasificación de datos.

#### 1. Carga de datos.

En este punto influyen muchas cosas pues se trata de la base de nuestro analizador, debemos tener datos de entrenamiento, etiquetas de datos (palabras que no queremos analizar) y un conjunto de datos de prueba.

#### 2. Preprocesamiento de datos.

Relacionado con el punto anterior, pues cada que se realiza la carga de datos es necesario procesarlos de tal forma que se le de estructura, además de parametrizar todos los datos que tenemos.

Posteriormente debemos mezclar los datos, aquí podemos tener sesgo sobre los datos al cargarlos pero una mezcla (random) nos ayudaría bastante a reducirlo. Finalmente debemos dividir los datos en dos partes, una de entrenamiento y otra de clasificación.

#### 3. Entrenamiento de clasificador.

Intentaremos utilizar una red neuronal de convolución para clasificar nuestro texto; ya que tenemos previamente etiquetado nuestro texto y contamos con datos de entrenamiento comenzamos con la canalización del texto, es en este paso donde agregaremos las etiquetas de “positivo” y “negativo”. Para esto nos apoyaremos de spaCy con textcat para tener nuestro ciclo de entrenamiento. Es necesario construir un ciclo de entrenamiento para entrenar textcat, esto se hace mediante lotes de datos, donde cada lote es un subconjunto de nuestro texto (para que no sea tan complejo procesarlos). Por último se debe realizar una evaluación de la formación del modelo. SpaCy tiene la función `evaluate_model()` que nos calcula los siguientes valores:

- Los verdaderos positivos: quiere decir que se clasifican como positivos y son correctos.

- Los falsos positivos: son los positivos predecidos incorrectamente que, de hecho, fueron negativos.
- Negativos verdaderos: son los que se clasifican como negativos y son correctos
- Falsos negativos: los que se calcularon incorrectamente como negativos pero en realidad son positivos.

Utilizamos una medida de puntuación llamada F-score, mide la precisión de nuestro modelo, se encuentra en el intervalo entre 0 y 1, 0 es el más bajo y 1 el más alto.

#### 4. Clasificación de los datos.

Con el entrenamiento completado es necesario probarlo, esto se realizará con el conjunto de datos disponible. De acuerdo a los datos proporcionados y el límite de estos tendremos una precisión, pero entre más datos más tiempo y esto mejorará nuestro F-score.

### ENFOQUE DE LA INTELIGENCIA ARTIFICIAL

El enfoque de la inteligencia artificial que hemos tomado para nuestro proyecto es el de un sistema que piensa racionalmente<sup>[13]</sup>. Como principales obstáculos nos encontramos con el hecho de que no es simple transformar el conocimiento formal como lo son los tweets a los términos formales que necesitamos para procesarlos, además si el conocimiento con el que se entrena al agente es limitado es muy seguro que las respuestas obtenidas sean incorrectas. Por último pero no menos importante podemos encontrarnos con problemas en los que se requiera de muchos recursos para procesar todos estos datos como conocimiento por lo que el obtener un buen entrenamiento para el analizador de sentimientos puede demorar horas.

### TIPO DE AGENTE, REAS, ENTORNO DE TRABAJO Y PROPIEDADES

Proponemos el uso de un agente basado en *objetivos*, un agente de este tipo primero aprende con el conjunto de entrenamiento que se le pasa, en este caso el conjunto brinda ejemplos de Tweets que han sido previamente etiquetados como positivos/negativos. Una vez que ha “asimilado” su entrenamiento es posible pasarle nuevos datos que no estén presentes en el conjunto que se usó para el entrenamiento, es decir, podremos pasar Tweets que no han estado en el conjunto de entrenamiento. Para este punto el agente ha aprendido a diferenciar un tweet positivo de uno negativo gracias a su entrenamiento y con base en dicho aprendizaje podrá tomar una decisión para etiquetar este nuevo tweet y así cumplir su objetivo.

Mientras más grande sea el conjunto de entrenamiento existen más probabilidades de que la etiqueta brindada al nuevo tweet sea correcta.

