



Pontificia Universidad
Católica del Ecuador

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

ESCUELA DE SISTEMAS

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
MAGISTER EN DATA SCIENCE**

Análisis de sentimientos en redes sociales con aplicaciones en política ecuatoriana

AUTOR:

Miguel Tupiza

DIRECTOR:

Alfredo Calderón Serrano

Contenido

1. Introducción.....	4
1.1. Planteamiento del problema.....	4
1.2. Objetivos	4
1.3. Conceptos previos.....	4
1.4. Estado del arte	7
1.4.1. Democracia participativa	7
1.4.2. Situación actual del análisis de sentimientos en la política ecuatoriana.....	8
2. Marco teórico.....	9
2.1. Ciudadanía digital.....	9
2.2. Política 2.0.....	10
2.3. Análisis de redes sociales	11
3. Diseño de la propuesta	13
3.1. Antecedentes	13
3.2. Narración	13
3.3. Diagrama del proceso.....	13
4. Desarrollo y resultados	15
4.1. Obtención de los datos	15
4.2. Limpieza de los datos	16
4.3. Análisis de sentimientos	16
4.4. Visualización de resultados	16
4.5. Evaluación.....	17
5. Conclusiones y recomendaciones	20
5.1. Conclusiones	20
5.2. Recomendaciones	20
6. Bibliografía	21

Tabla de ilustraciones

Figura 1. Cuadro de diálogo para instalar NLTK.....	5
Figura 2. Diagrama de dispersión de la ocurrencia de una palabra.....	6
Figura 3. Frecuencia de la ocurrencia	7
Figura 4. Tema y encabezado de encuesta	8
Figura 5. Cuantos saben lo que es AS	
Figura 6. Cuantos usan tecnología en política.....	9
Figura 7. Ejemplo de bloqueo de comentario ofensivo (notigram.com).....	10
Figura 8. Proyección en millones de usuarios (insiderintelligence, 2022).....	12
Figura 9. Páginas web más visitadas (Hootsuite, 2022).....	12
Figura 10. Diagrama de flujo de la propuesta	14
Figura 11. Versión de API y nivel de privilegio.	15
Figura 12. Relación subjetividad-polaridad	17
Figura 13. Gráfico de barras sobre los resultados	17
Figura 14. Tweet obtenido de la página fuente	18
Figura 15. Diagrama de dispersión de @ZULNELEA	19
Figura 16. Gráfico de barras resultado AS.....	19
Figura 17. Tweet calificado como negativo	20

1. Introducción

1.1. Planteamiento del problema

La constante evolución y transformación de la realidad del mundo en el que vivimos trae consigo nuevas maneras de interactuar, informarnos y participar en casi todos los aspectos en los que antes lo hacíamos de manera restringida, por ejemplo, la política. La vía para informarnos solía ser la prensa, radio y televisión, limitando la fuente de información a espacios de horarios fijos, a márgenes de opiniones según las ideologías y empatías políticas de los medios de comunicación, la participación ciudadana se limitaba a las entrevistas con sus respectivas ediciones.

Varios avances tecnológicos han traído consigo reducir esas brechas y permitir a las personas incluirse un poco más en temas que son de su interés. La capacidad de informarse está muy cerca, se la puede hacer en cualquier momento con solo utilizar un smartphone, es posible también opinar y debatir sobre los temas que generan interés o preocupación. No obstante, era de esperarse que el público pudiese ser muy diverso y podría ocasionar todo tipo de comentarios y reacciones entre ellos los poco acertados u ofensivos. Tampoco se puede dejar de lado que se pueda recurrir a trucos digitales para el aumento de la popularidad y seguidores falsos según la plataforma de difusión de contenido.

Entonces se hace necesario combatir problemas nuevos con soluciones actuales, es por esto por lo que se plantea un modelo de participación ciudadana digital en asuntos de política nacional para identificar los posibles perfiles falsos y realizar un análisis de sentimientos de los participantes con el fin de tener una visión más realista de sus opiniones para hacerlas llegar a los interesados. Esto podría ayudar a un actor político a tener una realidad sobre la aceptación o no de sus electores para trabajar sobre aquello, manejándose todo sobre el marco adecuado de ética.

1.2. Objetivos

Encontrar un personaje en el área de la política que tenga interacción con la comunidad a la que pretende dirigir u ocupar un cargo público para que estos generen los datos necesarios que servirán para la producción del modelo utilizando herramientas tecnológicas que arrojarán los sentimientos de sus participantes.

Extraer esta información proveniente de la red social Twitter para su posterior análisis y demostrar la posibilidad de que sea factible utilizarla para producir conocimiento que apoye en la toma de decisiones en ciertos ámbitos.

1.3. Conceptos previos

- Procesamiento de lenguaje natural

Para propósitos del presente trabajo, se puede mencionar que existen dos tipos de lenguaje: el lenguaje natural y el lenguaje formal. El lenguaje natural es aquel que ha venido evolucionando con el paso del tiempo de forma natural y hemos usado para comunicarnos unos con otros, formas de este tipo de lenguaje son el escrito, el hablado y los signos. No considera reglas gramaticales aún, estas se desarrollan en el proceso. Por otro lado; el lenguaje formal constituye reglas definidas para expresar una idea. Un tipo de este lenguaje es el científico, el matemático y se los incluye también a los lenguajes de programación.

El procesamiento del lenguaje natural involucra a ambos; es decir, el procesamiento del lenguaje natural mediante el uso de lenguajes de programación donde su principal aplicación sería poder interactuar con las máquinas mediante el uso del lenguaje natural.

Procesamiento de lenguaje natural mediante Python

Es uno de los lenguajes de programación (formales) más poderosos y utilizados para diversas aplicaciones en la actualidad, uno de ellos (y para el fin propuesto) es el procesamiento de datos lingüísticos a través de su biblioteca de código abierto *Natural Language Toolkit* (NLTK).

