



**UNIVERSIDAD DE GUAYAQUIL  
FACULTAD DE INGENIERIA INDUSTRIAL  
CARRERA DE INGENIERIA EN TELEINFORMÁTICA**

**TRABAJO DE TITULACIÓN  
PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERA EN TELEINFORMÁTICA**

**ÁREA  
TECNOLOGÍA DE LOS ORDENADORES**

**TEMA  
USO DE MINERÍA DE DATOS PARA ANÁLISIS DE  
SENTIMIENTOS EN TWEETS DEL FÚTBOL  
ECUATORIANO.**

**AUTORA  
AVILA LÓPEZ MERLY CAROLINA.**

**DIRECTOR DEL TRABAJO  
ING. COMP. PLAZA VARGAS ANGEL MARCEL, MG.**

**GUAYAQUIL, SEPTIEMBRE 2022**



**ANEXO XI.- FICHA DE REGISTRO DE TRABAJO  
DE TITULACIÓN  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA DE INGENIERÍA EN TELEINFORMÁTICA**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA			
FICHA DE REGISTRO DE TRABAJO DE TITULACIÓN			
<b>TÍTULO Y SUBTÍTULO:</b>	Uso de minería de datos para análisis de sentimientos en tweets del fútbol ecuatoriano.		
<b>AUTOR(ES)</b> (apellidos/nombres):	Avila López Merly Carolina.		
<b>TUTOR/ REVISOR</b> (apellidos/nombres):	Ing. Comp. Plaza Vargas Angel Marcel, Mg. / Ing. Sist. Pincay Bohórquez Freddy Steve, Mg		
<b>INSTITUCIÓN:</b>	Universidad de Guayaquil		
<b>UNIDAD/FACULTAD:</b>	Facultad de Ingeniería Industrial.		
<b>MAESTRÍA/ESPECIALIDAD:</b>			
<b>GRADO OBTENIDO:</b>	Ingeniera en Teleinformática.		
<b>FECHA DE PUBLICACIÓN:</b>	28 de septiembre 2022	<b>No. DE PÁGINAS:</b>	114
<b>ÁREAS TEMÁTICAS:</b>	Tecnología de los Ordenadores		
<b>PALABRAS CLAVES/ KEYWORDS:</b>	Tweets, Análisis de sentimiento, hinchas, procesamiento de datos, API, polaridad, lenguaje natural, Jupyter.		
<p><b>RESUMEN</b></p> <p>En este trabajo se procederá a realizar un análisis de sentimiento en tweets, creando algoritmos en Jupyter Notebook que ayuden a la extracción, recolección, selección, limpieza y validación de datos que permita aplicar la minería de datos mediante dicho análisis de sentimientos. El uso masivo de las redes sociales en la actualidad genera una cantidad de datos enormes, la idea principal de este análisis es clasificar y determinar la polaridad (positivos, negativos, neutros) de cada Tweets realizados por usuarios acerca de cuatro equipos del fútbol ecuatoriano; Barcelona Sporting Club, Club Sport Emelec, Liga de Quito Universitaria e Independiente del Valle, este análisis de sentimiento aplicado en este ámbito deportivo, tiene la finalidad de crear estrategias para el marketing de los equipos, ayudando de esta manera a fortalecer su marca o nombre, a saber y tener conocimiento acerca de lo que piensan sus seguidores y en algunos casos ajustar o cambiar acciones basados en las preferencias de sus hinchas.</p> <p><b>ABSTRACT</b></p> <p>In this work we will proceed to perform a sentiment analysis on tweets, creating algorithms in Jupyter Notebook that help the extraction, collection, selection, cleaning and validation of data that allows the application of data mining through sentiment analysis. The massive use of social networks nowadays generates a huge amount of data, the main idea of this analysis is to classify and determine the polarity (positive, negative, neutral) of each tweet made by users about four Ecuadorian soccer teams; Barcelona Sporting Club, Club Sport Emelec, Liga de Quito Universitaria and Independiente del Valle, this sentiment analysis applied in this sporting environment, aims to create strategies for marketing teams, thus helping to strengthen its brand or</p>			

name, to know and have knowledge about what their followers think and in some cases adjust or change actions in the preferences of their fans.		
ADJUNTO PDF:	SI X	NO
CONTACTO CON AUTOR/ES:	Teléfono: 593-939669306	E-mail: <a href="mailto:merly.avilal@ug.edu.ec">merly.avilal@ug.edu.ec</a>
CONTACTO CON LA INSTITUCIÓN:	Nombre: Ing. Ramón Maquilón Nicola	
	Teléfono: 593-2658128	
	E-mail: <a href="mailto:direccionTi@ug.edu.ec">direccionTi@ug.edu.ec</a>	



**ANEXO XII.- DECLARACIÓN DE AUTORÍA Y DE  
AUTORIZACIÓN DE LICENCIA GRATUITA  
INTRANSFERIBLE Y NO EXCLUSIVA PARA EL USO  
NO COMERCIAL DE LA OBRA CON FINES NO ACADÉMICOS**



**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

---

**LICENCIA GRATUITA INTRANSFERIBLE Y NO COMERCIAL DE LA OBRA CON  
FINES NO ACADÉMICOS**

Yo, **AVILA LÓPEZ MERLY CAROLINA**, con C.C. No. **0956920870**, certifico que los contenidos desarrollados en este trabajo de titulación, cuyo título es **“USO DE MINERÍA DE DATOS PARA ANÁLISIS DE SENTIMIENTOS EN TWEETS DEL FÚTBOL ECUATORIANO.”** son de mi absoluta propiedad y responsabilidad, en conformidad al Artículo 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN\*, autorizo la utilización de una licencia gratuita intransferible, para el uso no comercial de la presente obra a favor de la Universidad de Guayaquil.

A handwritten signature in blue ink that reads "Merly Avila Lopez".

---

**AVILA LÓPEZ MERLY CAROLINA**  
C.C. No. 0956920870



**ANEXO VII.- CERTIFICADO PORCENTAJE  
DE SIMILITUD  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**



Habiendo sido nombrado **ING. PLAZA VARGAS ANGEL MARCEL, MSC.**, tutor del trabajo de titulación certifico que el presente trabajo de titulación ha sido elaborado por **AVILA LÓPEZ MERLY CAROLINA**, con mi respectiva supervisión como requerimiento parcial para la obtención del título de **INGENIERA EN TELEINFORMÁTICA**.

Se informa que el trabajo de titulación: **USO DE MINERÍA DE DATOS PARA ANÁLISIS DE SENTIMIENTOS EN TWEETS DEL FÚTBOL ECUATORIANO**, ha sido orientado durante todo el periodo de ejecución en el programa Antiplagio **TURNITIN** quedando el **2%** de coincidencia.

<https://ev.turnitin.com/app/carta/es/?lang=es&o=1897148176&s=1&u=1133714257>



Firmado electrónicamente por:

**ANGEL MARCEL  
PLAZA VARGAS**

**ING. COM. PLAZA VARGAS ANGEL MARCEL, MG.  
DOCENTE TUTOR  
C.C. 0915953665  
FECHA: 11/9/2022**



**ANEXO VI. - CERTIFICADO DEL DOCENTE-TUTOR  
DEL TRABAJO DE TITULACIÓN  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**



Guayaquil, 11 de septiembre del 2022

Sr (a).

**Ing. Annabelle Lizarzaburu Mora, MG.**

Director (a) de Carrera Ingeniería en Teleinformática / Telemática

**FACULTAD DE INGENIERÍA INDUSTRIAL DE LA UNIVERSIDAD DE  
GUAYAQUIL**

Ciudad. –

De mis consideraciones:

Envío a Ud. el Informe correspondiente a la tutoría realizada al Trabajo de Titulación: **“USO DE MINERÍA DE DATOS PARA ANÁLISIS DE SENTIMIENTOS EN TWEETS DEL FÚTBOL ECUATORIANO”** del estudiante **AVILA LÓPEZ MERLY CAROLINA**, indicando que ha (cumplido con todos los parámetros establecidos en la normativa vigente:

El trabajo es el resultado de una investigación.

El estudiante demuestra conocimiento profesional integral.

El trabajo presenta una propuesta en el área de conocimiento.

El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se adjunta el certificado de porcentaje de similitud y la valoración del trabajo de titulación con la respectiva calificación.

Dando por concluida esta tutoría de trabajo de titulación, **CERTIFICO**, para los fines pertinentes, que la estudiante está apta para continuar con el proceso de revisión final.

Atentamente,



Firmado electrónicamente por:  
**ANGEL MARCEL  
PLAZA VARGAS**

**ING. COMP. PLAZA VARGAS ANGEL MARCEL, MG.**

**CC: 0915953665**

**Fecha: 11 de septiembre del 2022**



**ANEXO VIII. - INFORME DEL DOCENTE REVISOR  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**



Guayaquil, 22 de septiembre de 2022.

Sr (a).

**Ing. Annabelle Lizarzaburu Mora, MG.**

Director (a) de Carrera Ingeniería en Teleinformática / Telemática

**FACULTAD DE INGENIERÍA INDUSTRIAL DE LA UNIVERSIDAD DE  
GUAYAQUIL**

Ciudad. -

De mis consideraciones:

Envío a Ud. el informe correspondiente a la REVISIÓN FINAL del Trabajo de Titulación **“USO DE MINERÍA DE DATOS PARA ANÁLISIS DE SENTIMIENTOS EN TWEETS DEL FÚTBOL ECUATORIANO”** del estudiante **AVILA LÓPEZ MERLY CAROLINA**. Las gestiones realizadas me permiten indicar que el trabajo fue revisado considerando todos los parámetros establecidos en las normativas vigentes, en el cumplimiento de los siguientes aspectos:

Cumplimiento de requisitos de forma:

El título tiene un máximo de 14 palabras.

La memoria escrita se ajusta a la estructura establecida.

El documento se ajusta a las normas de escritura científica seleccionadas por la Facultad.

La investigación es pertinente con la línea y sublíneas de investigación de la carrera.

Los soportes teóricos son de máximo 5 años. La propuesta presentada es pertinente.

Cumplimiento con el Reglamento de Régimen Académico:

El trabajo es el resultado de una investigación.

El estudiante demuestra conocimiento profesional integral.

El trabajo presenta una propuesta en el área de conocimiento.

El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se indica que fue revisado, el certificado de porcentaje de similitud, la valoración del tutor, así como de las páginas preliminares solicitadas, lo cual indica el que el trabajo de investigación cumple con los requisitos exigidos.

Una vez concluida esta revisión, considero que el estudiante está apto para continuar el proceso de titulación. Particular que comunicamos a usted para los fines pertinentes.

Atentamente,



Firmado electrónicamente por:

**FREDDY STEVE**

**PINCAI**

**BOHORQUEZ**

**ING. SIST. PINCAI BOHORQUEZ FREDDY STEVE, MG**

**C.C:1308497682**

**FECHA: 22 de septiembre de 2022**

### **Dedicatoria**

En primer lugar, a Dios por haberme guiado en cada paso de este proyecto y sobre todo por darme sabiduría.

A mi padre Manuel Avila que ahora es un ángel en mi vida, su motivación y confianza en mí hicieron esto posible, este logro es más suyo que mío y por haber sido el hombre con más paciencia y amor en el mundo y haberse esforzado hasta el último momento de su vida para que su familia tenga lo que él no pudo tener.

A mi madre Lorena López por ser mi fortaleza y brindarme sus consejos cada día, sin su presencia en mi vida este logro no sería posible.

A mis hermanas Giselle y María Rosa por brindarme su tiempo y estar siempre cuando las necesito.

A mis familiares y amigos que sin esperar nada a cambio estuvieron acompañando en mis tristezas y alegrías. Y a todas aquellas personas que de alguna manera contribuyeron cosas buenas en mi vida.



### **Agradecimiento**

A Dios por darme salud y la oportunidad de alcanzar uno de los objetivos en mi vida, por bendecirme con una familia maravillosa que me apoya en todo momento.

A mi papá Manuel Avila que, aunque ya no esté físicamente estoy segura de que me bendice y esta orgullo desde el cielo. A mi mamá Lorena López por la paciencia, consejos y amor para no desistir en cumplir mis metas.

A mi hermana Giselle y mi cuñado Abel por haberme regalado la luz que le faltaba a mi vida, mi pequeño Manuelito. A mi hermana María Rosa por acompañarme durante todas las madrugadas en la realización de este proyecto y hacerme reír con sus ocurrencias en mis momentos de estrés.

Gracias a mi maestro y amigo Ing. Javier Domínguez por haberme brindado su apoyo moral en el momento más difícil de mi vida y por ser un apoyo fundamental durante todos estos años de mi carrera universitaria.

También me parece correcto agradecer a mi tutor y a mi revisor por su paciencia y por compartir sus conocimientos profesionales.

A todos mis maestros que durante todo este tiempo dedicaron sus esfuerzos y parte de sus conocimientos. Familiares y amigos más cercanos por su apoyo durante este trayecto en mi vida.

## Índice general

N°	Descripción	Pág
	Introducción	1

### Capítulo I El problema

N°	Descripción	Pág.
1.1	Planteamiento del problema	3
1.2	Situación de conflicto.	4
1.3	Delimitación del problema	4
1.3.1	Aspecto metodológico	4
1.3.2	Delimitación temporal	5
1.4	Formulación del problema	5
1.5	Justificación e importancia	5
1.6	Objetivos generales y específicos	6
1.6.1	General	6
1.6.2	Específicos	6
1.7	Hipótesis prospectiva	6
1.8	Variables	6
1.8.1	Variable independiente	6
1.8.2	Variable dependiente	6
1.8.3	Conceptualización y operacionalización de las variables	6
1.8.4	Indicadores de las variables.	7
1.9	Alcance	8

### Capítulo II Marco Teorico

N°	Descripción	Pág.
2.1	Antecedentes	9
2.1.1	Fútbol en el ecuador	9
2.1.2	Historia	9
2.1.3	Ecuador en el Mundial Corea-Japón.	10
2.1.4	Historia del Campeonato Ecuatoriano del Futbol.	11

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
2.1.5	Clubes de futbol de la serie A en el Ecuador.	12
2.1.5.1	Equipos con mayor popularidad.	14
2.2.	Marco Teórico	15
2.2.1	Antecedentes Nacional.	15
2.2.2	Antecedente Internacional.	16
2.2.3	Artículos basados en Minería de datos, Análisis de sentimientos.	17
2.3	Fundamentación Teórica.	19
2.3.1	Las Redes Sociales y el Fútbol.	19
2.3.2	Twitter	20
2.3.2.1	Api Twitter	20
2.3.2.2	Limitaciones de Twitter y su API.	21
2.3.3	Análisis de sentimiento en Twitter.	22
2.3.3.1	Aplicación del Análisis de sentimiento.	22
2.3.3.2	Polaridad	23
2.3.4	Procesamiento del lenguaje natural.	23
2.3.4.1	Lenguaje natural	23
2.3.4.2	Importancia del procesamiento del lenguaje natural.	24
2.3.4.3	El Pln y sus niveles de análisis.	24
2.3.4.4	Aplicación del PNL.	24
2.3.4.5	PNL aplicado al análisis de sentimientos.	25
2.3.4.6	Tokenización.	25
2.3.5	Minería de datos	25
2.3.5.1	Evolución de la minería de datos	25
2.3.5.2	Aplicación de la minería de datos.	26
2.3.5.3	Técnicas de la minería de datos.	27
2.3.5.4	Metodologías implementadas en la minería de datos.	28
2.3.5.4.1	Metodología KDD	28
2.3.5.4.1.2	Fases de la metodología KDD	28
2.3.5.4.2	Metodología CRISP-DM	31
2.3.5.4.3	Metodología SEMMA	33
2.3.5.5	Análisis comparativo de las metodologías KDD, CRISP-DM, SEMMA.	35
2.3.5.6	Herramientas para la minería de datos.	36
2.4	Definiciones Conceptuales.	37

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
2.4.1	Base de datos	37
2.4.2	Knime	37
2.4.3	R	38
2.4.4	Orange	38
2.4.5	Weka	38
2.4.6	Rapidminer	39
2.4.7	GNU octave	39
2.4.2	Python	39
2.4.3	Jupyter Notebook	40
2.4.4	Anaconda	40
2.4.5	Developer Twitter	40
2.4.6	Api Twitter	41
2.4.7	Token	41
2.5	Fundamentación legal	42

### **Capítulo III**

#### **Desarrollo de la propuesta**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
3.1	Metodología del proyecto.	43
3.1.1	Diseño de la investigación.	43
3.1.1.1	Enfoque cuantitativo.	43
3.1.1.2	Diseño cuasi-experimental.	43
3.1.1.3	Enfoque Longitudinal.	44
3.1.1.4	Diseño Documental.	44
3.2	Tipos de investigación.	44
3.2.1	Investigación exploratoria.	44
3.2.2	Investigación descriptiva.	45
3.2.3	Investigación cuasiexperimental.	45
3.3	Técnicas de investigación.	45
3.3.1	Técnica de muestreo.	45
3.4	Metodología de desarrollo.	46
	Anexos	49
	Bibliografía	92

**Índice de tablas**

<b>Nº</b>	<b>Descripción</b>	<b>Pág.</b>
1.	Definición de variables.	6
2.	Indicadores de las variables	7
3.	Fechas importantes en el fútbol del ecuador	9
4.	Trabajos nacionales referentes al tema.	16
5.	Trabajos internacionales referentes al tema.	17
6.	Artículos científicos referentes al tema.	18
7.	Características de la api.	21
8.	Áreas de aplicación de la minería de datos.	26
9.	Comparativa de las metodologías.	35
10.	Herramientas para la minería de datos.	36
11.	Hardware y software utilizados.	48
12.	Librerías utilizadas en el trabajo de investigación.	51
13.	Porcentajes concatenados de los cuatro equipos	73