## Entorno de trabajo y sus propiedades

- *Medidas de rendimiento:* El rendimiento que nuestro analizador de sentimientos puede tener está directamente relacionado con el tamaño del conjunto de entrenamiento, por lo cual este debe de ser robusto, de esta forma nuestra red neuronal puede aprender a diferenciar entre los dos escenarios que se pueden presentar (si un Tweet se considera negativo o positivo) y que posteriormente sea capaz de determinar si el Tweet que se le pasa como entrada es positivo o negativo para finalmente saber si esta respuesta obtenida es acertada en un porcentaje significativo. Obtener un 100% de acierto en este tipo de programas podría considerarse básicamente imposible, pues nosotros como humanos no empleamos un lenguaje formal, incluso en libros podemos encontrarnos con textos difíciles de comprender, ahora, tratándose de una plataforma tan informal tal como lo es una red social (Twitter) es común el uso de modismos, jerga regional, signos de puntuación para formar gestos :( :D ;), además del sarcasmo, ironía y un gran etcétera; todo esto altera en gran medida el aprendizaje del agente. Aún cuando nosotros dejemos “limpio” el tweet a analizar con procesos de tokenización o vectorización.
- *Entorno:* El entorno principal de nuestro agente (al ser pensado en el contexto de la pandemia) puede ser útil para conocer la situación general de cierta población y determinar si la enfermedad ha aumentado (muchos tweets con comentarios negativos) o decrecido (muchos tweets con emociones positivas), sin embargo, en general es variable pues puede aplicarse a múltiples campos, dependerá de quién use este, pues tal como se menciona en la justificación: el agente puede ser utilizado por una empresa que busque incrementar las ventas de sus productos y apoyarse de nuestro programa para conocer el sentimiento de los clientes hacia sus productos y con esto tomar una decisión informada sobre si es conveniente aumentar la producción (recibe muchos comentarios positivos) o si por el contrario es mejor bajar o eliminar el producto para evitar pérdidas
- *Actuadores:* El actuador del agente serían los Tweets, pues el programa se encarga de regresar una respuesta sobre el Tweet que recibe como entrada, para llegar a esta respuesta se usa el entrenamiento que también requiere de Tweets etiquetados como positivos y negativos. Como nuestro programa requiere en todo momento de los tweets estos pasan a convertirse en el actuador.
- *Sensores:* Para los sensores se requiere un teclado para ingresar un Tweet, el cual sería la entrada del programa, se requiere además de una pantalla para visualizar los datos, tanto los que se ingresan como la respuesta del agente.

\*El entorno es parcialmente observable<sup>[13]</sup> pues si bien la pantalla (sensor) recibe información sobre la parte que se está ejecutando para la limpieza del texto y

finalmente otorga una salida, ningún sensor recibe información sobre todos y cada uno de los pasos que se ejecutan para obtener dicha respuesta.

\*Es determinista<sup>[13]</sup> pues pese a ser parcialmente observable sabemos que dada la entrada de un tweet obtendremos la emoción del tweet, independientemente de la etiqueta que reciba el tweet no esperamos obtener ningún otro resultado que no sea una emoción.

\*Es secuencial<sup>[13]</sup> pues la acción presente afecta a la futura, es decir no podemos pasar al segundo proceso de limpieza si no hemos pasado por el primero y de la misma forma no podemos obtener la emoción del tweet si antes el texto no fue preparado de la manera correcta. Incluso si un error ocurriera en algún punto no podríamos continuar y obtener la emoción por lo que las acciones presentes afectan a las futuras.

\*Es estático<sup>[13]</sup>. El entorno del agente no cambiará nunca mientras este está realizando sus funciones. Una vez que el agente ha recibido su entrenamiento no debe verificar nada más.

\*Es discreto<sup>[13]</sup> pues tiene un número finito de estados: pasa de un proceso de limpieza a otro, pasa de una entrada a una salida. No tiene múltiples variables/factores que pueden generar un número infinito de estados.