Para instalar esta librería debemos ingresar estos comandos en el intérprete, en este caso Anaconda Jupyter

```
import nltk
nltk.download()
```

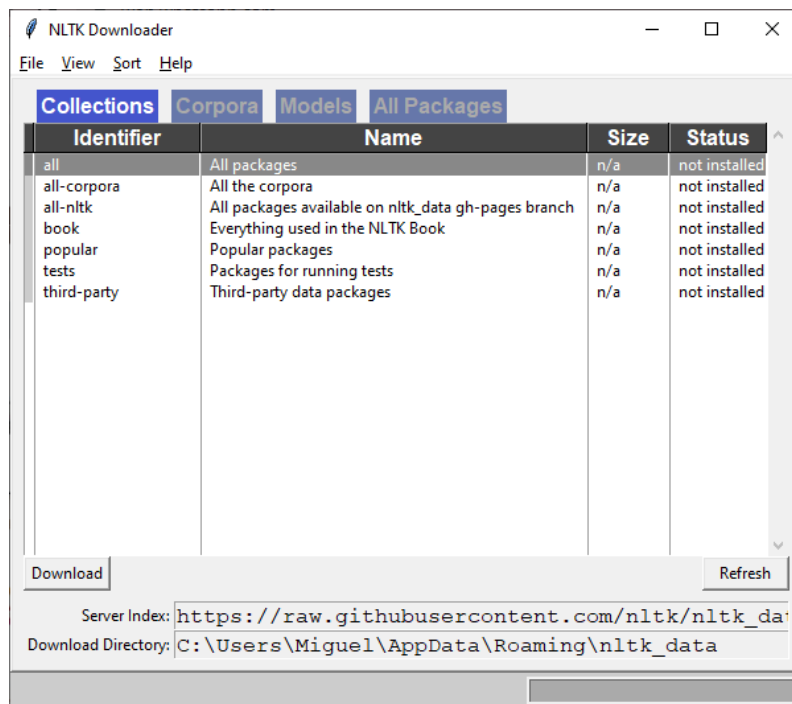


Figura 1. Cuadro de diálogo para instalar NLTK

Como demostrativo utilizaremos algunos ejemplos de comandos y utilidades que puede hacer esta librería, empezaremos cargando algunos textos de ejemplo mediante el comando

```
from nltk.book import *

*** Introductory Examples for the NLTK Book ***
Loading text1, ..., text9 and sent1, ..., sent9
Type the name of the text or sentence to view it.
Type: 'texts()' or 'sents()' to list the materials.
text1: Moby Dick by Herman Melville 1851
text2: Sense and Sensibility by Jane Austen 1811
text3: The Book of Genesis
text4: Inaugural Address Corpus
text5: Chat Corpus
text6: Monty Python and the Holy Grail
text7: Wall Street Journal
text8: Personals Corpus
text9: The Man Who Was Thursday by G . K . Chesterton 1908
```

Si queremos comprobar con qué frecuencia aparece una palabra dentro de un texto, por ejemplo, *horse* en *Moby Dick* escribimos

```
text1.concordance("horse")
```

Displaying 25 of 27 matches:

```
country also with a view of catching horse - whales , which had bon
upon broiled ibis and roasted river horse , that you see the mummi
, my neck feeling as if it were in a horse - collar ; and suddenly
in air , the beautiful and bountiful horse - chestnuts , candelabra
as the plungings of the Roman race - horse but so much the more str
es ; who didst hurl him upon a war - horse ; who didst thunder him
ds on legendary record of this noble horse , that it was his spirit
the evangelist , rides on his pallid horse . Therefore , in his oth
great hull through the water , as a horse walks off with a cart .
ct out of the water , like a rearing horse . From the ship , the sm
ort of life that lives in a dog or a horse . Indeed , in other resp
ve been only a large seal , or sea - horse ; bearing all this in mi
anted before the ark of Israel , his horse ' s head and both the pa
n , then ? even as a man who rides a horse is called a horseman ? P
he try - works . First comes white - horse , so called , which is o
severed from the whale , the white - horse is first cut into portab
ularly chopping it into the portable horse - pieces . This spade is
That office consists in mincing the horse - pieces of blubber for
ich is conducted at a curious wooden horse , planted endwise agains
side of the mast ; why , there ' s a horse - shoe nailed on that si
what sign will the sun then be ? The horse - shoe sign ; for there
opposite the gold . And what ' s the horse - shoe sign ? The lion i
horse - shoe sign ? The lion is the horse - shoe sign -- the roari
ed to one of them ; a regular circus horse he was , too , that went
e same number of months , mounted on horse instead of sailing in sh
```

Obtenemos una salida con las veces que aparece la palabra buscada junto con un poco de texto, en este sencillo ejemplo podemos ver como interactuar con el lenguaje natural mediante un programa computacional.

En el siguiente ejemplo, podemos interpretar de manera grafica las ocurrencias de una palabra en un texto, posterior se puede aplicar un análisis según el objetivo requerido, sea este comparar la ocurrencia de una palabra de un texto en diferentes épocas o comparar los estilos según el autor o la temática. Para ellos ingresamos.

```
text1.dispersion_plot([ "whales" , "horse" , "monstrous" , "noble" ])
```

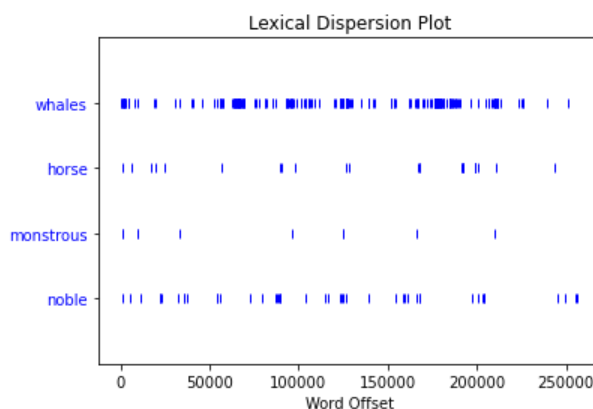


Figura 2. Diagrama de dispersión de la ocurrencia de una palabra

Con esto podríamos predecir la ocurrencia de una palabra o encontrar patrones repetitivos de la frecuencia de esta.

```
fdist1 = FreqDist(text1)
fdist1
FreqDist({' ': 18713, 'the': 13721, '.': 6862, 'of': 6536, 'and': 6024, 'a': 4569, 'to': 4542, ';': 4072, 'in': 3916, 'that': 2982, ..
.})
fdist1.plot(40, cumulative=True)
```