## Índice de figuras.

<b>Nº</b>	<b>Descripción</b>	<b>Pág.</b>
1.	Ecuador al mundial 2002	11
2.	Fechas históricas en el fútbol ecuatoriano,	12
3.	Equipos de la serie a,	13
4.	Gráfico de barras con porcentajes de los quipos más populares	14
5.	Logo de los equipos	14
6.	Esquema para detectar la polaridad en las opiniones	23
7.	Evolucion de la mineria de datos	26
8.	Descripción de las técnicas de la minería de datos	27
9.	Fases de la metodología kdd	30
10.	Explicación de las fases de la metodología	31
11.	Crisp-dm estructura de los 4 niveles	32
12.	Fases de la metodología crisp.	32
13.	Fases metodología semma	34
14.	Iteración del proceso semma	34
15.	Base de datos	37
16.	Logo knime	37
17.	Logo de r	38
18.	Logo de orange	38
19.	Logo weka	38
20.	Logo rapidminer	39
21.	Logo octave	39
22.	Logo python	39
23.	Logo jupyter	40
24.	Logo anaconda	40
25.	Interfaz api	41
26.	Logo developer	41
27.	Token	41
28.	Fases metodología aplicada	47
29.	Esquema general del análisis de sentimientos	49
30.	Identificación de datos	54

31.	Credenciales de la api	54
32.	Tweets extraídos de la red social	55
33.	Crear carpeta donde se almacena el csv	55
34.	Excel generado con los tweets	56
35.	Lectura del csv	56
36.	Codificación para la limpieza del texto	57
37.	Función stopwords para la eliminación de palabras	58
38.	Dataset del texto limpio	58
39.	Linea de código de la traducción de tweets	59
40.	Dataset con la columna tweet_en	59
41.	Crear el csv con los tweets traducidos	60
42.	Lectura de palabras que expresen sentimientos	60
43.	Importacion de la libreria a usar para el analisis de sentimientos.	61
44.	Ejecución del código para el análisis de polaridad	61
45.	Dataset con los resultados de la polaridad en cada tweet	62
46.	Crear nubes de palabras	63
47.	Nube de palabras	63
48.	Tweets extraídos del equipo de barcelonasc	65
49.	Análisis de polaridad	65
50.	Dataset con resultados de polaridad tweets barcelonasc	66
51.	Tweets del equipo de csemelec	67
52.	Dataset con la polaridad de los tweets del equipo de csemelec	69
53.	Tweets del equipo de ldu	69
54.	Dataset con los resultados de polaridad del equipo ldu	71
55.	Tweets del equipo independiente del valle	71
56.	Dataset con los resultados de polaridad de los tweets del equipo idv	73
57.	Nube de palabras positivas y negativas equipo barcelona	74
58.	Nube de palabras positivas y negativas del equipo de emelec	75
59.	Nube de palabras positivas y negativas del equipo ldu	75
60.	Nube de palabras positivas y negativas del equipo de idv	75

## Índice de gráficos

N°	Descripción	Pág.
1.	Gráfico de barras con la polaridad de los tweets.	62
<u>2.</u>	Diagrama pastel con la polaridad de tweet.	63
3.	Gráfico de barras equipo de barcelonasc.	66
4.	Diagrama pastel equipo barcelonasc.	67
<u>5.</u>	Gráfico de barras del csemelec.	68
6.	Diagrama pastel del csemelec.	68
7.	Diagrama de barras del equipo ldu.	70
8.	Diagrama pastel con la polaridad del equipo ldu	70
9.	Diagrama de barras del equipo idv	72
10	Diagrama pastel con la polaridad del equipo idv.	72
11.	Diagrama de barras con los datos de los 4 equipos	74



**Índice de anexos**

<b>Nº</b>	<b>Descripción</b>	<b>pág.</b>
1.	Ley de proteccion de datos personales	78
2.	Decreto ejefutivo	82
3.	Pasos para crear cuenta en el developer de twitter	83
4.	Código de programacion	87



**ANEXO XIII.- RESUMEN DE TRABAJO DE  
TITULACIÓN (ESPAÑOL)  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

---



**“USO DE MINERÍA DE DATOS PARA ANÁLISIS DE SENTIMIENTOS EN TWEETS  
DEL FÚTBOL ECUATORIANO”**

**Autor:** Avila López Merly Carolina.

**Tutor:** Ing. Comp. Plaza Vargas Angel Marcel, Mg.

**Resumen**

En este trabajo se procederá a realizar un análisis de sentimiento en tweets, creando algoritmos en Jupyter Notebook que ayuden a la extracción, recolección, selección, limpieza y validación de datos que permita aplicar la minería de datos mediante dicho análisis de sentimientos. El uso masivo de las redes sociales en la actualidad genera una cantidad de datos enormes, la idea principal de este análisis es clasificar y determinar la polaridad (positivos, negativos, neutros) de cada Tweets realizados por usuarios acerca de cuatro equipos del fútbol ecuatoriano; Barcelona Sporting Club, Club Sport Emelec, Liga de Quito Universitaria e Independiente del Valle, este análisis de sentimiento aplicado en este ámbito deportivo, tiene la finalidad de crear estrategias para el marketing de los equipos, ayudando de esta manera a fortalecer su marca o nombre, a saber y tener conocimiento acerca de lo que piensan sus seguidores y en algunos casos ajustar o cambiar acciones basados en las preferencias de sus hinchas.

**Palabras claves:** Tweets, Análisis de sentimiento, hinchas, procesado de datos, API, polaridad, lenguaje natural, Jupyter.



**ANEXO XIV.- RESUMEN DE TRABAJO DE  
TITULACIÓN(INGLÉS)**



**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

---

**“USING DATA MINING FOR THE ANALYSIS OF SENTIMENT ON TWEETS ABOUT THE ECUADORIAN FOOTBALL”**

**Author:** Avila López Merly Carolina.

**Advisor:** Ing. Comp. Plaza Vargas Angel Marcel, Mg.

**Abstract**

In this work we will proceed to perform a sentiment analysis on tweets, creating algorithms in Jupyter Notebook that help the extraction, collection, selection, cleaning and validation of data that allows the application of data mining through sentiment analysis. The massive use of social networks nowadays generates a huge amount of data, the main idea of this analysis is to classify and determine the polarity (positive, negative, neutral) of each tweet made by users about four Ecuadorian soccer teams; Barcelona Sporting Club, Club Sport Emelec, Liga de Quito Universitaria and Independiente del Valle, this sentiment analysis applied in this sporting environment, aims to create strategies for marketing teams, thus helping to strengthen its brand or name, to know and have knowledge about what their followers think and in some cases adjust or change actions in the preferences of their fans.

**Keywords:** Tweets, Sentiment analyses, fans, data processing, API, polarity, Jupiter, Natural language.

## **Introducción**

Hoy en día, el aumento de la interacción a través de las redes sociales brinda a los usuarios la oportunidad de compartir sus opiniones sobre cualquier tema favorito, ya sea un producto o una marca y a través de esto, las empresas pueden medir su impacto en la sociedad.

Las redes sociales han cambiado la forma en que nos relacionamos con nuestro entorno, sobre todo porque han brindado un espacio en el que las personas pueden expresar libremente sus sentimientos y opiniones sobre temas de diversa índole que tiene un impacto significativo en la forma en que se comunica la información o los mensajes.

La red social Twitter es una de las redes más populares que permite enviar mensajes SMS cortos de 280 caracteres llamados tweets. El uso principal de los tweets es para que los usuarios expresen sus opiniones mediante estas publicaciones. Estos son muy útiles, ya que se utilizan en muchas encuestas como análisis de sentimientos en este caso.

La información seleccionada de esta manera se conoce como Minería de Opinión y es por ello por lo que la mayoría de las empresas conocen las ideas de los usuarios sobre el producto y así obtienen más ventajas competitivas en su campo de actividad o en diferentes áreas.

Para desarrollar el análisis de sentimiento en tweets del fútbol ecuatoriano se aplicará la técnica de minería de datos. Haciendo uso de metodologías en fases: Identificando, filtrando y realizando la validación y visualización de datos. Para acceder a dichos datos de la red social se utilizará la API de Twitter que utiliza el lenguaje de programación Jupyter con Python utilizando varias librerías que permiten la minería de datos y así poder realizar análisis de minado de datos y presentar los resultados a través de gráficos.

Se detalla a continuación la explicación de cada capítulo como finalidad de llegar al cumplimiento de cada uno de los puntos de la propuesta del proyecto.

Capítulo I Este capítulo está compuesto por los siguientes puntos: Planteamiento del problema, situación de conflicto, delimitación del problema, aspecto metodológico, delimitación temporal, formulación del problema, justificación e importancia, objetivos, variable dependiente e independiente, operacionalización de variables y alcance.

Capítulo II Se realizará el marco teórico citando los siguientes puntos: Los antecedentes de la investigación, fundamentación teórica, definición conceptual y como punto final se desarrollará la fundamentación legal.

Capítulo III En este capítulo se realiza el desarrollo de la propuesta, describiendo la metodología aplicada para el análisis y recopilación de información que será utilizada en el presente estudio. Se expondrán los enfoques de investigación, técnicas que fueron de gran utilidad para la recolección de los datos, las estrategias empleadas para obtener y procesar la información de este trabajo.

# **Capítulo I**

## **El problema**

### **1.1 Planteamiento del problema**

La necesidad que tienen las personas de expresarse de alguna forma y especial en temas que le llaman la atención como los deportes o la música, en este caso se va a escoger el fútbol ecuatoriano.

Actualmente las organizaciones deportivas, disponen del uso de redes sociales como una herramienta de comunicación y marketing, ya que es el medio por el cual pueden atraer de manera más eficaz a la audiencia. Las redes sociales cumplen un rol importante permitiendo al público, dialogar de modo directo con dirigentes de su club favorito. Sirviendo de gran ayuda para así a determinar la popularidad de un club mediante esta red social.

Los deportes en este caso el fútbol mueve una cantidad considerable de personas, haciendo notar la emoción que tienen cada juega su equipo del que son hinchas ya que no solo demuestran su apoyo asistiendo a estadios sino de forma virtual haciendo uso de redes sociales, debido a que el internet es el medio más eficaz de hacer llegar una opinión o comentario.

Un Análisis de Sentimientos hace referencia a métodos que ayudan a la extracción e identificación de la información existente en el mundo del internet. Este análisis ayuda a determinar si la información que se seleccionó en las redes sociales contiene opiniones negativas, positivas o neutras. Con un Análisis de Sentimiento, se pretende entender cuál es la intención apropiada de un mensaje u opinión. Mejorando las estrategias empresariales y facilitando la gestión de una reputación en línea, llevando a cabo un plan estratégico de marketing en línea.

El trabajo de organizar un escrito en lenguaje natural en un sentimiento según su polaridad tiende a ser muy complicado. Ya sea porque la interpretación personal de un usuario es diferente de la de los demás, la tarea se hace un poco más difícil si el texto es muy corto o no haya sido escrito de forma correcta, como lo es en los mensajes de la red social Twitter.

## **1.2 Situación de conflicto.**

El uso de las redes sociales ha sido el medio principal de las personas donde pueden expresar lo que piensan sin ninguna prohibición, ya sean opiniones buenas o malas. generando así una gran cantidad de datos. En donde la mayoría de estos no han sido procesados a través de un análisis y conforme a esto, causan una pérdida a la identificación de nuevas características basadas en comentarios para definir el sentimiento de una persona, permitiendo de esta manera buscar soluciones o notar por medio de terceras personas que cosas deben mejorar o que estrategias deben pulir tanto en los jugadores o equipos de serie A dentro del fútbol ecuatoriano.

## **1.3 Delimitación del problema**

La implementación de este análisis de sentimientos a usuarios se llevará a mediante tweets, El análisis consiste en lograr entender una frase exacta, desarrollando así mejores estrategias, toma de decisiones y facilitar la reputación online. Para esto se aplican técnicas de minería de datos que ayudan a conocer la perspectiva en base a la opinión de los usuarios en las redes sociales, se consultara una base de datos obtenidas de Twitter para luego ser procesada por varias herramientas necesarias y haciendo uso de programas que ayudaran a la filtración de la información.

Se basa en determinar qué es lo que se va a realizar, mediante lo cual se indica hasta qué punto el investigador abarca el problema a investigar.

Área: Equipos de fútbol de la Serie A en Ecuador.

Aspecto: Minería de Datos aplicado a Análisis de sentimientos en Twitter.

### **1.3.1 Aspecto metodológico:**

- 1) Se revisará inicialmente los tweets de todos los equipos de fútbol de la Serie A en Ecuador
- 2) Se Hará uso de las páginas oficiales del club con más tendencia en el medio.
- 3) Se determinará un grupo de 4 equipos con mayor popularidad en el fútbol ecuatoriano.
- 4) Se desarrollará el análisis sobre los Tweets relacionados a los equipos seleccionados.

### **1.3.2 Delimitación temporal.**

El análisis de sentimiento que se realizará en Twitter será durante un periodo de tiempo determinado de 3 meses, haciendo uso de la técnica de minería de datos.

### **1.4 Formulación del problema.**

Al tomar como referencia el planteamiento de problema surgen la siguiente interrogante:

¿Es posible determinar la popularidad de un club deportivo mediante un análisis de sentimiento en Twitter?

¿Cuáles son las herramientas adecuadas para realizar una extracción de datos en redes sociales?

### **1.5 Justificación e importancia.**

El internet se ha vuelto una parte fundamental en algunos aspectos de la vida, como lo es en el fútbol. El fanatismo y la pasión con la que un aficionado vive su relación con sus equipos y jugadores preferidos, no se compara con muchas cosas. No solo con ir a los estadios o ver los partidos por TV, sino que ha llegado al terreno de las redes sociales con un gran impacto. Trayendo como ventaja conocer la percepción de su equipo, haciendo mención si el desempeño fue optimo si esté está contento o no con el resultado de su último partido.

Realizando el uso correcto de las herramientas y tecnologías necesarias que ayuden al análisis de una cantidad considerable de información que almacenada dentro de la red social. Con la herramienta seleccionada se pretende realizar una extracción de datos en Twitter a cerca de opiniones, tweets positivos o negativos emitidos por aficionados de los equipos de su preferencia. Una vez obtenidos los datos se procede a aplicar la técnica de minería de datos haciendo uso del lenguaje de programación Python y sus librerías para realizar dicho análisis.

Al analizar el sentimiento en tweets se hace mención de cómo se sienten las personas en el medio digital. En vez de realizar una estadística sencilla post o mensajes, abarca los sentimientos y opiniones. El desarrollo de este análisis implica compilar y analizar información de las publicaciones que los usuarios comparten sobre un tema en específico en la red social. Ayudando a determinar la reputación de los jugadores de cada club, mejorando



la experiencia dentro de su estadio y lo más importante tener a la competencia en el radar, ya sea para comparar el rendimiento o la medición de jugadores.

## 1.6 Objetivos Generales y específicos

### 1.6.1 General

Desarrollar un análisis de sentimientos de datos obtenidos mediante Tweets relacionados al fútbol ecuatoriano.

### 1.6.2 Específicos

- Seleccionar las herramientas de acuerdo con un análisis inicial que permita la búsqueda de datos de los equipos en Twitter.
- Especificar la herramienta para la extracción de datos.
- Determinar metodologías para analizar el sentimiento.
- Mostrar los resultados obtenidos del análisis realizado mediante gráficos.

## 1.7 Hipótesis prospectiva

¿Es posible realizar un análisis de sentimiento del fútbol ecuatoriano mediante tweets, haciendo uso de minería de datos?

## 1.8 Variables

### 1.8.1 Variable independiente

- Perfil del aficionado en la red social.
- Datos generados por la red social Twitter.

### 1.8.2 Variable dependiente

- Particularidad de cada aficionado al expresar su opinión.

### 1.8.3 Conceptualización y operacionalización de las variables.

**Tabla 1.** Definición de Variables.

Independiente		Dependiente	
Variable	Perfil del aficionado	Datos obtenidos de la red social	Particularidad

	en la red social.	Twitter	de cada aficionado al expresar su opinión.
<b>Definición</b>	Conducta que tiene una persona al momento de expresarse de su equipo o jugador. Esto puede variar según la satisfacción que haya tenido al ver el fútbol.	Información obtenida de la red social que se encuentran a la vista de todos los usuarios ya sea expresando sus emociones u opiniones.	Cuando muestran su afición por el deporte permitiendo así definir su identidad online.
<i>Información Obtenida mediante investigación elaborada por Merly Avila López.</i>			

#### 1.8.4 Indicadores de las variables.

**Tabla 2.** Indicadores de las variables

Variables	Indicadores	Descripción	Tipo de Indicador
<b>Perfil del aficionado en la red social.</b>	Activo, edad, comenta sobre el equipo de su preferencia.	La forma en que comparte su idea u opinión.	Cualitativo.
<b>Datos obtenidos de la red social Twitter</b>	Conectados, dispositivos, follow, ciudad, tweet, retweet, user name.	Información general por extraer de la red social.	Cuantitativo.
<b>Particularidad de cada aficionado al expresar su opinión.</b>	Comparte su opinión de forma rigurosa acerca de su equipo en línea.	La manera en que un usuario escriba su comentario, ya sea positivo o negativo.	Cualitativo.
<i>Información Obtenida mediante investigación elaborada por Merly Avila López.</i>			

## 1.9 Alcance

- El presente trabajo se realizará en un periodo de tiempo limitado de 3 a 4 meses, mediante la extracción de datos.
- El análisis por realizar ayudará a conocer así el sentimiento del hincha ecuatoriano en redes sociales.
- Para esto se consultará una base de datos obtenida de Twitter y luego esta información será procesada por programas y herramientas que ayuden a la filtración de datos, opiniones más relevantes de los usuarios acerca del fútbol en el Ecuador y así conocer las características de este a la hora de expresarse por el medio digital.

## Capítulo II

### Marco Teórico

#### 2.1 Antecedentes

##### 2.1.1 Fútbol en el Ecuador.

##### 2.1.2 Historia.

La investigación realizada indica que la Selección Ecuatoriana es un equipo formado con jugadores de nacionalidad ecuatoriana y representa a la república del Ecuador y a la FEDER en las competencias realizadas por la Conmebol y la FIFA. Teniendo su primer enfrentamiento contra Bolivia en Bogotá en agosto de 1938.