\*Es un multiagente<sup>[13]</sup> pues a pesar de que todo el trabajo lo realiza solo, con ayuda de herramientas, requiere comunicarse con otro agente para obtener su respuesta, este “agente extra” es el propio usuario humano.

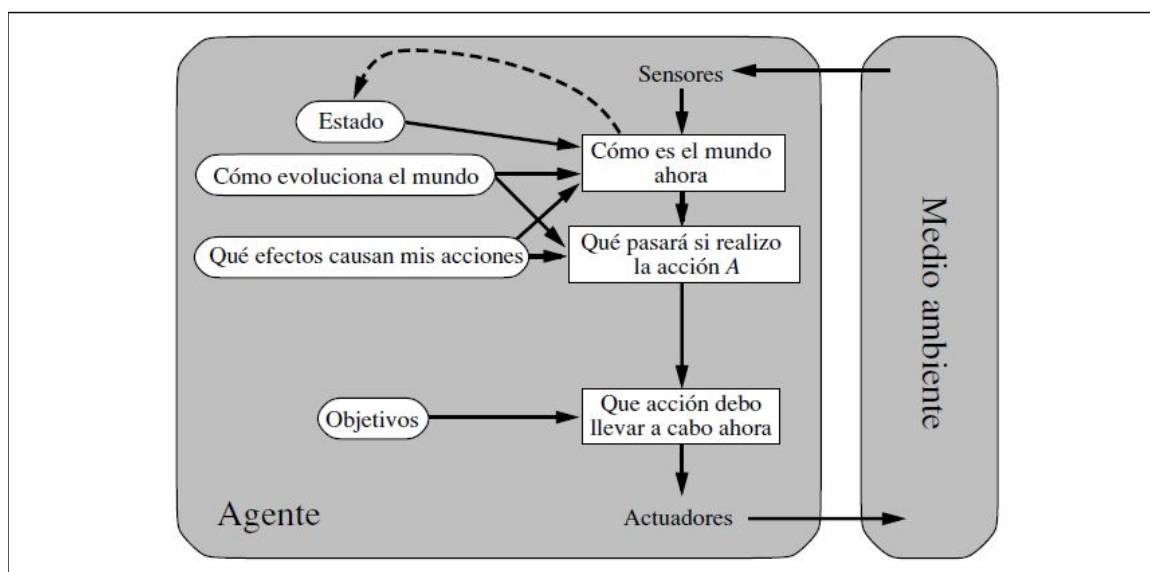
## **COMPORTAMIENTO DEL AGENTE RESPECTO A OBJETIVOS**

Tras terminar por completo la implementación de nuestro analizador consideramos que nuestro agente propuesto resultó efectivo para ayudarnos a cumplir los objetivos propuestos en etapas tempranas del desarrollo de este proyecto pues logramos obtener el sentimiento de un tweet como una emoción negativa o positiva, además, al ser un agente que aprende mediante un entrenamiento podría ser relativamente sencillo modificar el conjunto de entrenamiento y con ello el tipo de texto que se puede analizar y obtener su emoción.

## **ESQUEMA DE REPRESENTACIÓN DEL CONOCIMIENTO**

Al hacer uso de muchas herramientas tales como spaCy o la propia API de Twitter contamos con una parte del trabajo resuelto.

A continuación se presenta un modelo de un agente basado en objetivos (y modelos):<sup>[13]</sup>



**Figura 2.13** Un agente basado en objetivos y basado en modelos, que almacena información del estado del mundo así como del conjunto de objetivos que intenta alcanzar, y que es capaz de seleccionar la acción que eventualmente lo guiará hacia la consecución de sus objetivos.

En nuestro caso introducimos y vemos los tweets y la respuesta generada con ayuda de los sensores, el estado actual del mundo está dado por el conjunto de datos de entrenamiento que no son sino tweets catalogados como positivos y negativos con anterioridad.

El objetivo en general es uno: obtener la etiqueta (respuesta) del nuevo tweet, sin embargo, aquí podemos tener dos vertientes y ambas son una respuesta válida (negativo y positivo).