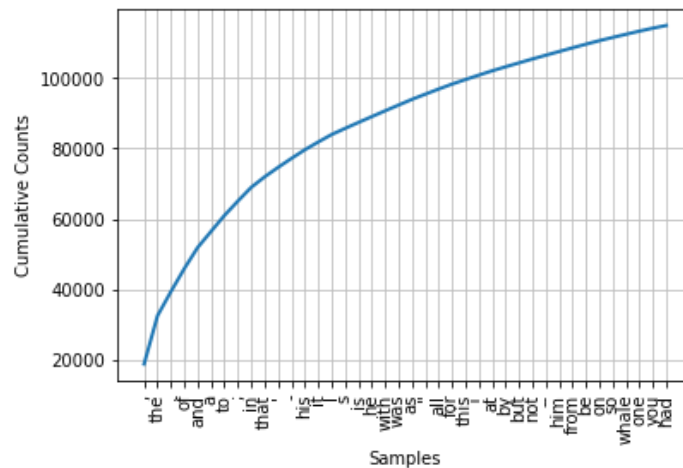


Figura 3. Frecuencia de la ocurrencia

Hemos observado la frecuencia de las palabras con más concurrencia y se podría tener una idea sobre el tema que estamos analizando, existen también las palabras que ocurren una sola vez que, junto con otras técnicas nos ayudan a entender el contexto, diferenciar el estilo para saber de qué autor se trata o la época; una aplicación de esto podría ser una tarea automatizada de clasificación de libros digitales o encontrar respuestas en buscadores. Para los objetivos del presente trabajo de investigación esto nos ayudará a extraer el contexto de una sentencia.

- Machine Learning

El aprendizaje de máquina evoluciona como un subconjunto de la inteligencia artificial, en el que las máquinas aprenden sin ser explícitamente programadas a partir de grandes volúmenes de información. En estos se buscan patrones, relaciones o asociaciones para predecir, clasificar o decidir sobre nuevas entradas en base a lo que aprendió con los datos anteriores. Se dice que se obtiene conocimiento a partir de los datos. (Pincay, 2022)

1.4. Estado del arte

1.4.1. Democracia participativa

En el Ecuador, desde la constitución del 2008 en el artículo 95 dice que, los ciudadanos “participarán de manera protagónica en la toma de decisiones, planificación y gestión en los asuntos públicos”. Entonces se reforma el sistema democrático para implementar el concepto de democracia participativa cuyo propósito es la participación de la ciudadanía en la toma de decisiones y rendición de cuentas de los gobiernos, la llamada soberanía popular. Se trata del

control de los poderes del estado mediante los medios de participación ciudadana para que los ciudadanos estén obligados a ejercer el control social en: la concepción, planificación, ejecución, gestión y evaluación de las políticas públicas como parte integral de ese “proceso permanente de construcción del poder ciudadano” (Chunga, 2014).

Una manera si se puede decir indirecta de este tipo de participación, es el expresar mediante opiniones en temas de política en plataformas digitales sobre asuntos de relevancia, existen medios de comunicación digitales que aprovechan el poco control para la difusión de sus opiniones o empatías sociales en los así llamados canales o periodismo digitales. Los ciudadanos aprovechan estos espacios y plataformas para expresar sus opiniones sobre los personajes y temas políticos, es ahí donde los interesados podrían encontrar si de verdad están representando a sus dirigidos y si estos se sienten representados.

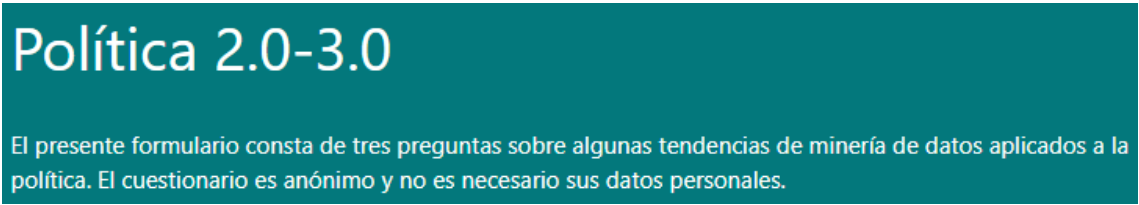
1.4.2.Situación actual del análisis de sentimientos en la política ecuatoriana

Existen pocos trabajos formales en repositorios educativos en internet sobre el análisis de sentimientos que sean aplicados a la política ecuatoriana, uno de ellos analiza puntualmente un período de tiempo sobre la carrera del expresidente Rafael Correa, trabajo donde se concluye que no existe una concordancia sobre las opiniones extraídas de redes sociales y los resultados electorales. Menciona también la complejidad de dicho trabajo por la falta de desarrollo de los algoritmos para el idioma español y algún modismo ecuatoriano. (Gómez-Torres, Jaimes, Hidalgo, & Luján-Mora. 2018).

Tampoco se conoce de una metodología de trabajo que sea aplicada comúnmente en el seguimiento de las campañas de los políticos antes durante y después de las carreras electorales, empresas o negocios que brinden este tipo de servicios no están disponibles si no es a manera de investigación o desarrollo.

Con la intención de indagar si en efecto los stakeholder de este presente trabajo están al tanto del tema planteado, se ha elaborado una sencilla encuesta y se ha enviado a un grupo pequeño de involucrados en temas políticos locales a fin de conocer la situación actual de lo que ellos conocen o no.

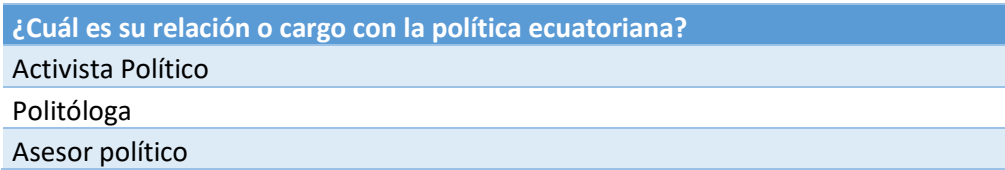
Los resultados son:



Política 2.0-3.0

El presente formulario consta de tres preguntas sobre algunas tendencias de minería de datos aplicados a la política. El cuestionario es anónimo y no es necesario sus datos personales.

Figura 4. Tema y encabezado de encuesta

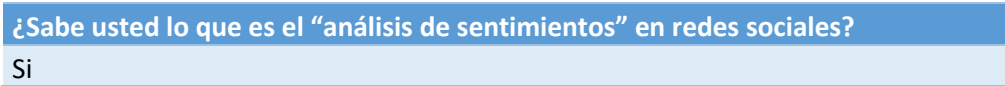


¿Cuál es su relación o cargo con la política ecuatoriana?

Activista Político

Politóloga

Asesor político



¿Sabe usted lo que es el “análisis de sentimientos” en redes sociales?