Los primeros años se dice que fue una de las selecciones con menos productividad de América, En 1988 se realizó un cambio de jugadores por Dusan Draskovic y gracias a eso se inició un periodo de transformación dentro del futbol en el Ecuador consiguiendo quedar en 4to lugar en la Copa América en 1993. (Chato970, 2015)

El fútbol en el Ecuador tuvo cabida cuando el guayaquileño Juan Alfredo Wright volvió al Ecuador desde Inglaterra junto a su hermano Roberto Wrigth en 1899. Motivando a jóvenes de la ciudad a practicar el deporte.

Se enlistarán varias de las fechas importantes dentro del futbol.

**Tabla 3.** Fechas importantes en el Fútbol del Ecuador

Año	Actividad
1899	Fundación del primer club deportivo en Guayaquil, nombrado “Guayaquil Sport Club”
1900	Registro de los primeros encuentros en el futbol.
1902	Fundación del club Sport Ecuador.
1922	Fundación de la Federación Deportiva Guayaquil.
1925	Guayas y Pichincha organizan sus primeros encuentros.

1926	Organización de las olimpiadas Nacionales en Riobamba. Ecuador se afilia a la FIFA de manera oficial.
1927	Ecuador se une a la CONMEBOL
1950	Creación de los primeros equipos en Guayaquil (Emelec, Barcelona, Norte América, 9 de octubre, entre otros.)
1951	Inician los campeonatos provinciales en Guayaquil.
1953	Inicio de campeonatos provinciales en Pichincha.
1967	Finalizan los campeonatos provinciales.
1971	Se crea la Serie B como 2do nivel del futbol en el Ecuador, creándose así el nivel de ascensos en la serie A.
1978	Reformación de reglamentos el nombre de la federación Ecuatoriana de Fútbol cambia.
1989	El número de quipos en la primera división varia, entre 16 en un año y El torneo se reforma dejando 12 equipos en la serie A.

---

*Información obtenida mediante investigación realizada por Merly Avila.*

### **2.1.3 Ecuador en el Mundial Corea-Japón.**

Según (FEF, 2012) Fue en el 2011 cuando la Conmebol confió a la F.E.F la organización del Sudamericano “Juventud América”. Las ciudades Ambato, Cuenca, Guayaquil, Latacunga y Portoviejo se prepararon para recibir a las promesas del continente, en un torneo considerado como uno de los trascendentales en el ámbito de la Conmebol por actuar como medio internacional a las jóvenes estrellas surgidas de la cantera sudamericana.

La clasificación al mundial argentino llegó cuando la selección comenzaba a consolidarse rumbo a Corea-Japón 2002. Se ponía de manifiesto la superación del fútbol ecuatoriano. La gestión administrativa empezaba a rendir frutos.

Encarar al Sudamericano fue un verdadero reto para José María Andrade quien fue el encargado de dirigir al seleccionado que viajaría a Argentina con la misión de lograr uno de los cupos al mundial de Argentina.

El grupo E albergó al seleccionado junto a las selecciones de Etiopía, Costa Rica y Holanda. Un triunfo, un empate y una derrota hicieron que Ecuador avanzara a la ronda de Octavo de final.

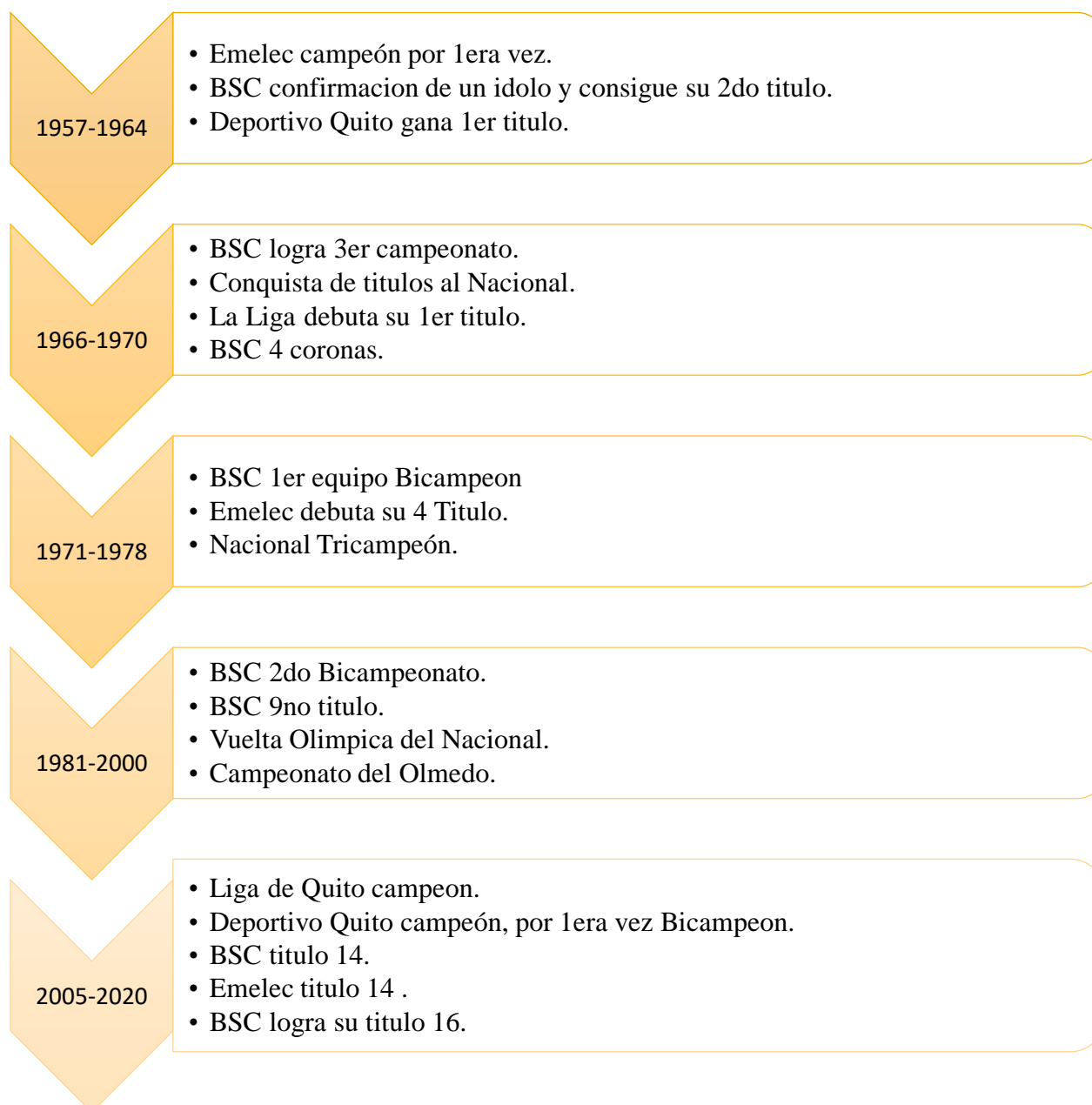
La selección de Ecuador estuvo conformada por figuras como Segundo Alejandro Castillo, Félix Borja, Jorge Guagua y Roberto Mina. Se completó con elementos también de enorme valía como: Daniel Viteri, Franklin Corozo, Walter Iza, Franklin Salas entre otros. Fecha histórica a la clasificación del mundial Corea-Japón 2002, 7 de noviembre del 2001.



*Ilustración 1. Ecuador al mundial 2002, Información obtenida mediante investigación. La historia del mundial contada con estampillas.*

### **2.1.4 Historia del Campeonato Ecuatoriano del Fútbol.**

Fue en el año 1957 que inicio este sueño unificador, llena de goleadores, técnicos con gran éxito y sobre todo campeones. A continuación, se muestra línea de tiempo del campeonato ecuatoriano.



*Ilustración 2. Fechas históricas en el fútbol ecuatoriano, Información obtenida mediante investigación realizada por Merly Avila.*

### 2.1.5 Clubes de futbol de la serie A en el Ecuador.

Conocida por su nombre popular LigaPro Betcris, en el Ecuador es el sistema de ligas de máxima categoría organizada hasta el 2018 por la F.E.F y en el 2019 por la liga de futbol profesional en el país.

Actualmente está entre las mejores ligas del mundo, teniendo el puesto 18 según la lista oficial de dicha institución en el año 2021. A lo largo de los años un total de 60 equipos han

sido participantes en la Serie A desde sus inicios en el año 1957, en los cuales 10 han sido campeones.

Según (LigaPro Betcris, 2022) la lista de los equipos de la Serie A son los siguientes.



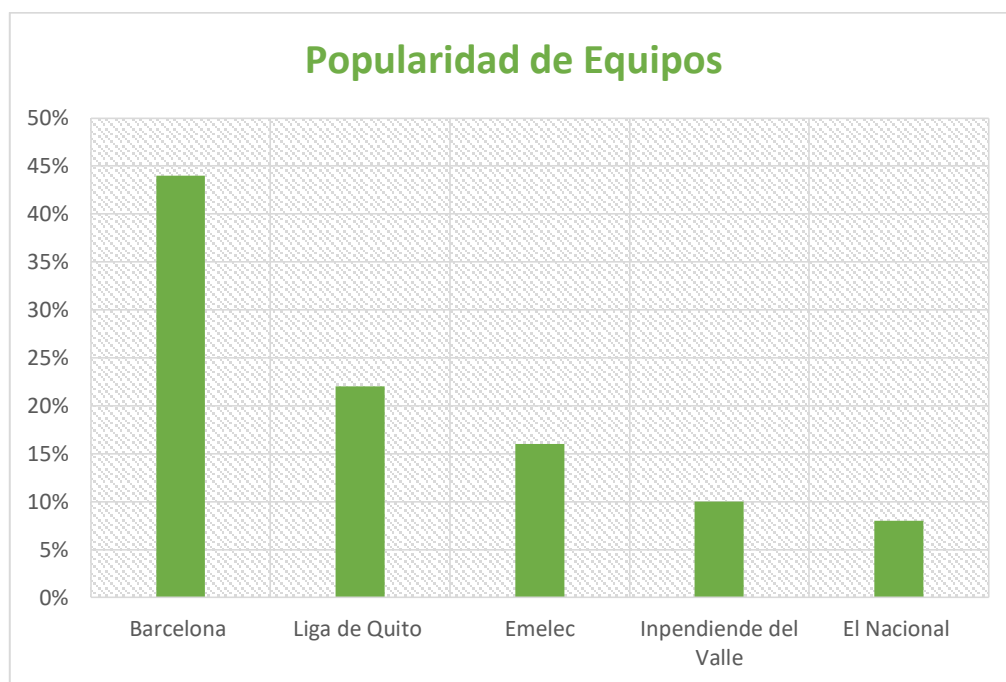
*Ilustración 3. Equipos de la Serie A, Información obtenida mediante investigación realizada por Merly Avila.*



### 2.1.5.1 Equipos con mayor popularidad.

En una encuesta realizada por la empresa líder mundial del Marketing Deportivo y mediciones estadísticas. Mostraron resultados, donde cuenta a Barcelona como el equipo que tiene más hinchada a nivel nacional con 44%, el 2do lugar de la lista lo tiene la Liga Deportiva Universitaria de Quito con un 22% y como 3er lugar Emelec con 16% siguiéndoles los equipos Independiente de Valle 10% y el nacional con 8%. (Gerardo Molina, 2020).

En la presente investigación se realizará un análisis de sentimientos a los cuatro equipos principales, quedando enlistado de la siguiente manera: Barcelona, Liga Deportiva Universitaria de Quito e Independiente del Valle.



**Ilustración 4.** Gráfico de barras con porcentajes de los quipos más populares, información obtenida del diario el telégrafo, realizado por Merly Avila.



**Ilustración 5** Logo de los equipos, información obtenida mediante investigación realizada por Merly Avila.

## 2.2. Marco Teórico

### 2.2.1 Antecedentes Nacional.

(Alvarado, 2016) En su trabajo *“Análisis de sentimiento en Twitter de los socios de un club de fútbol a través de la evaluación de herramienta que manejan gran volumen de información”*. Realiza un análisis de sentimientos en Twitter, en donde hace uso de técnicas que permite la extracción de información en grandes volúmenes, teniendo como objetivo la polaridad de la información que benefician la campaña para la atracción de nuevos socios al Barcelona SC. Los comentarios de los usuarios serán de gran ayuda para la elaboración de plan para nuevos socios. En este proyecto Alvarado hizo un estudio a través de Twitter con la nueva campaña para atraer socios haciendo uso de Hadoop con palabras claves y de esta manera identificar cuál de las fases mostradas a los socios tenía mejor recepción, y de esta puedan elegir cuál de estas fases se adecúa a su economía y así poder acceder a su plan como nuevo socio. Hadoop fue la herramienta a utilizar para analizar la información.

(Vega, 2020) En su trabajo *“Análisis de Sentimientos en Twitter para descubrir contenido xenófobo hacia los inmigrantes venezolanos en Ecuador”* tiene como propósito establecer la presencia de temas xenófobos en una recopilación de tweets, alusivo a los inmigrantes de Venezuela en Ecuador, este trabajo se realizó a través de fases de la metodología KDT, a su vez hizo uso del lenguaje de programación Python, con las librerías más destacadas como NLTK, y la aplicación de producción de un texto en un idioma natural para la traducción de los tweets.

(Narváez, 2017) es su trabajo titulado *“Aplicación de técnicas de minería de datos para el análisis de sentimientos en las redes sociales sobre productos de la marca fideos Cayambe”* Realizó una investigación de las técnicas de extracción de información, en donde le permita aplicar un análisis de sentimientos de la marca fideos Cayambe, los datos son extraídos mediante Facebook, los resultados extraídos serán de gran ayuda para la medición del grado de relación y opinión que tienen los seguidores con respecto a la marca de dicho producto. Hizo uso de la metodología de minería de datos CRIPS-DM. Para la aplicación de las técnicas seleccionadas, que procesarán los algoritmos, serán usadas herramientas de libre acceso, que cuentan con las características necesarias para su requerido uso.

Estos trabajos de investigación se asemejan a la investigación planteada, ya que trata de analizar los sentimientos, emociones y comportamiento de usuarios, sobre un tema de interés público. Teniendo como finalidad la obtención de datos y que estos puedan ser implementados en cualquier aspecto que sea requerido.

**Tabla 4.** Trabajos nacionales referentes al tema.

Referencia	Tema	Metodología	Técnica	Herramienta
(Alvarado, 2016)	Análisis de sentimiento en Twitter de los socios de un club de fútbol a través de la evaluación de herramienta que manejan gran volumen de información	No menciona.	Recopilación (Gestión de clústers)	Apache (framework Apache Spark, API Twitter)
(Vega, 2020)	Análisis de Sentimientos en Twitter para descubrir contenido xenófobo hacia los inmigrantes venezolanos en Ecuador	metodología CRISP-DM KDT.	Recopilación documental, Entrevista.	Spyder IDE Google Colab
(Narváez, 2017)	Aplicación de técnicas de minería de datos para el análisis de sentimientos en las redes sociales sobre productos de la marca fideos Cayambe.	CRISP-DM	árboles de clasificación, clustering y regresión lineal.	DSS(Decision Support Systems), OLAP (On-Line Analytical Processing)

Información obtenida mediante investigación. Esquema realizado por Merly Avila.

### 2.2.2 Antecedente Internacional.

(Ferreyra, 2020) En este trabajo “Análisis de sentimiento en Tweets de Fútbol Argentino”, el autor presenta un diseño de todas las etapas realizadas. Haciendo una recolección de tweets incluyendo la particularidad del español de su país Argentina, ya sean expresiones, palabras, dialectos propios de las personas. Realiza un filtrado de tweets se crea una guía para que se puedan separar tweets y que queden aquellos que hablan del fútbol argentino. Para el etiquetado de polaridad, diseñó una guía con criterios exactos, usando varios aspectos del lenguaje, para así poder identificar de mejor forma la polaridad de dichos tweets. En el proceso de etiquetado siguió con la forma semi automática seguida por la SEPLN, destacando las etiquetas de los tweets que se extrajeron fueron netamente puestas por usuarios en dicha red social.

(Montesinos, 2014) En este trabajo “Análisis de sentimientos y predicción de eventos en Twitter”, tiene como idea hacer uso de métodos que permitan realizar este análisis y en este caso se despliega un caso implementando parte de estas técnicas. La red social que se escogió para

realizar este trabajo es Twitter ya que en Chile se usa mucho esta app. Su idea principal fue enfocarse en las elecciones presidenciales primarias realizadas en la Alianza por Chile y sus candidatos el objetivo es predecir los resultados, analizando quien está a favor de un partido en específico o en contra de los candidatos. Para la obtención de resultados de la opinión de los usuarios.

(Córdova, 2019) En su tesis *“Metodología basada en Minería de Datos para la detección de usuarios Influencers en Twitter”* plantea un método para el Descubrimiento de usuarios Influencers en Twitter, siguiendo el proceso esencial de Minería de Datos, haciendo uso de diversas técnicas de recolección de datos y utilizando un conjunto de métricas de influencia para obtener una puntuación de influencia para cada usuario, que permita conseguir un ranking de los usuarios más influyentes para un público objetivo determinado. Estos trabajos de investigación tienen una similitud con la investigación en curso, ya que su objetivo es analizar sentimientos ya sean positivos o negativos de usuarios en Twitter y el uso de la Minería de Datos.

**Tabla 5.** Trabajos internacionales referentes al tema.

Referencia	Tema	Metodología	Técnica	Herramienta
(Ferreira, 2020)	Análisis de sentimiento en Tweets de Fútbol Argentino.	No menciona.	Gestión de clústers	Googletrans
(Montesinos, 2014)	Análisis de sentimientos y predicción de eventos en Twitter.	No menciona el autor.	Recopilación de datos.	API twitter, SQL, Perl.
(Córdova, 2019)	Metodología establecida en Minería de Datos para la detección de usuarios Influencers en Twitter.	CRISP-DM KDD	Análisis de algoritmos.	API twitter, web scraping.