En el caso de un agente basado en objetivos como este tenemos que son más flexibles pues el conocimiento que soporta su decisión está presentado explícitamente, además esta base de conocimiento puede modificarse al cargar un nuevo entrenamiento.

## CONCLUSIONES

Con el desarrollo de este proyecto estamos convencidos de una cosa: al hacer uso de las tecnologías que ya existen podemos desarrollar cosas muy interesantes, sin embargo, a veces es complicado lograr el acceso a estas tecnologías. En este caso fue un poco complicado pues Twitter al ser una empresa con relevancia a nivel mundial debe de cuidar los datos de sus usuarios y es claro que no puede confiárselos a cualquier persona de una forma tan fácil, por eso se debe de seguir un proceso de aplicación, aún así deben de existir procesos que sean más largos, detallados y complejos en general.

Al elegir esta forma para trabajar tuvimos algunas complicaciones: encontrar un dataset adecuado, querer hacer la implementación en español y llegar a la conclusión de que no tendríamos tiempo suficiente debido a la carga de trabajo de otras materias, además después del primer rechazo creímos que tal vez sería necesario cambiar el tema del proyecto, pues era probable que no nos dieran el acceso. Afortunadamente logramos el acceso y hemos solucionado los problemas a los que nos hemos enfrentado, quizá pudo haber una forma más fácil de desarrollar nuestro proyecto.

Igualmente aprendimos muchísimo, no solo al realizar la implementación sino también con toda la investigación que tuvo que realizarse y con el proceso de aplicación.

A continuación se encuentra la liga a nuestro repositorio de GitHub en el que se podrá encontrar el proyecto terminado así como las instrucciones necesarias para correrlo desde cero:

<https://github.com/inteligencia-artificial-2021-1/Comic-OS/tree/main/Proyecto3>

## BIBLIOGRAFÍA

- [1] Significados. Actualizado 2021. Marco Teórico. Recuperado el 25-Enero-2021 de: <https://www.significados.com/marco-teorico/>
- [2] Santos, Diego. 2020. HubSpot. 97 estadísticas sobre redes sociales en 2020. Recuperado el 25-Enero-2021 de: <https://blog.hubspot.es/marketing/estadisticas-redes-sociales>
- [3] Rosas, Rosana. (s.f.). Rosana Rosas. Análisis de sentimientos en redes sociales. Recuperado el 13 de enero de 2021 de: <https://rosanarosas.com/analisis-sentimiento-redes-sociales/>
- [4] IBM. Actualizado 2020. Natural Language Understanding. Recuperado el 26-Enero-2021 de : <https://natural-language-understanding-demo.ng.bluemix.net/>
- [5] Sentiment140. 2013. Sentiment140. Recuperado el 26-Enero-2021 de: <http://www.sentiment140.com/>
- [6] SocialOS. 2008. Peoplebrowsr. Recuperado el 26-Enero-2021 de: <http://brands.peoplebrowsr.com/>
- [7] Mislove et al. 2010. Pulse of the Nation: U.S. Mood Throughout the Day inferred from Twitter. Recuperado el 26-Enero-2021 de: <http://www.ccs.neu.edu/home/amislove/twittermood/>
- [8] Menegus, Bryan. 2018. Gizmodo. Twitter alerts some users to “Unusual” data leak. Recuperado el 26-Enero-2021 de: <https://gizmodo.com/twitter-alerts-some-users-to-unusual-data-leak-1831158698>
- [9] Twitter. 2021. Twitter Developer. Recuperado de: <https://developer.twitter.com/en>
- [10] spaCy. 2017. Industrial-strength Natural Language Processing. Recuperado el 22-Enero-2021 de: <https://spacy.io/>
- [11] Maas et al. (2011). Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011). Recuperado de: <http://www.aclweb.org/anthology/P11-1015>
- [12] Maas et al. (2011). Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011). Dataset. Recuperado de: <https://ai.stanford.edu/~amaas/data/sentiment>
- [13] Stuart Russell, Peter Norvig. 2004. Inteligencia artificial un enfoque moderno. 2da. Edición.