Si

No
Si

¿Utiliza o conoce usted, alguna herramienta tecnológica que sirva de apoyo o que tenga que ver con la manera de hacer política en el Ecuador?
Si
Si
No

En caso de que la respuesta a la pregunta 2 sea si, mencione cuál conoce.
Redes sociales
Medios de comunicación

Según esto; más de la mitad de quiénes contestaron afirman saber lo que es al análisis de sentimientos, más allá de si el concepto es completamente correcto, han escuchado o tienen idea de la propuesta. Así mismo, igual cantidad de personas utilizan algún medio tecnológico para ello.



Figura 5. Cuantos saben lo que es AS



Figura 6. Cuantos usan tecnología en política

Queda entonces un camino largo por recorrer hasta que en Ecuador y también en América Latina se adopte la costumbre de la implementación de minería de datos con aplicaciones de análisis de sentimientos como apoyo tecnológico para la toma de decisiones y nuevas estrategias de manejo de simpatizantes y votantes.

2. Marco teórico

2.1. Ciudadanía digital

De igual manera que nos comportamos con el entorno que nos rodea como individuos en una sociedad en el mundo físico, se espera un comportamiento similar en el mundo virtual debido a la popularización del uso del internet y sus servicios. Ser respetuosos y vivir en urbanidad se debe extender también en las plataformas digitales de opinión y comunicación dado que el mundo ha cambiado mucho y existen maneras diferentes (modernas) de ser ciudadanos. Existen varios foros

de debate sobre quiénes deben impartir estos hábitos a los nuevos ciudadanos, son estos los que llegan a un mundo donde la tecnología abarca todos los aspectos de nuestras vidas y son usuarios de la tecnología desde tan temprana edad. El aprender que hacer y que no hacer en Internet (Haro, 2020).

Edith González en su artículo “Formación Ciudadana, una nueva materia para el currículum escolar” da la siguiente definición sobre ciudadanía digital:

“Un ciudadano digital es aquella persona que utiliza tecnología de la información para mejorar su participación en la sociedad, la política y el gobierno, o sea, los que utilizan Internet regularmente y con efectividad”.

Tanta popularidad adopta este nuevo concepto que ya se elaboran leyes sobre la protección de datos personales y penas para los cyber criminales, nos enfrentamos también a los problemas que las nuevas tecnologías enfrentan con la misma o más complejidad que el mundo físico.

Parte de la utilidad que el análisis de sentimientos propone en este aspecto, es la utilización de algoritmos que pueden detectar expresiones ofensivas en redes sociales u otras plataformas para bloquearlas, preguntando en ocasiones si está de acuerdo en ver dicho contenido. En el caso del texto, esto se lo realiza mediante el PLN analizado anteriormente; es decir, text minning aplicado en beneficio del cuidado y su susceptibilidad. No lo realiza una persona de comentario en comentario, si no que el análisis de sentimientos polariza las expresiones de las personas reales o no y así determina contenido potencialmente ofensivo de manera automatizada.

Este comentario infringe nuestras Normas comunitarias sobre lenguaje que incita al odio.

ABIERTO

¿QUÉ PUEDES HACER?

Ver opciones

Obtén más información sobre este comentario y descubre qué puedes hacer

Figura 7. Ejemplo de bloqueo de comentario ofensivo (notigram.com)

Esto ayuda a la regulación del comportamiento del ciudadano digital. Otra aplicación del análisis de sentimientos, en el presente trabajo se realizará algo similar.

2.2. Política 2.0

Según Ortiz Marín, et. al, (2011) política 2.0 es “el esfuerzo que realizan los ciudadanos que buscan participar en la formulación, desarrollo y evaluación de las políticas públicas mediante el uso de la inteligencia colectiva plasmada en las redes sociales que se han formado para ese propósito. Por otra parte, a través de la política 2.0, el electorado recurre a internet para organizarse y participar masivamente en las campañas políticas de sus candidatos que, mediante esta modalidad, buscan ganar las elecciones”.

Es una nueva manera de hacer política que llega embebida con la popularización del Internet y las redes sociales en donde los candidatos expresan sus ideas y los electores pueden contribuir con una realimentación a las expresiones o propuestas, en la manera tradicional esto no era así, la prensa, radio y televisión es un medio de comunicación de un solo sentido. Alguien se dirigía a

un grupo de personas en un espacio de tiempo definido y no se sabía en tiempo real y mucho menos con exactitud la aceptación del público receptor pues no había una respuesta a las expresiones o propuestas vertidas, es donde llegan las nuevas tecnologías junto con conceptos emergentes como: sociedad de la información, economía de la información, etc.

Propone una comunicación más fluida y de doble sentido donde la comunidad responde y se expresa generando algo más que solo comentarios, generan datos, los mismos que pueden llevar a la información y estos al conocimiento (ciencia de datos). Los interesados encuentran ahí herramientas útiles que sirve de herramienta en sus decisiones.

No es una técnica nueva, el uso de Internet para hacer política se viene realizando desde hace ya unas décadas con algunos casos de éxito como el del expresidente Bill Clinton quién se apoyó en Internet para realizar una operación denominada “move on” que evitaba un juicio político en su contra por conducta sexual inapropiada mediante una campaña de correos electrónicos y páginas web que recolectaba firmas. El principio era proponer que se deje atrás (move on) el juicio y que los esfuerzos se centren en los problemas más urgentes de esa nación, la iniciativa recolectó más de 475.000 firmas en dicha página web.

2.3. Análisis de redes sociales

Según lo expuesto hasta el momento se hace necesario regresar la vista hacia las plataformas digitales para entender un panorama alterno a la realidad física a la que estamos acostumbrados, sean estas encuestas, opiniones o entrevistas en medios clásicos, para comparar la cantidad de participaciones que podría generarse en una red social con el fin de decidir si alguien que quiera participar en la política, centrará o no sus esfuerzos promoviendo su candidatura en estas plataformas.

En base a los datos de la empresa Hootsuite, cuyo trabajo consiste en la gestión de varias redes sociales en un solo sitio, las estadísticas más importantes de las redes sociales son:

- Proyección a futuro

La siguiente gráfica muestra la tendencia de la cantidad de usuarios (en millones) por plataforma en Latinoamérica.