Información obtenida mediante investigación. Esquema realizado por Merly Avila.

### 2.2.3 Artículos basados en Minería de datos, Análisis de sentimientos.

Los artículos por mencionar son trabajos de distintos países teniendo como finalidad un estudio al análisis de sentimiento en redes sociales en distintas áreas haciendo uso de la minería de datos.

(Rodriguez & Haber, 2020) En su artículo *“Análisis de sentimientos en Twitter aplicado al #impeachment de Donald Trump”*. Aborda los sentimientos que circulan en Twitter sobre *impeachment* de Donald Trump. La implementación de un enfoque mixto de naturaleza descriptiva arroja resultados sobre los aspectos emocionales relevantes para el discurso de los usuarios estadounidenses en público. Comprender el sentimiento relevante contenido en los tweets es una estrategia multidisciplinaria. Aplicamos técnicas de minería de datos y sus tipologías, así como análisis de discurso multimodal para extraer y evaluar valores de sentimiento. Proponemos tener en cuenta todo el soporte, código y prácticas de microblogging.

(Blasco & Coenders, 2020) *“Análisis de sentimiento de la agenda de los partidos políticos españoles en Twitter durante la Moción de Censura de 2018. Un enfoque de datos composicionales”* Los autores realizan este artículo con el fin de efectuar un análisis de sentimiento en el área de la política, haciendo uso de publicaciones de cuentas oficiales de las alineaciones políticas que representa el congreso. Aplica un análisis que está compuesto por datos y visualizando la información a través de Biplot. Muestra la veracidad de ciertos temas en la agenda y agrega un tercer componente como lo es analizar el sentimiento.

(Córdova, 2019) en su trabajo *“Técnicas para el Análisis de Sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength”* describe dos tipos de técnicas. La primera está basada en procesos de Aprendizaje Automático Supervisado. Para la aplicación de esta se necesita herramientas del Procesamiento de Lenguajes Naturales y como un punto de partida un corpus clasificado. La segunda está basada en diccionarios de polaridad. Hace uso de la herramienta de SentiStrength, que está siendo muy usada para estudios de Tweets en inglés. Señala ventajas y limitaciones de cada uno de estos enfoques para su aplicación a la investigación en comunicación política.

**Tabla 6.** Artículos científicos referentes al tema.

Referencia	Tema	Metodología	Técnica	Herramienta
------------	------	-------------	---------	-------------

---

(Rodriguez & Haber, 2020)	Análisis de sentimientos en Twitter aplicado al #impeachment de Donald Trump.	Análisis del discurso multimodal.	Clustering, Análisis Cualitativo y Cuantitativo.	Api rest, twitter, Stela.
(Blasco & Coenders, 2020)	Analizar el sentimiento en la agenda de los partidos políticos de España en Twitter durante la Moción de Censura de 2018. Un enfoque de datos composicionales	Metodología estadística estándar.	Recopilación de datos.	Biplot composicional, API twitter,
(Córdova, 2019)	Métodos para Analizar el sentimiento en Twitter: Aprendizaje automático Supervisado y SentiStrength	No menciona	Algoritmos de aprendizaje automático supervisado	SentiStrength.

---

*Información obtenida mediante investigación. Esquema realizado por Merly Avila.*

## 2.3 Fundamentación Teórica.

### 2.3.1 Las Redes Sociales y el Fútbol.

Las redes sociales es el medio que se usa en la actualidad para interactuar de forma virtual con otras personas, ayudan a expresar gustos, pensamientos, opiniones o sentimientos puede ser aplicada en varios campos como lo son; La educación, investigación, ciencia, marketing digital y deportes.

Un punto muy importante en el futbol es el fanatismo, ese grupo de personas que comparten creencias, jergas específicas y valores como una forma de expresarse.

Los clubes deportivos mantienen una presencia permanente en el mundo de la comunicación, dándole mayor cabida a las redes sociales, haciéndolas una fuente principal para la comunicación. Son de mucha ayuda para abarcar información sobre la entidad a

través de este medio permitiendo a usuarios estar al tanto de su equipo e interactuar con grupos de personas con su misma afinidad.

Las redes sociales contienen informaciones importantes, que mediante las API de cada una se puede extraer y analizarlos haciendo uso de la técnica minería de datos. En la actualidad una de la red social con más uso es Twitter, a continuación, se detallará sobre la misma.

### **2.3.2 Twitter**

Twitter es una red social gratuita que admite compartir textos de 280 caracteres estos pueden contener fotos, ideas, enlaces, popularmente llamados Tweets, sus usuarios se mantienen conectados al publicarlos, son visibles para cualquier usuario a menos que estos se encuentren protegidos. Al responder o compartir un tweet los usuarios interactúan entre sí en tiempo real sobre contenidos de su preferencia. Se puede acceder a la aplicación ya sea desde un computador o un dispositivo móvil.

Hoy en día Twitter cuenta con un total de 1.3 millos de cuentas en Twitter, con una cantidad de 500 tweets diarios. (Ahlgren, 2022)

Los usuarios se identifican con un nombre de usuario y estos pueden mencionar a sus seguidores usando el símbolo de arroba y para etiquetar el símbolo del hashtag. También comparte información de una forma más amplia ya sea a empresas, desarrolladores o aquellos usuarios que tienen acceso programático a los datos de la aplicación a través de su API.

#### **2.3.2.1 Api Twitter**

Una API es la forma en la que los programas informáticos solicitan la información y esta es proporcionada por la aplicación Twitter. La API de Twitter se puede utilizar para recuperar y analizar datos de Twitter de forma programática, así como para generar conversaciones en Twitter.

A lo largo de los años, la API de Twitter ha crecido al agregar niveles adicionales de acceso para que los desarrolladores e investigadores académicos puedan escalar su acceso para mejorar e investigar la conversación pública. Recientemente, se lanzó la API de Twitter v2. La API v2 de Twitter incluye una base moderna, funciones nuevas y avanzadas y una incorporación rápida al acceso esencial. Cuando se desea acceder a la API primero se debe registrar una aplicación. Esta solo permitirá obtener la información que es de origen público en Twitter, para así acceder a mensajes directos ya sea de los que reciben o envían deben

tener permisos adicionales, pero estos no son otorgados de forma predeterminada. (DeveloperTwitter, 2022)

La API de Twitter tiene cinco grupos de puntos de conexión que son:

- Cuentas y usuarios
- Tweets y respuestas
- Mensajes directos
- Anuncios
- Herramientas y SDK de editor.

**Tabla 7.** Características de la API.

API TWITTER			
¿Como acceder a la API?	¿Cómo hacer la primera solicitud en API?	Niveles de acceso y versiones de la API.	Formas de pago.
Se debe obtener una cuenta en el desarrollador.	Identificar que uso le quiere dar.	Access to Twitte API v2.	Algunas de las versiones se encuentran
Las claves y token de la aplicación. (API key and secret, Fichas de acceso de usuario, token de acceso a la aplicación.	Elegir una herramienta para realizar la solicitud. (Postman, Sample code). Verificar que tenga respuesta de la solicitud.	Access to standard v1.1. Access to premium v1.1. Access to Enterprise.	disponible de forma gratuita y otra de paga como en el caso de la versión
Se debe realizar la primera solicitud.	Ajusta la solicitud usando parámetros.	Access to the Ads API	premium o Access to Enterprise.

*Información obtenida mediante investigación realizada por Merly Avila.*

### 2.3.2.2 Limitaciones de Twitter y su API.

En Twitter se puede realizar un aproximado de 1000 actualizaciones por día, incluyendo a los tweets y retweets, limita a que el usuario envíe 250 mensajes por día. Pone límites a las cuentas de los usuarios por varias razones, con el fin de ayudar a prevenir spam y abuso. Estas limitaciones son cada hora, se puede actualizar algunas veces al día, pero el



usuario puede ser bloqueado de forma temporal de la actualización, si esta es realizada muchas veces en una hora y cuando esto sucede, solo tienen que esperar entre 2 o 24 en la cuenta para poder ser desbloqueado. (DeveloperTwitter, 2022)

Según la API de Twitter esta puede tener máximo 350 solicitudes por hora y esto es para aplicaciones que han sido autenticadas de no ser así, solo pueden realizar 150 solicitudes de API en la solicitud incluyen acceso a tweets, retweets y mensajes directos.

### **2.3.3 Analizar los sentimientos en Twitter.**

El análisis de sentimiento ayuda a clasificar opiniones, comentarios, ideas en función a la polaridad con que estos sean expresados por un usuario. El análisis de sentimiento se realiza con la finalidad de determinar si el público objeto acoge una marca, basándose en una evaluación automatizada de comentarios de usuarios y así establecer si este se ha escrito con una intención positiva o negativa para esto se usa la minería de datos, quiere decir, un análisis automático de texto escrito en lenguaje natural no está compuesto de listas con elementos ya sean positivos o negativos. (IONOS, 2020)

Una herramienta de análisis de sentimiento tiene que reconocer de forma específica el usuario y el entorno del producto en que se esté realizando, su tarea principal es tener una idea general de la opinión y en específico el estado de ánimo de un usuario en lo que respecta ya sea producto, marca o en este caso equipos de fútbol. Se lleva a cabo para definir emociones que hay detrás de un texto y analizar que quiso expresar realmente el usuario, permite medir el éxito de las campañas de marketing, comentarios relevantes sobre una marca. (IONOS, 2020)

El análisis de sentimientos está basado en el procesado de lenguaje natural, este pretende tener información subjetiva de un texto o documentos. (Garcia, 2019)

#### **2.3.3.1 Aplicación del Análisis de sentimiento.**

Realizar un análisis de sentimiento tiene beneficios múltiples, ayuda a empresas, gobiernos y organizaciones quienes en todo el mundo son los que tienen más interés en el avance en este campo de investigación. Conocer a través de este análisis que es lo que piensa la gente sobre productos, marcas. Es una de las herramientas muy valiosas que puede ofrecer muchas ventajas competitivas según el uso que se les dé. El uso de esta en las redes sociales es de gran ayuda para detectar la gestación de ciertos acontecimientos, ya sean quejas, opiniones positivos o negativos. (Sande J. S., 2018)

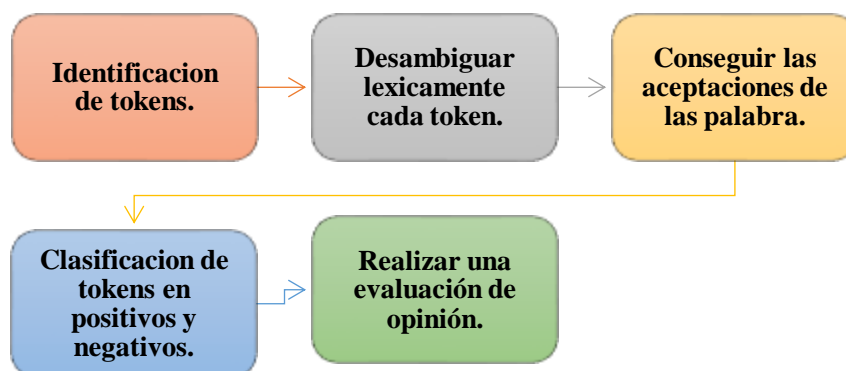
Puede ser aplicada en distintos ámbitos como son: Reputación política, corrección de opiniones, análisis del mercado financiero, posicionamiento de publicidad online, valoración de opinión de productos y servicios.

### 2.3.3.2 Polaridad

En el artículo “*Una herramienta para gestionar comentarios de la web*” menciona que:

Una de las principales tareas de extracción de opinión es la clasificación de la polaridad de opinión, que consiste en determinar si la opinión es positiva o negativa en relación con el tema al que se refiere. (Amores, Arco, & Artiles, 2015)

El análisis de sentimiento trabaja de la mano con la polaridad, esto consiste en clasificar textos, deducir si las opiniones de los usuarios son afirmativas o negativas. Esta puede ser medida a través de la subjetividad según su clasificación, de acuerdo con la fuente de donde proviene puede ser de un usuario, institución y que finalidad tiene su opinión.



**Ilustración 6.** Esquema para detección de la polaridad en comentarios. Información obtenida de la revista “*Una herramienta para gestionar comentarios de la web*”. Esquema realizado por Merly Avila.

### 2.3.4 Procesamiento del lenguaje natural.

El lenguaje es el conjunto de oraciones y secuencias de palabras a través de reglas gramaticales. Existe diferentes tipos de lenguaje según distintos criterios, entre ellos están el lenguaje natural, lenguaje formal y lenguaje artificial. En este caso nos enfocaremos en el lenguaje natural. (Gutiérrez-Herrera, 2015)

#### 2.3.4.1 Lenguaje natural

Es la lengua o idioma que nace de forma espontánea de un grupo de hablantes por necesidad de establecer una comunicación verbal. Se encuentra ligado a la cultura de cada

civilización y según como esta evoluciona dependiendo del uso que se le dé. Es de mucha importancia enmarcar el ámbito del empleo de la lengua natural en la comunicación. (Gutiérrez-Herrera, 2015)

#### ***2.3.4.2 Importancia del procesamiento del lenguaje natural.***

Es un campo enmarcado dentro del área de la inteligencia artificial y la lingüística. Tiene como objetivo fundamental facilitar la comunicación entre las personas y computadores haciendo uso de protocolos como los lenguajes naturales, estos son los más usados entre las personas para comunicarse entre sí, ya sea de forma oral o escrita, es más simple, directa y efectiva la comunicación entre elementos de la misma naturaleza, como personas, maquinas, que cuando es producida por entidades de diferentes orígenes.

Debido a la relación que hay entre las personas y sus computadoras es que se hace necesario el estudio y la búsqueda de protocolos que faciliten la interacción entre ambos para así poder mejorar las relaciones. Y el procesamiento del lenguaje natural es el encargado de facilitar esta tarea. (Sande J. S., 2018)

#### ***2.3.4.3 El PLN y sus niveles de análisis.***

Este tipo de sistemas deben llevar a cabo tareas que faciliten la comprensión entre la persona y el sistema. Estos trabajos consisten en un diseño de niveles en donde de forma secuencial se debe analizar e interpretar las oraciones hasta que sean entendidas y analizadas por el PLN. (Sande J. S., 2018) Existen 4 niveles de análisis, estos son:

- Nivel de análisis morfológico.
- Nivel de análisis semántico.
- Nivel de análisis sintáctico.
- Nivel de análisis pragmático.

Según el sistema a desarrollar se podrá implementar los diferentes niveles de análisis.

#### ***2.3.4.4 Aplicación del PNL.***

Tiene como objetivo mejorar la comunicación entre personas y computador, a continuación, se nombran algunas de las más importantes aplicaciones.

- Traducción automática de textos.
- Extracción de la información.

- Análisis de sentimientos.
- Recuperación de la información.

#### ***2.3.4.5 PNL aplicado al análisis de sentimientos.***

Según (Sande J. S., 2018), el sentimiento es una valoración cualitativa y cuantitativa a cerca de un beneficio, asistencia, individuo. Extraer de forma automática dicha información permite la creación de herramientas que ayuden a facilitar a conocer el sentir de las personas con relación a diferentes ámbitos de estudios.

#### ***2.3.4.6 Tokenización.***

Esta es una parte fundamental para la aplicación del lenguaje natural, un token es una unidad minima que una maquina puede entender y procesar. Este proceso se encarga de dividir una cadena sin procesar en pequeñas partes de un texto. (Perkins, Chopra, & Joshi, 2016)

### **2.3.5 Minería de datos**

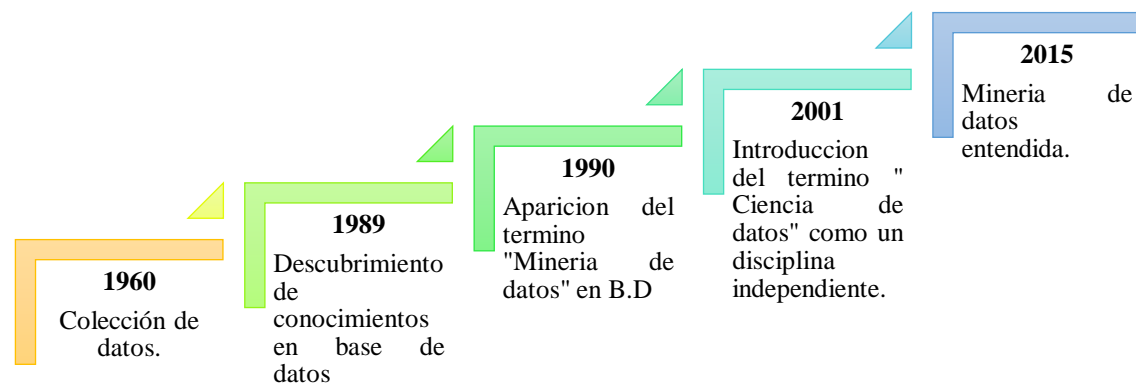
La minería de datos es un conjunto de métodos y tecnologías que permiten buscar en grandes bases de datos de forma automática o semiautomática para encontrar patrones, tendencias o reglas recurrentes que expliquen el comportamiento de los datos en un contexto determinado. Básicamente, la minería de datos intenta comprender el contenido del almacén de datos. Para ello utiliza métodos estadísticos y, en algunos casos, algoritmos de búsqueda cercanos a la inteligencia artificial y las redes neuronales. (Nectec, 2019)

#### ***2.3.5.1 Evolución de la minería de datos***

La minería de datos surge a inicios de los años 80 cuando en estados unidos se desarrolló un sistema de investigación para detectar fraudes a través sistemáticas de renacimiento de ciertos patrones. La expansión de este sistema se dio en los años noventa por las siguientes razones:

El aumento de los datos almacenados se ve favorecido por la automatización de trabajos, la observación con nuevas tecnologías y técnicas de almacenaje de datos como lo son las redes neuronales e inteligencia artificial. (Vallejo, Guevata, & Medina, 2018)

La evolución se distribuye de la siguiente manera, con las fechas más importantes.