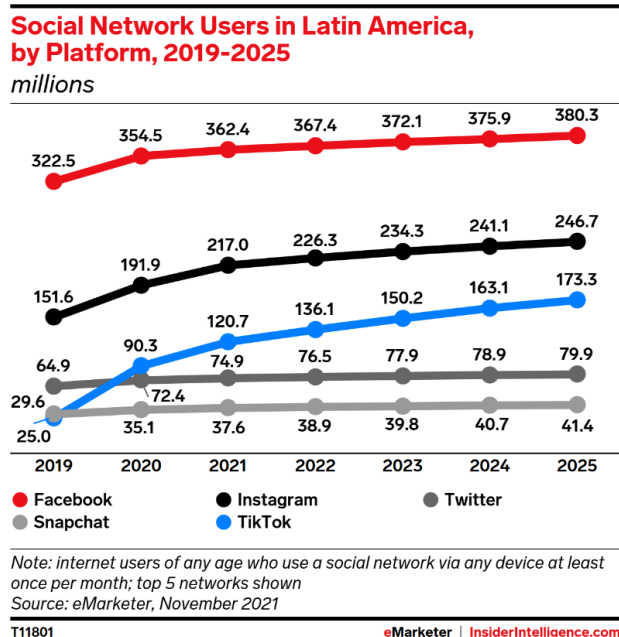


Figura 8. Proyección en millones de usuarios (insiderintelligence, 2022)

Claramente se observa un dominio de la red social Facebook sobre otras opciones.

- Páginas WEB más visitadas

Si bien es cierto el análisis se basa en redes sociales, las páginas web son de utilidad y se consideran brevemente en este estudio:

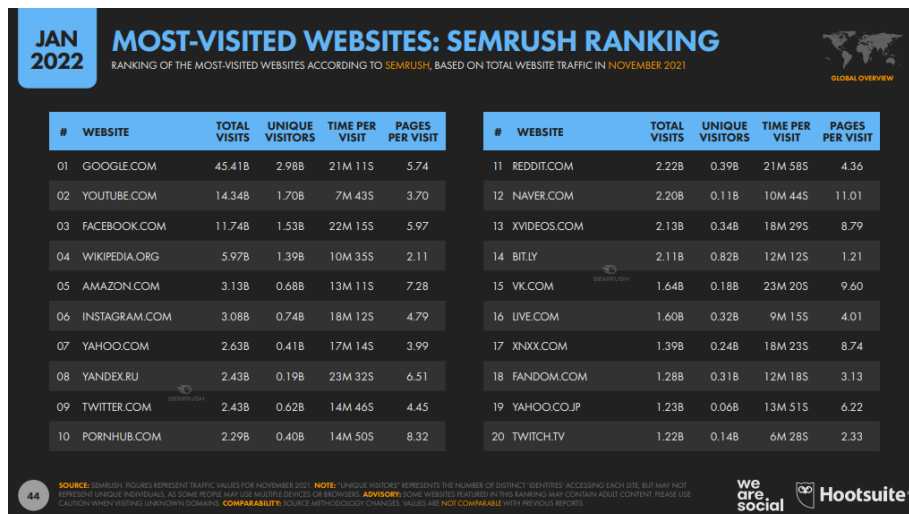


Figura 9. Páginas web más visitadas (Hootsuite, 2022)

Las páginas web de redes sociales se encuentran en la lista de las 10 primeras, estos datos son a nivel global.

Existe una tendencia al alza del uso de las redes sociales para diversos propósitos, en promedio la gente visita más de 10 minutos al menos una de las plataformas. Twitter registra 2,3 billones de visitas, un alcance significativo si se quiere llegar a un público objetivo.

3. Diseño de la propuesta

3.1. Antecedentes

En Ecuador están próximas las elecciones seccionales a realizarse en el 2023 para el balotaje de dignidades como alcaldes, prefectos y concejales. Se espera que todos los candidatos hagan uso de las redes sociales y plataformas informáticas para sus campañas y poder llegar a más electores con sus propuestas, los motivos son suficientes luego de haber expuesto lo analizado hasta el momento: gran cantidad de usuarios, libertad de expresión, participación en doble sentido, tiempo suficiente que pasa cada usuario en internet, etc.

Es un buen momento para plantear metodologías de trabajo basadas en ciencia de datos que permitan mejorar el conocimiento sobre la representación de los electores y su candidato. También, apoyar en la toma decisiones sobre las estrategias que pudiesen presentarse.

3.2. Narración

Para el propósito se recurrirá a alguien involucrado en política que participará o pretenda participar en las elecciones ya mencionadas para orientar el objetivo del presente trabajo en primer lugar a: que se familiarice en profundidad sobre las alternativas basadas en sistemas computacionales para implementar política 2.0, conozca si está representando a su público objetivo e informarse mediante redes sociales sobre la coyuntura local para así plantear una estrategia en base también a la red social Twitter.

Debido a que el presente trabajo de titulación finaliza antes que empiecen las campañas, el objetivo se orienta a extraer las opiniones polarizadas sobre los personajes que podrían ser de interés de él o los stakeholders.

Para llevar a cabo esta tarea, se propone cierta cantidad de procedimientos descritos en el siguiente diagrama de flujo.

3.3. Diagrama del proceso

No se propone realizar el análisis de sentimientos para un tema en específico, como se trata de crear una estrategia basada en ciencia de datos la intención es consultar al stakeholder qué tema o personaje podría ser de su interés y presentar los resultados en base a ese análisis para que el interesado realice las acciones correspondientes. Por lo expuesto, dicho trabajo se basa en una serie de tareas repetitivas que mostrarán una realidad en base a la red social Twitter por lo que el siguiente diagrama de flujo expuesto a continuación es de una sola tarea, misma que se llevará a cabo en más de una ocasión.

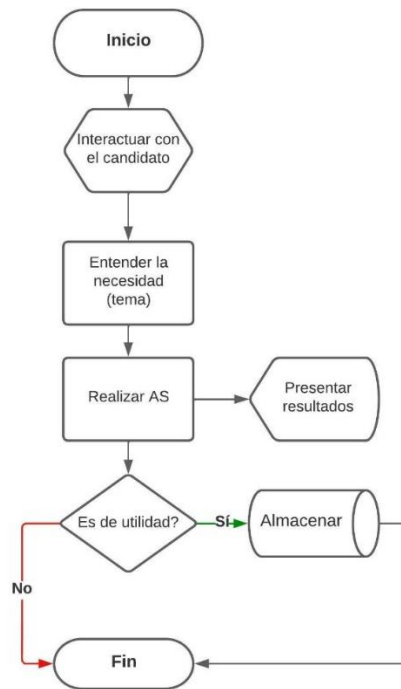


Figura 10. Diagrama de flujo de la propuesta

- Interactuar con el candidato

Dialogar con el interesado para planificar los pasos siguientes y entender su estrategia.

- Entender la necesidad

Concluir qué tema o personaje es de su interés o necesita saber la aceptación cuantitativamente de las opiniones de dicho tema o perfil.