*Ilustración 7. Evolución de la minería de datos, Información obtenida de la página TimeToast. Esquema realizado por Merly Avila.*

### 2.3.5.2 Aplicación de la minería de datos.

La minería de datos puede ser aplicada en distintos ámbitos que generen datos, entre varias de sus aplicaciones podemos enlistar las siguientes.

**Tabla 8.** Plazas de aplicación de la minería.

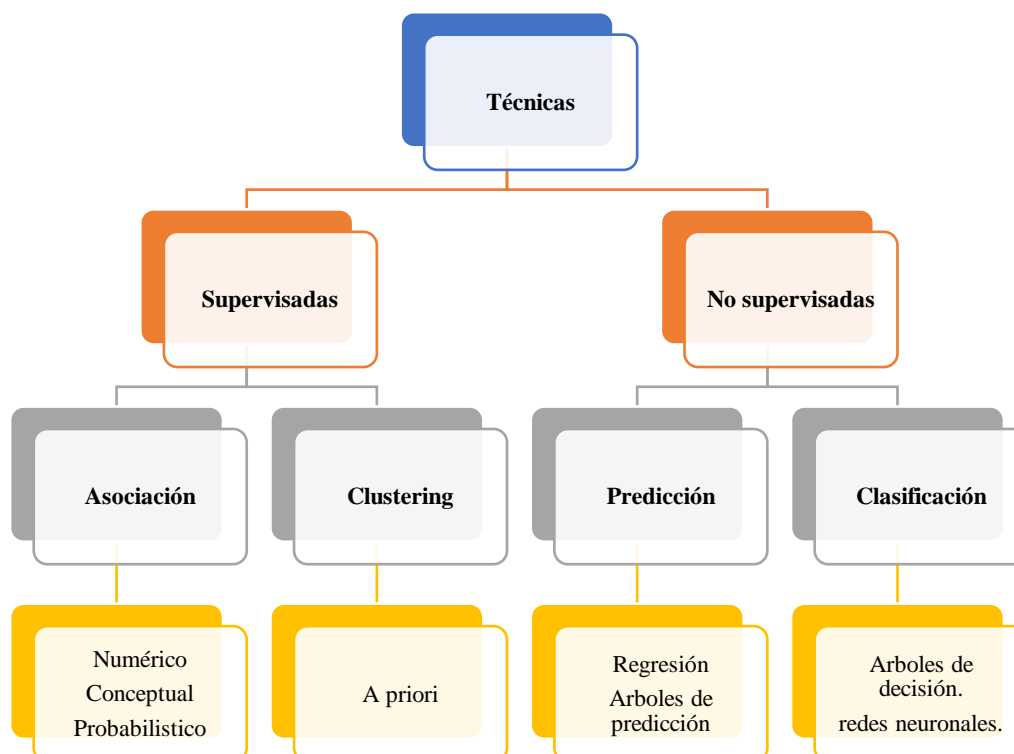
Áreas	Aplicación
Medicina	Buscar y encontrar alguna respuesta satisfactoria para algún tratamiento
Mercadotecnia	Identificar clientes dispuestos a responder ofertas de algún producto, servicio por correo, la fidelidad de este.
Banca	Analizar clientes, prestamos, determinar montos de crédito.
Industria y manufacturas	Diagnosticar fallas
Recuperación de información	Búsqueda e identificación de video, imagen, voz.
Geología	Identificar áreas de explotación minera en bases de datos de imágenes de satélites.

Agricultura	Identifica área de uso para cultivos.
Ciencias Ambientales	Identifica funcionamientos de ecosistemas ya sean naturales o artificiales para la mejora de su función.

*Información adaptada de la revista cubana de Ciencias informáticas, año 2009. Esquema realizado por Merly Avila.*

### 2.3.5.3 Técnicas de la minería de datos.

Se dice que la minería de datos es un conjunto de herramientas que ayudan a la obtención de un patrón de tendencia o comportamiento y puede ser utilizado en las tomas de decisiones. Pero esta no solo se basa en el uso de herramientas estándar, consiste más en una herramienta dinámica e iterativa y esta depende según al problema que se haya planteado, para esto es muy necesario conocer las técnicas que pueden usarse en esta búsqueda de patrones. Estas se dividen en:



**Ilustración 8.** Descripción de las técnicas de la minería de datos. Información obtenida mediante investigación. Esquema realizado por Merly Avila.

#### 2.3.5.4 Metodologías implementadas en la minería de datos.

(Cendejas, Acuña, Cortes, & Bolaños, 2017) Se debe tener en cuenta que el sistema de información es un conjunto de herramientas interconectadas que brindan acceso a las actividades de gestión y están especializadas en brindar información sobre cambios rápidos y oportunos; Los datos que solicita este programa son en su mayoría numéricos y cuantitativos, generalmente presentados gráficamente. Este tipo de sistemas brindan extracción de información a través de estándares, que son unidades o cadenas de información con estructuras de frecuencia conocidas. Algunos de los métodos aplicados se pueden enumerar dentro de estos estándares. A continuación, se enlistan las más importantes:

- Metodología KDD
- Metodología CRISP-DM
- Metodología SEMMA

Estas son las metodologías más implementadas en la minería de datos, porque proporcionan un orden en cada actividad realizada, teniendo como finalidad realizar trabajos de análisis para obtener resultados esperados y tomar decisiones adecuadas.

##### 2.3.5.4.1 Metodología KDD

La metodología KDD (Knowledge Discovery in Database) es un proceso que busca encontrar patrones, conocer la situación actual o futura de una organización por medio de modelos matemáticos. (Méndez, 2019).

El proceso KDD es un proceso para realizar automáticamente la extracción de conocimiento a partir de grandes cantidades de datos, que es de naturaleza iterativa y, por lo tanto, puede usarse varias veces hasta que se obtenga la información requerida.

##### 2.3.5.4.1.2 Fases de la metodología KDD.

Propone 5 fases que son: Selección, Limpieza, transformación, minería de datos e interpretación.

Las fases mencionadas son una parte del proceso KDD, pero FAYY96, incluye numerosos pasos y estos están estructurados en las siguientes etapas:

**Comprensión de dominio de la aplicación:** Se debe seleccionar el conocimiento relevante del dominio de la técnica y objetivos que el usuario quiera aplicar.

**Creación del conjunto de datos:** Almacenan los datos obtenidos y luego se procede a seleccionar los atributos.

**Limpieza y procesamiento de datos:** Se compone de la recolección de información más relevante, medir el tiempo y posibles cambios de la información.

**Reducción de los datos y proyección:** Depende del objeto del proceso, trata de encontrar las características más importantes de los datos.

**Escoger la minería de datos:** Se seleccionan la técnica en la cual se va a trabajar ya sea: Regresión, clasificación, agrupamiento.

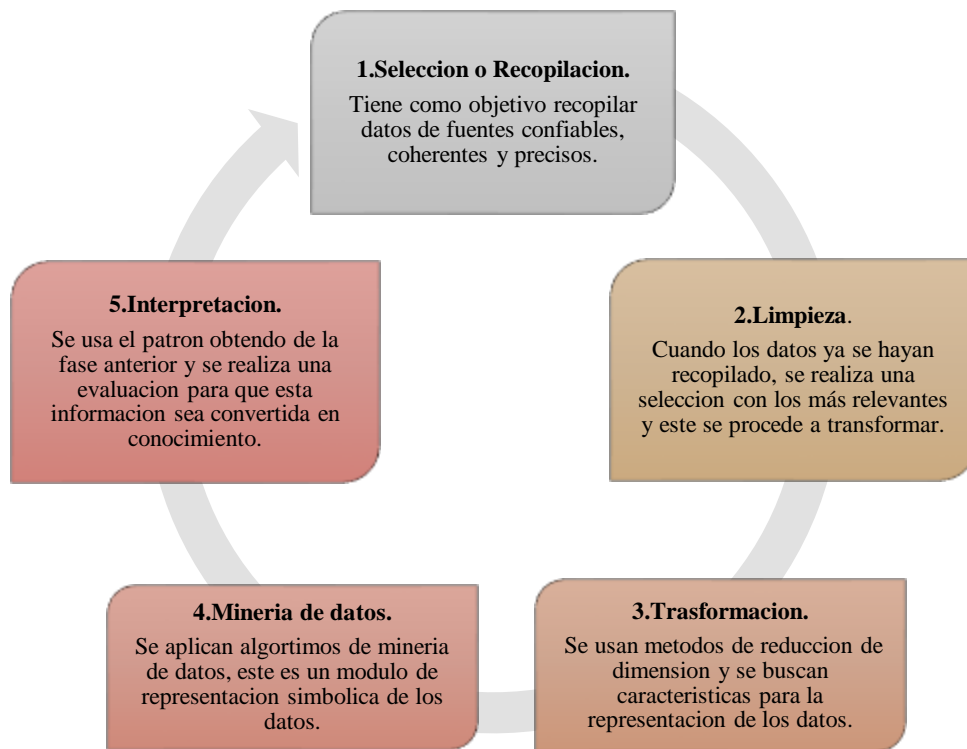
**Determinar algoritmos de la minería de datos:** Según la técnica escogida, se buscará un algoritmo ideal para la búsqueda del patrón de datos.

**Minar información:** Se basa en buscar modelos de utilidad según la metodología y algoritmo seleccionado.

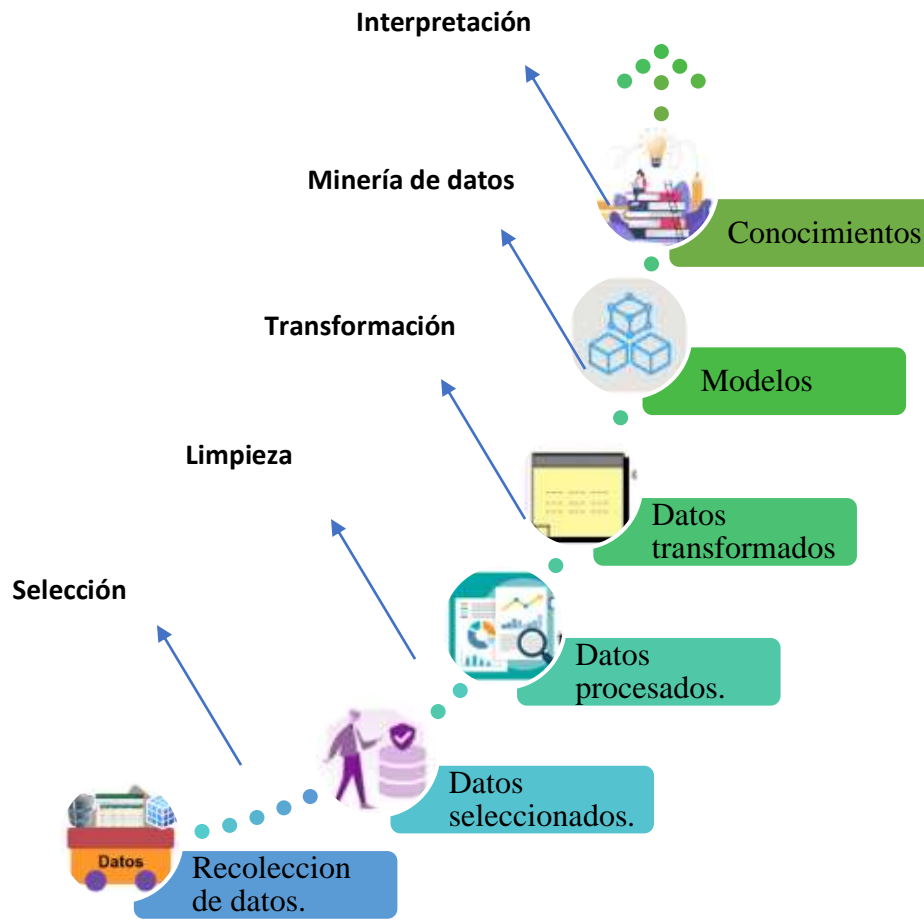
**Interpretación de patrones:** Muestra los resultados para entendimiento del usuario.

**Uso del conocimiento:** Se hace uso de la información que se a descubierto para la resolución de problemas.





***Ilustración 9.** Fases de la metodología KDD. Información obtenida mediante investigación. Esquema realizado por Merly Avila.*



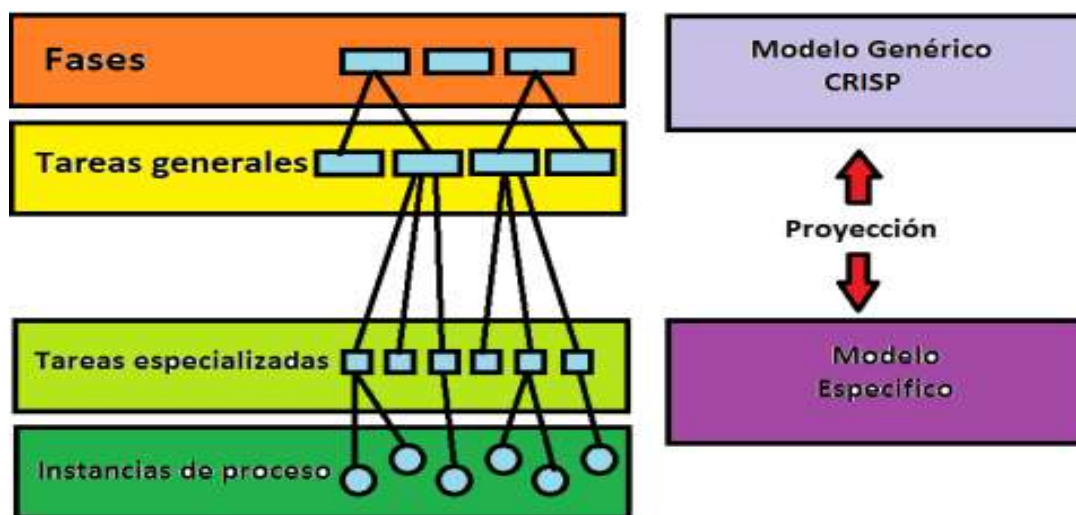
*Ilustración 10. Explicación de las fases de la metodología. Información obtenida mediante investigación.  
Esquema realizado por Merly Avila.*

#### 2.3.5.4.2 Metodología CRISP-DM.

CRISP-DM (Cross Industry Standard Process For Data Mining) Inicio en septiembre de 1996, permite que los proyectos trabajados sean replicados, muestra el ciclo de vida de un proyecto de una forma ordenada, abarca todas sus fases y relaciones entre las tareas asignadas. Esta metodología considera como un proyecto profesional a un proceso de análisis de datos implantando así un contexto más interesante en la elaboración de algún proyecto.

Tiene como ventajas:

- Es neutral con el uso de herramientas
- Tiene enfoque en el análisis técnico.
- Es una plataforma guía
- Contiene plantillas para realizar análisis.



*Ilustración 11. CRISP-DM estructura de los 4 niveles. Información obtenida de repositorio. Esquema realizado por autor.*

Esta metodología se divide en cuatro niveles de abstracción con una organización de representación escalonada mencionando asignaciones desde el nivel más global a las más específicas. (Arancibia, 2009)

La composición de su primer nivel consta de 6 fases:

Comprensión del negocio.	Comprensión de los datos.	Preparación de datos.	Modelado	Evaluación	Implementación
<ul style="list-style-type: none"> <li>• Determinar los objetivos del negocio.</li> <li>• Evaluación de la situación para cumplir con los objetivos.</li> <li>• Determinar el objetivo de la extracción de información.</li> <li>• Efectuar la guía de trabajo.</li> </ul>	<ul style="list-style-type: none"> <li>• Se recolectan los información inicial.</li> <li>• Descripción de la información recolectados.</li> <li>• Observación de la información.</li> <li>• Comprobación de la calidad de la información.</li> </ul>	<ul style="list-style-type: none"> <li>• Selección de información.</li> <li>• Limpiar la información.</li> <li>• Construir datos, atributos, registros y tablas.</li> <li>• Integrar la información.</li> <li>• Realización de formateo de información.</li> </ul>	<ul style="list-style-type: none"> <li>• Seleccionar la técnica de modelado.</li> <li>• Construir el diseño de prueba.</li> <li>• Construir el modelo.</li> <li>• Realizar una evaluación.</li> </ul>	<ul style="list-style-type: none"> <li>• Resultados</li> <li>• Verificación del proceso</li> <li>• Establecer los pasos siguientes.</li> </ul>	<ul style="list-style-type: none"> <li>• Planear la implementación.</li> <li>• Ejecutar un plan de seguimiento.</li> <li>• Construir informe final.</li> <li>• Revisar el proyecto.</li> </ul>

*Ilustración 12. Fases de la metodología CRISP. Información obtenida de repositorio. Esquema realizado por autor.*

**Comprensión del negocio:** Esta fase embarca el problema que tiene el negocio, permite analizar sus objetivos y requerimientos que este tiene. La importancia es comprender los objetivos ya que sin este ningún algoritmo podrá resolver el trabajo y así no obtener los resultados esperados.

**Comprender los datos:** Recopila la información referente al problema, teniendo como objetivo de sacar las primeras conclusiones y así crear conocimiento.

**Preparar los datos:** En esta fase se debe seleccionar la información más importante, a partir de los datos que se recolectaron en las fases iniciales.

**Modelado:** Realiza una elección de técnicas más adecuadas, para iniciar el proceso al conjunto de datos que se determinaron por medio de herramientas de minería de datos, y construir el modelo.

**Evaluación:** Se realiza la evaluación y comprobación de la funcionalidad del modelo creado, verificando los resultados que se muestren.

**Implementación:** Esta fase implementa ya el conocimiento que se obtuvo, realizando informes con los últimos resultados finales y que estos sean entendibles para el usuario.

#### 2.3.5.4.3 Metodología SEMMA.

Es una creación de SAS Institute, con el fin de trabajar con software SAS Enterprise Mine. Es conocida como una metodología, pero su ideal es ser una proposición de clasificación lógica de tareas de gran importancia en el SEMMA, tiene un conjunto de 5 fases para realizar un proceso de minería; Sample (Muestreo), explore (Exploración), modify (Modificación), model (Modelado), Assess (Evaluación). (Moine, 2013) (Ver ilustración 13)

Se orienta en el desarrollo del modelo de minería, dejando a un lado algunos aspectos de la investigación tales como el discernimiento del problema que se está estudiando o la planificación de la implementación de este. Según (Moine, 2013) Describe las fases de la siguiente manera:

**Muestreo:** Selecciona muestras de los datos que estén disponibles, para obtener información de relevancia deben tener un tamaño muy grande, y a su vez lo suficientemente pequeño para que el proceso ejecute de la manera correcta.

**Exploración:** Explora la información en busca de tendencias y relaciones inexploradas.

**Modificación:** En esta fase se preparan los datos, haciendo limpiando los valores que tengan anomalías. Se selecciona, crea, modifica las variables a trabajar

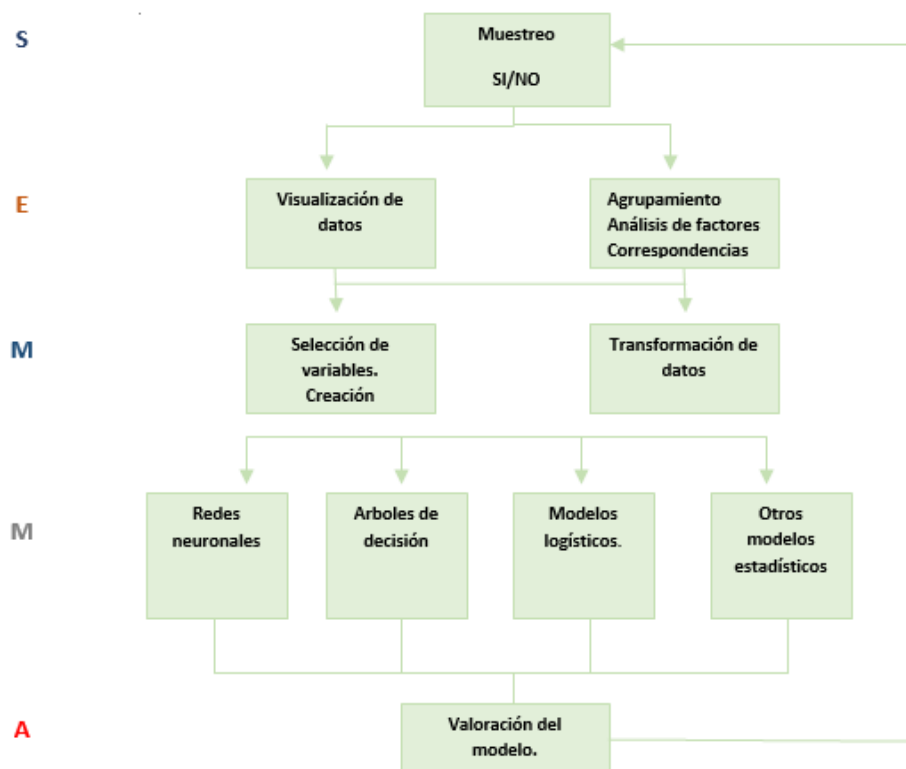
**Modelado:** En esta fase se crea el modelo que consentirá predecir las variables que se den de respuestas a variables explicativas. Hace uso de técnicas predictivas como redes neuronales, análisis de regresión y arboles de decisión.

Es un proceso de alto nivel, que no se diferencia con las otras metodologías, tal cual a la metodología KDD, SEMMA no facilita un plan de tareas que se deban realizar en cada una de las etapas. Por tal motivo hay una controversia acerca si SEMMA puede ser llamada una metodología.



*Ilustración 13. Fases metodología SEMMA. Información obtenida de repositorio. Esquema realizado por Merly Avila*

En la ilustración 14—se muestra un esquema de la iteración general del proceso SEMMA.



*Ilustración 14. Iteración del proceso SEMMA. Obtenido de repositorio. Esquema realizado por Merly Avila.*

### 2.3.5.5 Análisis comparativo de las metodologías KDD, CRISP-DM, SEMMA.







**Tabla 9.** Comparativa de las metodologías.

Metodologías	KDD	CRISP-DM	SEMMA
<b>Descripción</b>	Ejecuta modelos matemáticos. Busca patrones actuales o futuros.	Mayor control. Se encuentra en constante evaluación.	Se enfoca en la parte técnica. No analiza la problemática. Esta idealizado en medir la población.
<b>Fases</b>	Comprensión de dominio de la aplicación Creación del conjunto de datos. Limpieza y procesamiento de datos. Reducción de los datos y proyección Escoger la minería de datos Determinar algoritmos de la minería de datos. Minería de información. Interpretación de modelos. Uso del discernimiento.	Perspicacia del negocio. ➤ Perspicacia de la información. Preparación de la información. Modelado. Evaluación.	Muestreo Indagación Modificación Modelado. Evaluación.
<b>Herramientas</b>	Weka. Kmine Rapidminer	Weka Kmine Rapidminer Clementine	Weka. Kmine. Rapidminer. IBM SPSS.
<b>Diseñada para minería de datos</b>	SI	SI	SI
<b>Algoritmos</b>	Clústers. Series Arboles de decisión Regresión Arboles de decisión.	Regresión Redes neuronales Arboles de decisión.	Clouster. Arboles de decisión. Series de tiempo.

### 2.3.5.6 Herramientas para la minería de datos.

Hacer uso de datos previamente analizados por herramientas que tengan un enfoque estadístico para informar en problemas en el que estos datos puedan ofrecer una solución. Existe un sin número de herramientas para la aplicación de minería de datos entre ella se realiza una breve descripción de algunas de las más utilizadas.

**Tabla 10.** Herramientas para la minería de datos.

Herramienta	Logo	Tipo	Costo	Extensiones Librerías.
KNIME		Orientado a no programadores.	Puede ser gratuita/ El pago varía según su versión.	MySQL, H2O, Plotly.
R		Orientado a programadores.	Gratuita.	Plotly, H2O, dplyr.
ORANGE		Orientado a no programadores.	Gratuita.	Python, PostgreSQL.
PYTHON		Orientado a no programadores.	Gratuita.	Numpy, Scipy, Matplotlib.
WEKA		Orientado a no programadores.	Gratuita.	R, Deeplearnin, Scikit-learn.
GNU OCTAVE		Orientado a no programadores.	Gratuita	Nurbs, geometry.

**RAPID  
MINER**



Orientado a  
programadores. Gratuita.

Java,  
NoSQL, R, weka

---

*Información Obtenido mediante investigación. Esquema realizado por Merly Avila.*

## 2.4 Definiciones Conceptuales.

### 2.4.1 Base de datos

(Peiró, 2020 ) Menciona que; Una base de datos se refiere a un conjunto específico de datos o información que puede ser consultado de manera inteligente y comparte las características que desea resaltar para identificar la información que desea examinar.



*Ilustración 15. Base de datos. Información obtenida de la página web robótica holguinista. Elaborado por el autor.*

### 2.4.2 Knime.

Una plataforma de código abierto que ofrece a sus usuarios la visualización de flujos de datos y la ejecución selectiva de pasos analíticos, útiles para la integración, el procesamiento, el análisis y la investigación.



*Ilustración 16. Logo Knime. Información obtenida de knime.com. Elaborada por el autor.*



### 2.4.3 R

Es un entorno de software libre. Está inspirado por el lenguaje S su inicio fue en 1976 como ambiente de análisis, en 1988 se reescribió en C. Considerado un sistema planificado y coherente, es utilizado comúnmente para la computación y estadística.



*Ilustración 17. Logo de R. Información obtenida mediante investigación. Elaborado por el autor.*

### 2.4.4 Orange

Contiene componentes para preprocesamiento de datos. Cuenta con un versátil y rápida interfaz de programación.



*Ilustración 18. Logo de Orange. Información obtenida mediante investigación. Elaborado por el autor*

### 2.4.5 Weka

Utilizado para el aprendizaje automático y minería de datos, esta implementado con Java y compila en cualquier plataforma. Es fácil su uso gracias a la interfaz gráfica que posee.



*Ilustración 19. Logo Weka. Información obtenida mediante investigación. Elaborado por el autor*

### 2.4.6 Rapidminer

Basada en la minería de datos, implementa 500 técnicas o más de procesamiento de datos, visualización de datos y métodos de pruebas de modelos.



*Ilustración 20. Logo Rapidminer. Información obtenida mediante investigación. Elaborado por el autor*

### 2.4.7 GNU octave

Lenguaje de programación de alto nivel escrito en C++ es utilizado para la realización de cálculos numéricos.



*Ilustración 21. Logo Octave. Información obtenida mediante investigación. Elaborado por el autor*

### 2.4.2 Python

Para (Robledano, 2019) Es un lenguaje de programación multiplataforma y multiparadigma que destaca por su código legible y limpio. Una de las razones de su éxito es que cuenta con una licencia de código abierto que permite su uso en cualquier escenario. Es ideal para trabajar con grandes cantidades de datos, ya que admite la minería y el procesamiento como multiplataforma, razón por la cual las empresas de big data lo eligen.



*Ilustración 22. Logo Python. Información obtenida de la página de Python.org. Elaborado por el autor.*

### 2.4.3 Jupyter Notebook

Según (Cabrera & Díaz, 2018) Jupyter Notebook es una interfaz web de código abierto que permite importar texto, video, audio, imágenes y ejecución de código desde un navegador en varios idiomas. Esta interfaz independiente del lenguaje (de ahí su nombre, porque combina 3 lenguajes de programación de código abierto en el campo de la ciencia: Julia, Python y R) puede, por lo tanto, estandarizar la presentación del contenido de los cursos científicos, sin asumir ningún objeto en un idioma es limitado.



*Ilustración 23. Logo Jupyter Información obtenida de la página de Jupyter.org. Elaborado por el autor.*

### 2.4.4 Anaconda

Tiene como ventaja ayudar a simplificar la gestión e implementación de paquetes. La distribución incluye paquetes de “data science” adecuados para los sistemas operativos Windows, Linux y macOS. (Rondón, 2022)



*Ilustración 24. Logo Anaconda. Información obtenida de la página web anaconda.com. Elaborado por el autor.*

### 2.4.5 Developer Twitter

Developer Twitter consiente en aprovechar la plataforma abierta, en tiempo real e histórica de Twitter dentro de sus propias aplicaciones. Proporcionar herramientas, recursos, datos y productos API que lo ayuden a integrar y expandir su influencia en Twitter a través de investigaciones, soluciones y más. (Twitter, 2022)

La Plataforma para desarrolladores de Twitter le permite aprovechar el poder de la plataforma abierta, global, en tiempo real e histórica de Twitter dentro de sus propias aplicaciones.



*Ilustración 25. Interfaz API. Información obtenida de la página web developer twitter. Elaborado por el autor.*

#### 2.4.6 Api Twitter

La API de Twitter es un conjunto de puntos de programación que se pueden utilizar para comprender o crear conversaciones en Twitter. Esta API le permite buscar, buscar, interactuar o crear varios recursos, incluidos: Tweets, Usuarios, Mensajes directos, Tendencias, Lugares. (Twitter, 2022)



*Ilustración 26. Logo developer. Información obtenida de la página web spiralking. Elaborado por el autor.*

#### 2.4.7 Token

Es una unidad de valor emitida por entidades privadas. (Blockchain, 2021) define que; Son activos digitales únicos que no son intercambiables porque no hay dos iguales. Cualquier cosa que pueda representarse digitalmente puede convertirse en una NFT: desde tweets hasta memes y obras de arte. La criptografía de "token" permite que un solo propietario sea dueño de la pieza original.



*Ilustración 27. Información obtenida de la página web alamy. Elaborado por el autor.*

## 2.5 Fundamentación legal

La (Ley Organica de Proteccion de datos Personales, 2021) hace mención en el 1er Cap Ámbito de Aplicación Integral.

**Artículo 1.-** Objetivo y Finalidad. – “El objeto y finalidad de la presente ley es garantizar el ejercicio del derecho a la protección de datos personales, que incluye el acceso y decisión sobre información y datos de este carácter, así como su correspondiente protección. Para dicho efecto regula, prevé y desarrolla principios, derechos, obligaciones y mecanismos de tutela”.

**Artículo 2.-** Ámbito de aplicación material. – “La presente ley se aplicará al tratamiento de datos personales contenidos en cualquier tipo de soporte, automatizado o no, así como a toda modalidad de uso posterior”.

Según el (Decreto Ejecutivo N-1014, 2008) Acerca del uso de patrones abierto y software libre, como herramientas informáticas.

**Artículo 1.-** Establecer política pública para que las unidades del gobierno central utilicen software libre en sus sistemas y dispositivos informáticos.

**Artículo 3.-** Se comprende por software libre, a programas de computación que se pueden utilizar y distribuir sin restricción alguna, que permitan su acceso a los códigos fuentes y que sus aplicaciones puedan ser mejoradas.

Estos programas de computación tienen las siguientes libertades:

- Utilizar el programa con cualquier propósito.
- Distribuir copias sin ninguna restricción.
- Publicación del programa mejorado.

## **Capítulo III**

### **Desarrollo de la propuesta.**

#### **3.1 Metodología del proyecto.**

En este capítulo tiene como finalidad desarrollar la propuesta, aplicando cada una de las metodologías que sirvan de ayuda para la resolución del problema a estudiar. Como lo es en este caso analizar el sentimiento en Twitter.

##### **3.1.1 Diseño de la investigación.**

En este trabajo se plantea usar una investigación cuantitativa, principalmente siendo causi-experimental y longitudinal. Teniendo como propósito descubrir los sentimientos de usuarios a través de tweets publicados acerca del fútbol ecuatoriano, haciendo enfoque en los 4 primeros equipos con mayor hinchada de la serie A del fútbol en el Ecuador.

A su vez este trabajo, una investigación bibliográfica ya que para cumplir con los objetivos propuestos se perpetró una investigación de varias fuentes con la finalidad de indagar estudios, metodologías, técnicas, procedimientos necesarios y relacionados para analizar el sentimiento en Twitter.

##### ***3.1.1.1 Enfoque cuantitativo.***

Mediante la recolección de datos, este enfoque mide la realidad que se investiga, no solo busca medir sino también determinarlo, examinando la realidad objetiva, las técnicas de recolección de datos son estandarizadas. Hace uso a las estadísticas mediante esto se puede mostrar los resultados mediante graficas.

(Álvarez, 2011) menciona que el propósito de este enfoque es analizar y buscar una definición a los fenómenos fundando órdenes, esto es, encontrar estatutos que expongan la conducta social. Teniendo como finalidad valerse exclusivamente de la investigación directa de la demostración y la práctica. Este discernimiento debe ser basado en un análisis de hechos existentes, por lo cual se debe efectuar una representación neutral, completa y objetiva.

##### ***3.1.1.2 Diseño cuasi-experimental.***

El objetivo de este diseño es poner a prueba una suposición causal haciendo uso de una variable donde por razones logísticas o éticas no se puede determinar las unidades de investigación aleatoriamente a los grupos. (Fernandez, Vallejo, Livacic, & Tuerco, 2014)

Permite considerar situaciones y fenómenos tal y como se muestren, aquellos que se manifiestan ya sea de forma natural o directa. Este diseño se define como el que casi alcanza el nivel práctico, que limita para que llegue a ese nivel es que no hay una manera que asegure un nivel inicial de los grupos a experimentar.

### ***3.1.1.3 Enfoque Longitudinal.***

Según (Loeber & Farrington, 2015) Este tipo de enfoque, son aquellos que realizan una recolección de datos sobre un grupo de sujetos, siempre la muestra, en distintos momentos a lo largo del tiempo.

Este tipo de estudio tiene como objetivo estudiar y evaluar a un grupo de personas por un periodo de tiempo prolongado, está definido como un análisis de cambios que se producen en el tiempo en una misma muestra.

### ***3.1.1.4 Diseño Documental.***

Es aquella investigación que depende de la información seleccionada mediante, documentos, revistas, meta-análisis.

Para (Alfonso, 1995) Una investigación documental es una técnica científica, un desarrollo ordenado de búsqueda, recolectar, organizar, analizar e interpretar la información a un tema determinado.

## **3.2 Tipos de investigación.**

Para el desarrollo de la propuesta, se hace uso de varias metodologías y técnicas, por lo cual, se debe definir el tipo de investigación que se aplicará durante el proceso.

### **3.2.1 Investigación exploratoria.**

La investigación exploratoria es de gran utilidad, para adquirir información sobre la posibilidad de efectuar una exploración integra acerca de un argumento particular. Investiga inconvenientes de la conducta humana, identificar conceptos o variables promisorias, establecer prioridades para investigaciones posteriores o sugerir afirmaciones (postulados) verificables. (Sampieri, Collado, & LucioP, 1997)

Se hace uso de esta investigación en el presente trabajo, es importante conocer la información relacionada con Twitter y para ser precisos en información que muestre la aplicación de un análisis de sentimientos a usuarios en redes sociales, medir las opiniones que se generen, investigar algoritmos que se puedan utilizar en el desarrollo para interpretar emociones y

todos los puntos que estén relacionadas a la propuesta. Es de gran importancia argumentar las definiciones conceptuales investigadas en diferentes fuentes.

### **3.2.2 Investigación descriptiva.**

Este estudio describe la población, situación o fenómeno en el que se desarrolla el estudio. Proporcionar información sobre qué, cómo, cuándo y dónde es relevante para la pregunta de investigación. Como su nombre indica, este tipo de investigación no proporciona una explicación, sino una descripción de la situación. (Jervis, 2020)

### **3.2.3 Investigación cuasiexperimental.**

Esta investigación nos ayudará a obtener los datos relacionados a las emocionales que se muestren en usuarios en la red social Twitter.