- Realizar el análisis de sentimientos

Con la cuenta objetivo, realizar el análisis de sentimientos correspondiente y extraer las cifras de las opiniones polarizadas.

- Presentar los resultados

Enviar los resultados de manera legible, gráfica y en un formato poco técnico obtenidos del análisis de sentimientos y si es necesario solventar las inquietudes.

- ¿Es de utilidad?

Es el candidato quién aceptará o no la utilidad de los resultados en base a su experiencia. De ser el caso, podría confirmar si los resultados estuviesen lejos de la realidad por su conocimiento de las opiniones directas o por otros medios. Por supuesto; se trata de abrir la visión más allá de su universo.

En caso de ser útil o correcto, se almacena en un repositorio por si se lo necesita en un análisis posterior. Luego se da por finalizada la tarea.

4. Desarrollo y resultados

Twitter es una red social que más allá de las características conocidas por sus millones de usuarios, cuenta con una API que permite la extracción de datos para su posterior análisis o en tiempo real. Para acceder a estas funciones es necesario crear una cuenta de desarrollador a través de una cuenta de usuario. Dicha plataforma ha actualizado su versión de la API 1.0 a 2.0 (lanzada en agosto de 2020) en donde se pueden observar algunos cambios significativos como nuevos niveles de privilegio el cuál influyó en la elaboración del presente trabajo; por esto, ha sido útil solicitar la elevación en los permisos disponibles pues hay temas relacionados con este inconveniente que pudiese presentarse y se evita con anticipación.

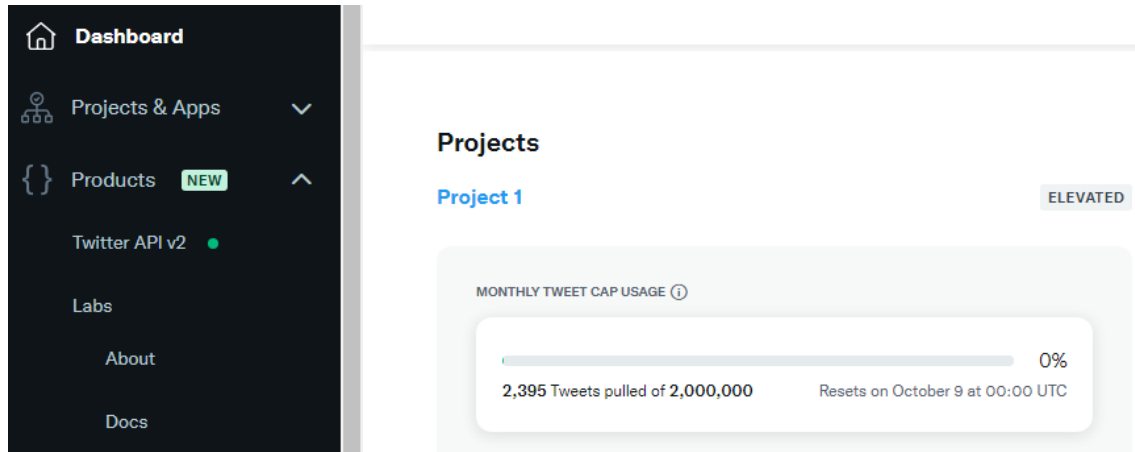


Figura 11. Versión de API y nivel de privilegio.

Otro cambio significativo son las credenciales que se obtienen para la conexión con la API, aunque Twitter menciona que son compatibles con versiones anteriores.

a) Interactuar con el candidato

Oscar Hoyos es quién realizará las solicitudes sobre temas o perfiles de interés, ha mencionado su intención en participar en las elecciones próximas, a la fecha de la elaboración (septiembre de 2022) solicita saber la opinión de Jorge Yunda @LoroHomero, pues ha presentado su candidatura oficial y es tema de coyuntura

b) Entender la necesidad

En este caso es la polaridad de los tweets de una cuenta mas no de un tema

c) Realizar análisis de sentimientos

La metodología de trabajo propuesta es la siguiente.

4.1. Obtención de los datos

Obtenemos los últimos 400 Tweets relacionados con el perfil @LoroHomero, generados por el propietario o etiquetados, para colocarlos en un data frame

```
221 # Extracción de n tweets de una cuenta especificada
222 posts = api.user_timeline(screen_name = "LoroHomero", count = 400, tweet_mode = "extended")

227 # Generamos un data frame que contiene los Tweets
228 df = pd.DataFrame([tweet.full_text for tweet in posts], columns=["Tweets"])
```

4.2. Limpieza de los datos

Definimos la clase

```

33 def clean_text(txt):
34     ...
35     Clean the tweet text
36     - Remove:
37         1. # Hash character (Not the whole hashtag)
38         2. Hyperlinks
39         3. Usernames
40         4. Emojis
41         5. RT (Retweets)
42     ...
43     txt = re.sub(r"RT[\s]+", "", txt)
44     txt = txt.replace("\n", " ")
45     txt = re.sub(" +", " ", txt)
46     txt = re.sub(r"https?:\/\/\/\S+", "", txt)
47     txt = re.sub(r"(@[A-Za-z0-9_]+)|(^[\w\s]|\#)", "", txt)
48     txt = emoji.replace_emoji(txt, replace='')
49     txt = txt.strip()

```

Aplicamos sobre el dataframe creado que contiene los textos.

```

230     # Limpieza de los textos de los tweets
231     df["Tweets"] = df["Tweets"].apply(clean_text)

```

4.3. Análisis de sentimientos

Definimos las clases y colocamos los datos en dos data frames uno para la polaridad y otro para la subjetividad.

```

54 def get_subjectivity(txt):
55     ...
56     grado de opinión en el texto del tweet [0, 1] donde 0 representa
57     información fáctica y 1 representa una opinión personal
58     ...
59
60     return TextBlob(txt).sentiment.subjectivity
61
62
63 def get_polarity(txt):
64     ...
65     número de punto flotante [-1, 1] que representa
66     declaraciones donde -1 es negativo, 0 es neutral y 1 es positivo
67     ...
68
69     return TextBlob(txt).sentiment.polarity

```

```

233     # Se crean dos columnas adicionales para subjetividad y polaridad:
234     df["Subjectivity"] = df["Tweets"].apply(get_subjectivity)
235     df["Polarity"] = df["Tweets"].apply(get_polarity)
236     pd.set_option('display.max_columns', 4)

```

d) Presentar resultados

4.4. Visualización de resultados

Como se había mencionado anteriormente, según solicitud del interesado, se muestran los resultados en gráficos.

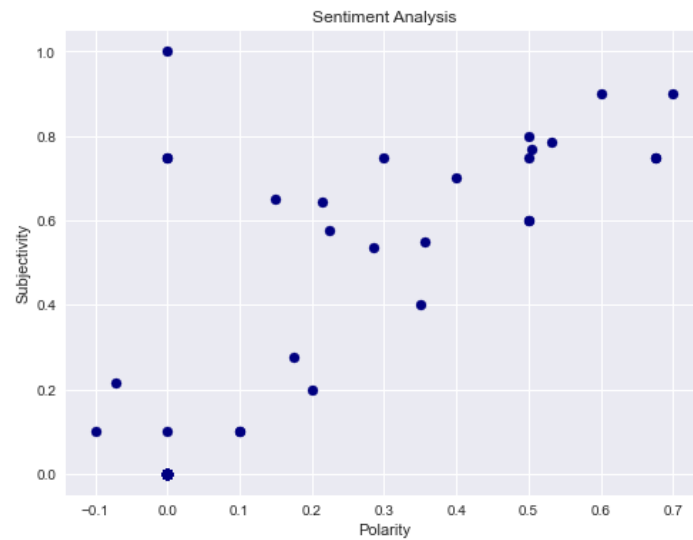


Figura 12. Relación subjetividad-polaridad

Este gráfico muestra un diagrama de dispersión de la relación subjetividad – polaridad de las columnas contenidas en el data frame elaborado.

% OF POSITIVE TWEETS: 11.11
% OF NEGATIVE TWEETS: 1.01

El porcentaje de tweets positivos es mayor a los negativos en 10 a 1.

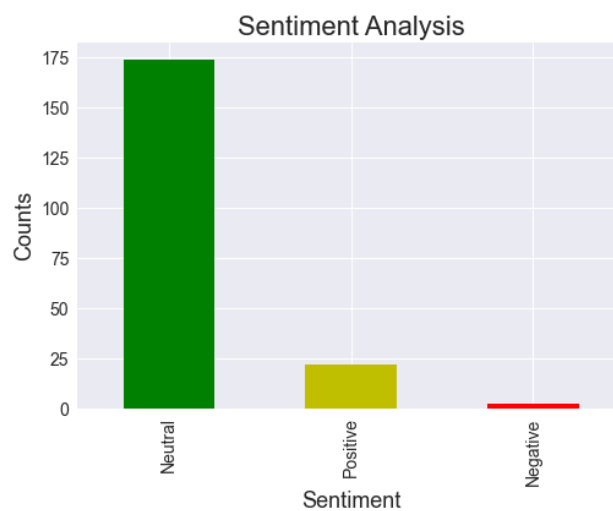


Figura 13. Gráfico de barras sobre los resultados

Existe una gran cantidad de tweets neutros.

4.5. Evaluación

De cierta manera se pretende evaluar el trabajo de la librería textblob implementada mediante observación, es necesario dos pasos para esto: comprobar que los tweets obtenidos son de la cuenta correcta y que la polaridad es apropiada. Para lo primero recurrimos a una herramienta

disponible en las hojas de cálculo de Google que permite extraer tweets según un objetivo parametrizado lo que demostraría la funcionalidad de la API y por simple observación en la cuenta objetivo. Para lo segundo, interpretamos directamente el tweet. Esto se lo hará de manera inferencial; es decir, se toma un tweet al azar.

- Tweets positivos obtenidos del data frame del modelo propuesto

ONLY POSITIVE TWEETS:

TWEET #1: The most beautiful encounter

TWEET #5: Hoy me reencontré a los tiempos con los amigos representantes de bares discotecas terminales terrestres Red de Mercados Unión de Ligas Barriales de Quito y Asociación de Ligas Full apoyo gracias panas

TWEET #14: Look at her so adorable tails When Mommy comes Love or not Dachshund Dachshunds

- Tweets obtenidos de la hoja de cálculo sin polarizar

Date	Screen Name	Full Name	Tweet Text
21/09/2022 17:07	@LoroHomero	Jorge Yunda Machado	Difícil conciliar el sueño después de las macabras noticias en torno a María Belén, tanto misterio a su alrede
21/09/2022 14:57	@LoroHomero	Jorge Yunda Machado	Hoy me reencontré a los tiempos con los amigos representantes de bares, discotecas, terminales terrestres,

- Tweets de la página oficial



Figura 14. Tweet obtenido de la página fuente

Podemos comprobar que la API se conecta de manera adecuada en los dos casos en que se usa Twitter para desarrollo y obtiene los tweets relacionados con la cuenta objetivo.

Se procede entonces a leer los tweets publicados para interpretarlos de manera deliberada. Por ejemplo:

TWEET #5: Hoy me reencontré a los tiempos con los amigos representantes de bares discotecas terminales terrestres Red de Mercados Unión de Ligas Barriales de Quito y Asociación de Ligas Full apoyo gracias panas

Textblob lo califica como positivo, el contexto del mensaje es un reencuentro con amigos y gratitud lo que en efecto se considera positivo.

- e) ¿Es de utilidad?

Se retoma con el interesado, indica su aprobación pues su objetivo es tener más aceptación o relación con tweets positivos que la cuenta usada para el análisis. Se almacena el resultado en el presente documento. Se cierra el proceso.

Se reinicia el procedimiento ahora con una nueva solicitud, la cuenta de Luz Elena Coloma @ZULNELEA y los resultados son:

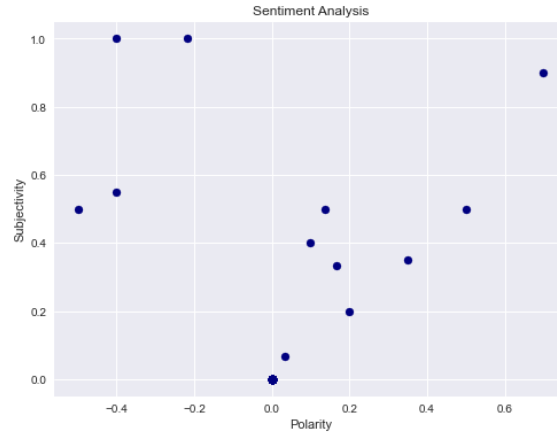


Figura 15. Diagrama de dispersión de @ZULNELEA

% OF POSITIVE TWEETS: 4.0

% OF NEGATIVE TWEETS: 2.0

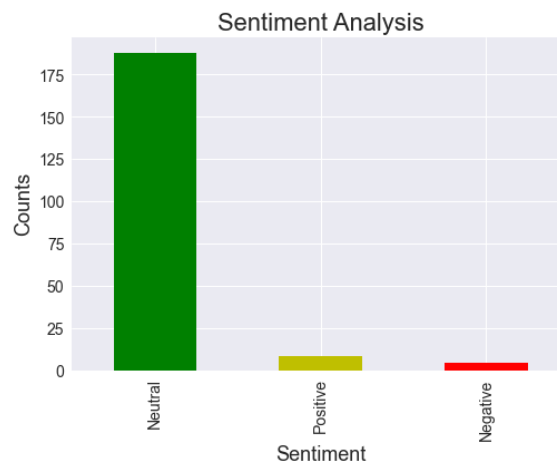


Figura 16. Gráfico de barras resultado AS

También se observa una alta cantidad de resultados neutros, pero con un ligero balance entre los resultados positivos y negativos, para comprobar tomamos con el mismo método inferencial pero ahora con un tweet cuya librería haya calificado como negativo.

ONLY NEGATIVE TWEETS:

TWEET #142: Opinión Quito está herida vulnerable y es ahí donde es necesario que habite la esperanza y la posibilidad de repens

El post en Twitter es:

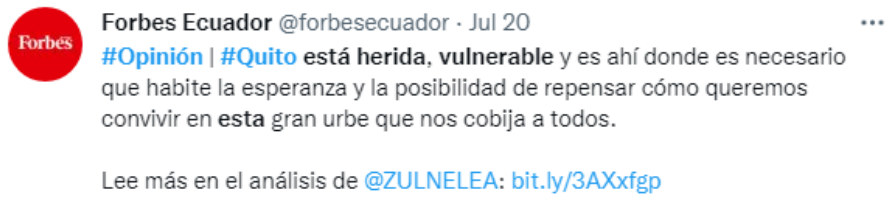


Figura 17. Tweet calificado como negativo

Se trata de un retweet donde se etiqueta la cuenta objetivo, se puede decir que es un comentario que habla de un tema negativo, pero no necesariamente en contra de una persona (ruido).

El interesado acepta la utilidad de la información y plantea participar en entrevistas sobre temas positivos para aumentar la percepción de optimismo en la gente como también, superar la cantidad de tweets positivos. Se almacena el resultado y se finaliza el proceso.

5. Conclusiones y recomendaciones

5.1. Conclusiones

La coyuntura presente al momento de realizar el análisis de sentimientos es de suma importancia para una correcta interpretación de los resultados, puesto que las opiniones de los usuarios podrían no ser negativos necesariamente contra una persona si no contra un tema o acontecimiento en específico al cuál el usuario al que se lo realiza el AS se refiere.

Se presenta una alta y considerable cantidad de resultados polarizados como neutros debido a que algunas cuentas “re-tweetean” contenido en gran parte informativo o en otro idioma o son incluidos en hilos que hablan sobre algún tema, en general, neutro.

Textblob y Vader son librerías enfocadas para el idioma inglés, textblob tiene desarrollo para el idioma español, aunque algunas implementaciones recomiendan la traducción primero al inglés y luego realizar el análisis de sentimientos, esto podría presentar un inconveniente si el texto original usa jerga o modismos propios de un país en especial Ecuador, aunque si el objetivo es político o educativo esto casi no se presentaría.

Al momento de la realización del presente trabajo se percibe cierta preferencia del electorado por el candidato Jorge Yunda con 11% de Tweets positivos (Figura 13), mientras que Luz Elena Coloma presenta 4% de Tweets positivos (Figura 16), por supuesto el análisis se lo realizó sobre dos cuentas objetivo y a meses de las elecciones, no queda más que esperar los resultados y comprobar el margen de error obtenido del presente análisis.

5.2. Recomendaciones

Es importante tener claro el objetivo al momento de implementar un análisis de sentimientos en base a lo que se pueda obtener de una red social, las plataformas permiten configurar la privacidad y los perfiles para que, por ejemplo, no se pueda comentar. O por último el propietario del perfil podría borrar mensajes que no le convienen, esto provocaría lo que en ciencia de datos se conoce como, la calidad de los datos o ruido, arrojando un resultado de una falsa aceptación.

Siempre es necesario en cualquier implementación de un modelo de ciencia de datos o IA, algún método de evaluación para contrastar algún margen de error, las librerías para NLTK en español

tienen aún mucho por desarrollar y los modismos o el sarcasmo podrían presentar falsos positivos o viceversa. En este caso se lo hizo mediante la lectura e interpretación directa, pero se podría trabajar aún más en esto último.

Debido a que el presente trabajo se lo realiza meses antes de las campañas para las elecciones y las elecciones en sí, queda abierta la posibilidad de la implementación de los análisis de tweets en tiempo real (librería y función disponible en la API 2.0) el día mismo de las votaciones, podría contrastarse con encuestas realizadas en redes sociales y los resultados de los métodos convencionales como exit poll, el objetivo es actualizar los tradicionales métodos de encuestas que aumente la transparencia de estas y comprobar la funcionalidad de AS.

6. Bibliografía

Bunge, M. (2012). *Filosofía de la tecnología y otros ensayos*. Lima: Fondo editorial.

developer.twitter.com. (s.f.). Obtenido de developer.twitter.com: <https://developer.twitter.com/en/updates/changelog#:~:text=On%20August%2012th%2C%2020%2C%20we,Twitter%20API%20v2%3A%20Early%20access.>

Chunga Arteaga, N. E. (2014). *LA DEMOCRACIA PARTICIPATIVA EN EL ECUADOR, NUEVO MODELO DE DEMOCRACIA EN LA CONSTITUCION DEL 2008* (Master's thesis).

Gómez-Torres, Estevan, Jaimes, Roger, Hidalgo, Orlando, & Luján-Mora, Sergio. (2018). Influencia de redes sociales en el análisis de sentimiento aplicado a la situación política en Ecuador. *Enfoque UTE*, 9(1), 67-78. <https://doi.org/10.29019/enfoqueute.v9n1.235>

Ortiz Marín, Manuel , & Gómez Castellanos, Rodolfo M. , & Concepción Montiel, Luis Enrique (2011). Tecnologías de la comunicación y política 2.0. *Espacios Públicos*, 14(30),72-84.[fecha de Consulta 18 de Agosto de 2022]. ISSN: 1665-8140. Disponible en: <https://www.redalyc.org/articulo.oa?id=67618934006>

Steven Bird, E. K. (2009). *Natural Language Processing with Python*. O'Reilly Media.