## **3.3 Técnicas de investigación.**

Son un conjunto de metodologías su objetivo es asegurar la funcionalidad de un proceso investigativo. Es decir, permite obtener información y conocimiento para encontrar soluciones a nuestras dudas. (AliatUniversidades, 2020)

### **3.3.1 Técnica de muestreo.**

Según (Díaz, 2006) Estas técnicas de muestreo están basadas en un conjunto de estadísticas que analizan la forma de elegir una muestra característica de la población. Quiere decir que este personifique a la población a la que se pretende extrapolar.

Esta técnica, permite seleccionar información, datos, comentarios, referentes a un grupo ya sea grande o pequeño de la población con el propósito de conseguir resultados a lo que requiera en la investigación.

En esta investigación se seleccionarán datos de la red social Twitter de las cuentas oficiales @BarcelonaSC, @CSEmelec, @LDU\_Oficial, @IDV\_EC con la finalidad de analizar los datos y definir comentarios, opiniones positivas o negativas referente al fútbol ecuatoriano.

#### **3.3.1.1 Muestreo no probabilístico**

Este tipo de técnica permite la seleccionar un grupo de la población de forma aleatoria, este muestreo muestra varios tipos, como lo son el muestreo por conveniencia, muestreo consecutivo o muestreo por cuotas.



### **3.3.1.2 Muestreo por conveniencia.**

Esta técnica selecciona una muestra de la población que está disponible para el investigador, y esta puede ser aplicada cuando este no consideró realizar una selección de muestra que represente a toda la población en general. En la investigación a realizarse se debe realizar una muestra que represente en específico una muestra de la población, cuando en la investigación esta sea demasiado grande para ser evaluada.

En este proyecto se aplicará este muestreo porque se seleccionará una muestra de la usuarios en específico para poder realizar el análisis de sentimientos en los tweets publicados referente a los equipos seleccionados.

### **3.4 Metodología de desarrollo.**

La metodología que se va a aplicar en este trabajo se basa en el ciclo de analítica de datos, en varias investigaciones esta metodología trabaja con un gran volumen de datos que no estén estructurado. Para que este se cumpla se debe aplicar la secuencia de pasos, basándose en la organización de tareas relacionadas a la adquisición y análisis de datos. Pero no se pretende cumplir todos los pasos, ya que solo se seleccionará las fases que se puedan adaptar al proceso de análisis que se quiere realizar con técnica basada en el data mining.

Esta metodología se compone de nueve fases, pero para la realización de este análisis que contiene un gran volumen y variedad de datos que se procesen, pero en este proyecto solo se utilizaran seis, según los investigado en el capítulo anterior referente a los métodos implementados en el data mining, escogiendo aquello que está basado en el ciclo de análisis de datos, ajustando las fases a lo que requiere el proyecto.

Implementando seis fases de analítica que tiene enfoque a la minería de datos se va a realizar una identificación en la red social Twitter, donde se hará uso de una cantidad de tweets que sean publicados por usuario en las cuentas oficiales de los equipos de fútbol Barcelona, Emelec, Idv, Liga de Quito se realizará una extracción de tweets realizados por usuarios, haciendo uso de la API de Twitter mediante la Jupyter Notebook, basado en el lenguaje de programación Python, los datos que se generen se guardaran en un archivo CSV en un libro de Excel. Luego de la extracción de dichos datos se procede a realizar una limpieza y validación en la misma notebook. Las sies fases ayuda a preparación de datos para realizar el análisis de sentimientos.

A continuación, se enlistan las fases a utilizar:



*Ilustración 28. Fases Metodología aplicada. Información obtenida mediante investigación. Elaborado por Merly Avila.*

Los datos que se adquieran y pasen por el proceso mencionado serán utilizados para el análisis a realizarse, se hará uso de la misma notebook en Jupyter, en esta etapa se procede a realizar el análisis de sentimientos haciendo uso de códigos que permitan realizar el desarrollo de los algoritmos de forma automática. Todos los datos que se obtengan se guardaran en un archivo CSV, para así comprender la lectura al igual que su almacenamiento. Para la visualización de la información que se requiere, se realizará mediante gráficos que muestren lo que se requiere de manera comprensible para el lector, por ende, también se generará una nube de palabras y gráficos que muestren la polaridad en cada tweet indicando cuales son los comentarios positivos y negativos.

### **3.4.1 Descripción del proyecto.**

Esta propuesta consiste en analizar el sentimiento en Twitter haciendo uso de minería de datos. Teniendo como objetivo identificar y extraer cierta cantidad de tweets mediante la Api de Twitter y a estos datos realizarle un proceso de limpieza y validación, los resultados obtenidos serán de ayuda para realizar el análisis de sentimientos en el ámbito del deporte,

siendo de gran ayuda ya sea para medir la interacción de los hinchas por el medio cibernético, realizar este tipo de análisis permite saber cómo se sienten o que tanto dicen las personas sobre el fútbol

### 3.4.2 Factibilidad Técnica.

Para la extracción y recolección de datos se trabaja con la Api de la red social seleccionada, teniendo acceso mediante contraseñas (Se debe crear una cuenta en el developer de la aplicación. Vea anexo 1). Dichos datos se procederán a almacenar para la realización de la limpieza y filtrado, para esto se utiliza la plataforma web Jupyter Notebook del lenguaje de programación libre y gratuito Python que cuenta con una gran cantidad de librerías y es de fácil adiestramiento siendo esta una de las razones por la cual es más factible realizar el análisis en esta plataforma. Al finalizar los resultados de polaridad de cada equipo, se presentarán mediante gráficos de barra y diagramas pastel.

En la siguiente tabla se realizará una breve descripción de los software y hardware que fueron utilizado para la elaboración de este análisis de sentimientos.

**Tabla 11.** Hardware Y software utilizados.

Hardware	Características
Dell Inspiron 14-3000	Windows 10 Proc Intel® Core™i5- 7200U CPU @ 2.50GHz 2.70 GHz Ram 8,00 GB Almacenamiento 931GB
Software	Características
Anaconda Navigator	Anaconda 3 (64 Bit)
Jupyter Notebook	6.4.5
Api Twitter	-
Formatos Csv	-
Excel	-

*Información tomada de la presente investigación. Elaborada por Merly Avila.*

### 3.4.3 Factibilidad Operacional.

En este proyecto al analizar el sentimiento no pretende infringir ninguna ley para su desarrollo como la extracción de los tweets ya que esta información es pública. Respetando así al artículo de la Ley de Protección de datos personales. A su vez la mención en el capítulo

anterior y uso de hardware y software libre como se especifica en el decreto Ejecutivo 1014 2008 Art 1 y 2, respaldados en la Constitución Nacional de la república del Ecuador.

### 3.5 Esquema general del análisis de sentimientos de datos.

Este esquema detalla el proceso como fue el desarrollo del análisis de sentimientos. Teniendo como objetivo analizar la opinión de usuarios en la red social y mostrar su polaridad (Ver ilustración 29) Se puede visualizar el cumplimiento de las fases de la metodología aplicada, clasificada en diferentes etapas como el procesamiento, almacenamiento, análisis de sentimientos aplicando aquí la extracción, visualización y resultados de la información mediante gráficos.



*Ilustración 29. Esquema general del análisis de sentimientos. Información por datos de la investigación.  
Elaborado por Merly Avila*

#### 3.5.1 Recursos utilizados para la realización del análisis de sentimientos.

Para este proceso fue importante conocer el entorno en el que se iba a desarrollar, así como la implementación de herramientas, lenguaje de programación ya mencionados en el capítulo anterior. Se escogieron las más factibles que permitieron desarrollar un trabajo eficiente y lograr cumplir con uno de los objetivos planteados.

### **3.5.1.1 Anaconda**

Software principal utilizado para la realización del proyecto, contiene Jupyter y el lenguaje de programación Python.

### **3.5.1.2 Twitter Api**

Herramienta de mucha importancia, permite el acceso a los tweets publicados por usuarios, pero se realiza mediante llaves de acceso que son generadas una vez se haya creado una cuenta en el Developer Twitter. (Ver anexo 3)

### **3.5.1.3 Tweets.**

Textos que contienen información acerca de un argumento en específico, como en este trabajo el Fútbol. Para este análisis se realiza una extracción de tweets de 4 equipos de la serie A del fútbol ecuatoriano. Los tweets tienen 4 tipos; Con enlaces, sin enlaces, especiales y automáticos. En este caso se utilizan los tweets especiales y este se clasifica en Tweets de menciones, de respuesta y Retweet.

### **3.5.1.4 Entorno Jupyter Notebook.**

Software libre con entorno desarrollado para trabajar en línea, será ejecutado en el navegador Avast Secure Browser, herramienta de gran utilidad para la realización de este análisis de sentimientos se escribirán las líneas de códigos y se ejecutarán bloque por bloque para mejor desempeño y así obtener respuestas de una manera más rápida. Este permitirá la limpieza, validación, y aplicación de la minería de datos y a su vez permite la visualización de los resultados esperados.

### **3.5.1.5 Librerías a utilizar en Python.**

Python es un lenguaje de programación estructural con una variedad de librerías que permiten obtener respuestas de una forma rápida alguna de ellas ayuda a la identificación y uso de modelos del lenguaje de procesamiento natural, es lo que se busca realizar en los datos obtenidos, a su vez contiene distintas librerías según las necesidades que se presenten durante la elaboración del proyecto. El uso de este lenguaje en el proyecto es debido su sencillo y fácil aprendizaje.

En la siguiente tabla describe las librerías que se emplearan para el desarrollo del análisis de sentimientos.

**Tabla 12.** Librerías utilizadas en el trabajo de investigación.

<b>Librerías</b>	<b>Utilidad</b>
<b>Tweepy</b>	Es de gran importancia, tiene la función como “wrapper” permitiendo la interacción con la Api rest de Twitter, por medio de esta librería se tendrá acceso a los token y credenciales en el Developer para realizar la extracción de Tweets.
<b>Path</b>	Sirve para obtener una ruta absoluta de algún modulo en Python.
<b>os</b>	Permite realizar operaciones que estén sujetas del sistema operativo, ya se crear carpetas, listas de contenidos.
<b>csv</b>	Valores separados por comas, formato utilizado para la exportación e importación de bases de datos y hojas de cálculo.
<b>TextBlob</b>	Esta librería de procesamiento de texto en Python es de suma importancia en este análisis ya que permite realizar tareas del PNL.

<b>WordCloud</b>	Útil para la reducción de contenidos importantes de un texto. Mientras mas frecuencia tenga una palabra, más grande se muestra en la nube de palabras.
<b>pandas</b>	Permite manipular tablas numéricas o series temporales es una librería con extensión a numpy.
<b>numpy</b>	Esta librería permite la creación de vectores y matrices y manejo de datos de una forma eficaz.
<b>re</b>	Especifica conjunto de cadenas, permite verificar si una cadena particular es igual a una expresión regular.
<b>matplotlib.pyplot</b>	Sirve para crear visualizaciones interactivas, animadas o estáticas.
<b>nltk</b>	Librería muy útil para el trabajo en lingüística computacional en Python, facilita el trabajo con datos de lenguaje humano.
<b>SentimentIntensityAnalyzer</b>	Determina de forma computacional la polaridad de un texto. (Positivo, negativo, neutro).
<b>tqdm</b>	Hace que los bucles muestren de forma instantánea un medidor de

progreso.

<b>stopwords</b>	Ayuda a cumplir la función de eliminar palabras vacías, aquellas que no muestran significado al texto.
<b>word_tokenize</b>	Divide cadenas en una lista de tokens.
<b>unidecode</b>	Enumera caracteres escritos en lenguaje humano y le otorga su propio código único a cada carácter.
<b>deep_translator</b>	Sirve para la traducción de los datos ya que en este tipo de análisis se usan librerías reciben la información en inglés para mejor una predicción.

---

*Información tomada de la presente investigación. Elaborada por Merly Avila.*

### **3.6 Aplicación de la metodología.**

Este proceso de análisis está adaptado a la metodología antes mencionada que consta de seis fases, una de ellas cumple con la aplicación del análisis de datos relacionada a la minería de datos con la finalidad de cumplir con los objetivos planteados.

Se procede a explicar la aplicación de cada fase.

#### **3.6.1 Identificar datos.**

Esta fase principal nos permite reconocer los datos que se requieren analizar en este caso, se busca analizar la polaridad en los tweets publicados en la red social según su perspectiva en el ámbito del deporte, y del equipo el cual es hinchas.

Siendo Twitter la red social de opinión pública, se caracteriza por ser una de las usadas por las personas y porque las cuentas en su mayoría son de acceso público, se realizará una extracción de gran cantidad de datos generados por usuarios de las cuentas de; BarcelonaSC, CSEmelec, LDU, IDV y proceder a analizarlos.





*Ilustración 30. Identificación de datos. Información tomada de la presente investigación. Elaborado por Merly Avila.*

### 3.6.2 Adquirir y Filtrar datos.

Para adquirir los datos se debe conectar a los servidores de Twitter y eso es mediante la API, esto se realiza en Jupyter Notebook haciendo uso de las librerías necesarias que ayuden el ingreso a los servidores para tener obtener las contraseñas del developer que permitan la extracción de datos, las credenciales son: clave del consumidor, secreto del consumidor, token de acceso, token de acceso secreto.

```

#Reading Credentials
#Change Credentials for you user
credentials = pd.read_csv('./Credentials/CredentialsStudent.csv')
consumerKey = credentials['key'][0]
consumerSecret = credentials['key'][1]
accessToken = credentials['key'][2]
accessTokenSecret = credentials['key'][3]

# Create authentication
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
# Add tokens for connection
authenticate.set_access_token(accessToken, accessTokenSecret)
# Create connection with Twitter - API
api_twitter = tweepy.API(authenticate)

```

*Ilustración 31. Credenciales de la API. Información tomada de la presente investigación. Elaborado por Merly Avila.*

Con estas credenciales se tuvo acceso a Twitter y de esta manera se pudo extraer la información con la que se realizaría el análisis de las cuentas oficiales de los equipos; @BarcelonaSC, @CSEmelec, @IDV\_EC, @LDU\_Oficial. Con la función tweepy. Cursor y el modulo de search con la query a usar, se pueden visualizar la cantidad de tweets extraídos mediante la función print.



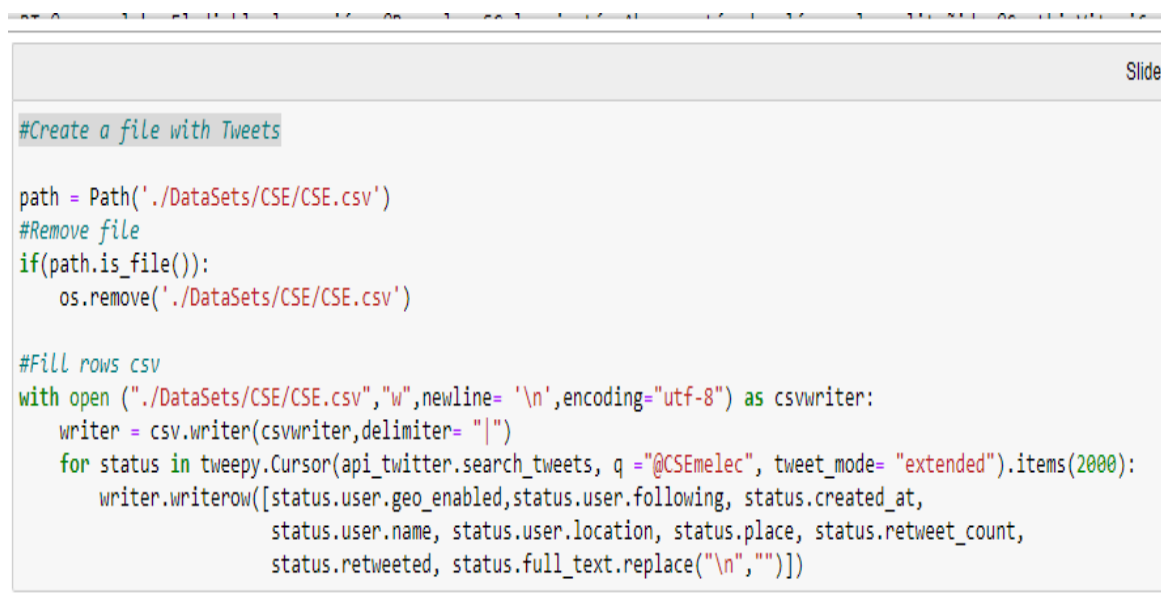
```
In [36]:
# Get Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@CSEmelec").items(100):
    print(status.text)

RT @BarcelonaSC: 🇹🇿 Presidente Carlos @AlfaroFloreno: "Damos la bienvenida al Estadio Banco Pichincha, nuestra casa, a Sudamérica y el mund...
RT @omarmaluk: El diablo los crió y @BarcelonaSC los juntó. Ahora están de pláceme la peliteñida @CynthiaViteri6 y el huevona sso @LassoGuill...
RT @BarcelonaSC: 🇹🇿 Alejandro Domínguez, Presidente @CONMEBOL: "El Estadio Banco Pichincha es hermoso, es un templo del fútbol sudamericano...
RT @ccarrasco750: La directiva de @BarcelonaSC muestra su inconformidad con la Comisión Nacional de Arbitraje, luego de que a l club no se l...
RT @billyelectric: Mientras la alcaldesa @CynthiaViteri6 y el banquero bachiller @LassoGuillermo despilfarran el dinero del pueblo en su eq...
@omarmaluk @BarcelonaSC @CynthiaViteri6 @LassoGuillermo @Libertadores Ya saben ni el primer partido del mundial gan... https://t.co/EyVu93OmSp
RT @omarmaluk: El diablo los crió y @BarcelonaSC los juntó. Ahora están de pláceme la peliteñida @CynthiaViteri6 y el huevona sso @LassoGuill...
@omarmaluk @BarcelonaSC @CynthiaViteri6 @LassoGuillermo @Libertadores XXXXXX
@omarmaluk @BarcelonaSC @CynthiaViteri6 @LassoGuillermo @Libertadores Totalmente de acuerdo XXXXXX incluso yo siendo hin... https://t.co/m3KntQLoxV
@omarmaluk @BarcelonaSC @CynthiaViteri6 @LassoGuillermo @Libertadores Par de ratas inundas
```

**Ilustración 32.** Tweets extraídos de la red social. Información tomada de la presente investigación. Elaborado por Merly Avila.

Luego de visualizar los tweets se procede a almacenar la información en un archivo csv, se ordena la ruta en la que se guardara el CSV mediante las funciones path y os, permitiendo la realización de las carpetas donde se almacenara. Con un “with open” se abrirá el doc csv y su ruta sin espacios y con el lenguaje Utf-8.

En un “writer” se le asignaran los nombres a los campos delimitados con separador, se le asignan atributos a cada Tweet, como se muestra en la ilustración 33.



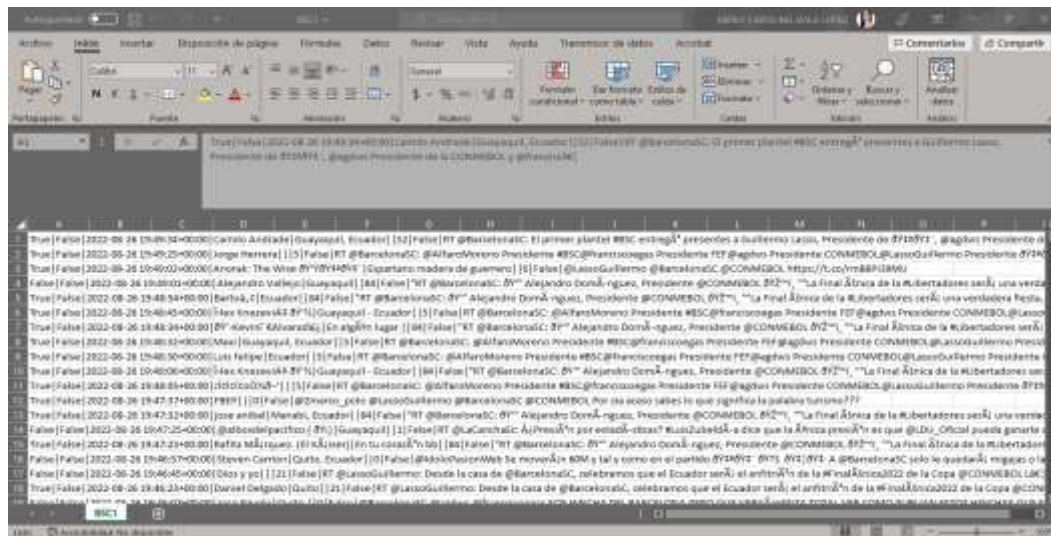
```
#Create a file with Tweets

path = Path('./DataSets/CSE/CSE.csv')
#Remove file
if(path.is_file()):
    os.remove('./DataSets/CSE/CSE.csv')

#Fill rows csv
with open ("./DataSets/CSE/CSE.csv","w",newline= '\n',encoding="utf-8") as csvwriter:
    writer = csv.writer(csvwriter,delimiter= "|")
    for status in tweepy.Cursor(api_twitter.search_tweets, q="@CSEmelec", tweet_mode="extended").items(2000):
        writer.writerow([status.user.geo_enabled,status.user.following, status.created_at,
            status.user.name, status.user.location, status.place, status.retweet_count,
            status.retweeted, status.full_text.replace("\n","")])
```

**Ilustración 33.** Crear carpeta donde se almacena el Csv. Información tomada de Jupyter notebook. Elaborado por Merly Avila.

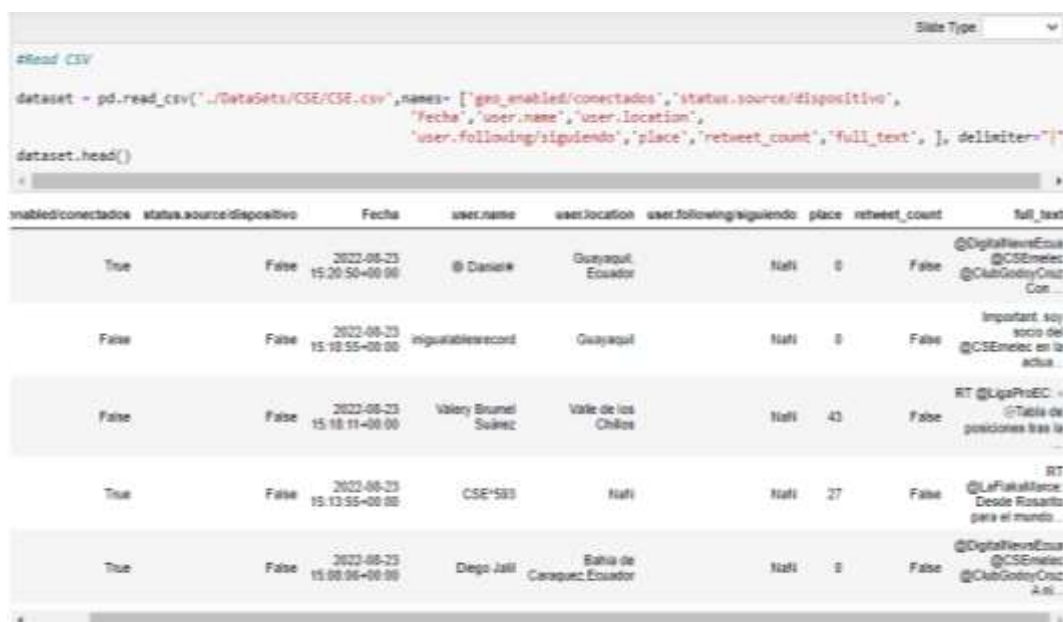
En la siguiente ilustración se muestra el archivo Excel de los tweets guardados en formato Csv.



**Ilustración 34.** Excel generado con los tweets. Información tomada de la presente investigación. Elaborado por Merly Avila.

### 3.6.3 Extracción de datos del csv.

Para la realización de este paso se utiliza la librería pandas, procediendo a asignarle los mismos nombres a las tablas, pero de forma abreviada de los atributos antes mencionados a su vez con el delimitador para poder obtener una separación de columnas, los datos se mostraran en datos generados en archivos data frame haciendo uso de la función “dataset”, permitiendo la visualización de los datos en forma ordenada y con cada uno de los atributos asignados, para mostrar la tabla del data frame se utiliza la función “dataset.head()”. Como se muestra en la ilustración # 35



**Ilustración 35.** Lectura del CSV. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

### 3.6.4 Limpieza y Validación de información.

Limpiar y validar la información, es una de las fases importantes de la propuesta de proyecto ya que establece reglas de validación y elimina datos no válidos conocidos que pueden causar distorsión en el análisis y presentar resultados erróneos. (Erl, Khattak, & Buhler, 2016)

Esta fase es de mucha importancia, ya que se procede a eliminar del texto; los signos, pronombres, símbolos, dígitos, urls ya que son datos innecesarios. Con la función Stopwords se eliminan las palabras que no son de mucha utilidad o no cumplen ninguna función dentro del texto. En la ilustración # se puede visualizar las funciones para crear los dataset del texto limpio, almacenándose en archivos csv. Para la visualización del dataset se utiliza la función “dataset.head” mostrando la tabla con la fila “tweets” donde se muestra el texto total mente limpio( Ver ilustración # 36)

```
# Stringify text
import cleantext
import unicode
import re

def cleanText(text):
    #Remove accent
    text = unicode.unidecode(text)
    #Remove Tags and hashstags
    text = re.sub(r'@[A-Za-z0-9]+', '', text)
    text = re.sub(r'#[A-Za-z0-9]+', '', text)
    #Remove others
    text = cleantext.clean(text, no_emoji=True, lower=True)
    text = cleantext.replace_currency_symbols(text, '')
    text = cleantext.replace_urls(text, '')
    text = cleantext.replace_emails(text, '')
    text = cleantext.replace_digits(text, '')
    text = cleantext.remove_punct(text)
    return text
```

**Ilustración 36.** Codificación para la limpieza del texto. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

```

#Function StopWords
import unicode

#Link Stopwords https://countwordsfree.com/stopwords

def removeStopWords(text,lang='en'):
    stopwords_free = []
    language = 'english'

    if lang == 'es':
        language = 'spanish'
        stopwords_free = [cleanText(line.strip()) for line in open("./Resources/stopwords_es.txt", 'r',encoding='utf-8')]
    else:
        stopwords_free = [cleanText(line.strip()) for line in open("./Resources/stopwords_en.txt", 'r',encoding='utf-8')]

    filtered_list = []
    text = cleanText(text)
    stop_words = stopwords.words(language)
    stop_words.extend(stopwords_free)
    words = word_tokenize(text)
    for w in words:
        if w.lower() not in stop_words:
            filtered_list.append(w)

    return ' '.join(filtered_list)

```

**Ilustración 37** Función stopwords para la eliminación de palabras. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

```

#Clean empty and NaN tweets
dataset.dropna(subset=['full_text'], inplace=True)
indexes = dataset[dataset['full_text'] == ""].index
dataset.drop(indexes, inplace=True)

#Remove StopWords of BSC.csv
dataset['tweet'] = dataset['full_text'].apply(lambda x : removeStopWords(x,'es'))

```

```

#Create CSV Clean
dataset.to_csv('./DataSets/CSE/CSE-Clean.csv', sep='|')

```

```

dataset.head()

```

ctados	status.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count	full_text	tweet
True	False	2022-08-23 15:20:50+00:00	Daniel*	Guayaquil, Ecuador	NaN	0	False	@DigitalNewsEcu @CSEmelec @ClubGodoyCruz Con ...	español mierda messi
False	False	2022-08-23 15:18:55+00:00	inigualablesrecord	Guayaquil	NaN	0	False	Important, soy socio del @CSEmelec en la actua...	important socio actualidad condicion cancelar ...
False	False	2022-08-23 15:18:11+00:00	Valery Brumel Curaçao	Valle de los Chillón	NaN	43	False	RT @LigaProEC: * @Tabla de	rt tabla posiciones fecha fase

**Ilustración 38.** Dataset del texto limpio. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Con el texto limpio se procede a realizar la traducción de este al idioma inglés, ya que el análisis a desarrollar en este proyecto usa librerías que para tener mejor predicción de

información recibe los datos en inglés. A continuación, en la ilustración 39 se muestra la línea de código que se usó para la traducción.

```
# Function translate tweets
from deep_translator import GoogleTranslator

def translate_Text(text):
    return GoogleTranslator(source='es', target='en').translate(text)

def translate_Tweet(text):
    if not text :
        return ""
    else:
        return translate_Text(text)
```

**Ilustración 39.** Línea de código de la traducción de tweets. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Una vez finalizado el proceso de traducción que dura aproximadamente de 15 a 30 min (Varía según la cantidad de tweets), se muestra el dataset con la última columna llamada “Tweet\_EN”

Slide Type

```
#Tranlate full_text
dataset["tweet_EN"] = dataset['tweet'].apply(translate_Tweet)
dataset.head()
```

us.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count	full_text	tweet	tweet_EN
False	2022-08-23 15:20:50+00:00	Daniel*	Guayaquil, Ecuador	NaN	0	False	@DigitalNewsEcu @CSEmelec @ClubGodoyCruz Con...	espanol mierda messi	spanish fuck messi
False	2022-08-23 15:18:55+00:00	inigualablesrecord	Guayaquil	NaN	0	False	Important, soy socio del @CSEmelec en la actua...	important socio actualidad condicion cancelar ...	important partner current condition cancel ins...
False	2022-08-23 15:18:11+00:00	Valery Brumel Suárez	Valle de los Chillos	NaN	43	False	RT @LigaProEC: » Tabla de posiciones tras la ...	rt tabla posiciones fecha fase ligapro betcris...	rt table positions date phase ligapro betcris...
False	2022-08-23 15:13:55+00:00	CSE*593	NaN	NaN	27	False	RT @LaFlakaMarce: Desde Rosarito para el mundo...	rt rosarito mundo directamente ecuador pte eme...	rt rosarito world directly ecuador pte emelext...
False	2022-08-23 15:08:06+00:00	Diego Jalil	Bahia de Caraquez, Ecuador	NaN	0	False	@DigitalNewsEcu @CSEmelec @ClubGodoyCruz Ani...	q vayan	go

**Ilustración 40.** Dataset con la columna Tweet\_EN. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Cuando el texto se traduzca se crea un archivo Csv con la recolección de todos los tweets ya en inglés, estos serán un utilizados para realizar el análisis de sentimientos.



```
#Create csv translate
dataset.to_csv('./Datasets/CSE/CSE-EN.csv', sep="|")
```

**Ilustración 41.** Crear el csv con los tweets traducidos. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Luego se hace una lectura de sentimientos traducidos.

```
#Read sentiment translate
sentiment = pd.read_csv('./DataSets/CSE/CSE-EN.csv', delimiter="|")

#clean csv
sentiment.dropna(subset=['tweet', 'tweet_EN'], inplace=True)
#del. barcelona sentiment['Unnamed: 0']
sentiment = sentiment.rename(columns={'Unnamed: 0': 'ID'})
#Remove Stopwords of BSC.csv
sentiment['tweet_EN'] = sentiment['tweet_EN'].apply(removeStopWords)
sentiment.head()
```

is.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count	full_text	tweet	tweet_EN
False	2022-08-23 15:20:50+00:00	@Daniel*	Guayaquil, Ecuador		NaN	0	@DigitalNewsEcu @CSEmelec @ClubSodoyCruz Con...	español mierda messi	spanish fuck messi
False	2022-08-23 15:18:55+00:00	Inigualablesrecord	Guayaquil		NaN	0	Important soy socio del @CSEmelec en la actua...	important socio actualidad condicion cancelar...	important partner current condition cancel me
False	2022-08-23 15:18:11+00:00	Valery Brunel Suárez	Valle de los Chillos		NaN	43	RT @LigaProEC: + @Tabla de posiciones tras la	rt tabla posiciones fecha fase ligapro before...	rt table positions date phase ligapro before...

**Ilustración 42.** Lectura de palabras que expresen sentimientos. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Una vez realizado todo este proceso, empieza la fase en donde se realizará el análisis de sentimientos de la información extraída. Se busca la polaridad de cada uno de los datos con el objetivo de observar la interacción y conformidad de los usuarios en Twitter sobre los equipos seleccionados.

### 3.6.5 Análisis de datos (data mining): Analizar el sentimiento.

Al proceso de analizar información, se puede catalogar como un análisis exploratorio, estando vinculado con la minería de datos.

El análisis de sentimiento es una forma técnica de análisis de texto está enfocada en determinar las emociones de individuos. Determina la actitud del autor del texto mediante el análisis del texto dentro del contexto del lenguaje natural. El análisis de sentimiento no solo facilita información sobre cómo se sienten los individuos, sino también la intensidad del sentimiento del mismo. Las aplicaciones comunes para el análisis de sentimientos incluyen identificar la satisfacción del cliente o insatisfacción temprana, midiendo el éxito o fracaso del producto y detectando nuevas tendencias. (Erl, Khattak, & Buhler, 2016)

Para iniciar con esta fase, como primer paso del análisis de sentimiento se debe utilizar ciertas librerías (Ver ilustración 46) y códigos de programación que le asigne una polaridad a cada tweet que fue extraído conociendo de esta manera la opinión y satisfacción del usuario que redactó el tweet sobre ese equipo.

```
#Resources Stopwords
nltk.download('stopwords')
nltk.download('punkt')

[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!

True
```

```
nltk.download('vader_lexicon')

[nltk_data] Downloading package vader_lexicon to
[nltk_data] C:\Users\Dell\AppData\Roaming\nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
```

**Ilustración 43.** Importación de la librería a usar para el análisis de sentimientos. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Se define un procedimiento para reconocer automáticamente su polaridad. Las palabras clave extraídas se procesan mediante un análisis de polaridad, para así poder determinar las connotaciones positivas/negativas de los datos.

```
#Analysis Polarity
#nltk analysis
sia = SentimentIntensityAnalyzer()
result = {}
for i, row in tqdm(sentiment.iterrows(), total = len(sentiment)):

    text = row['tweet_EN']
    id = row["ID"]
    result[id] = sia.polarity_scores(text)

print(result)
```

**Ilustración 44.** Ejecución del código para el análisis de polaridad, Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En la ilustración 45 se muestra el código que se usó para la asignación de polaridad a cada uno de los tweets los resultados se mediante un medidor de progreso para ello se usa la

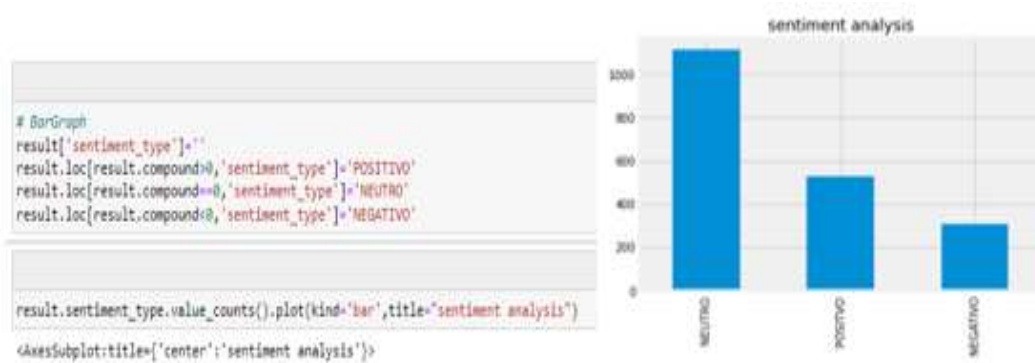


librería “tqdm” Para generar estos datos en un dataset se utiliza la función que se muestra en la ilustración # y se muestra a través de la función “result.head()”.

Slide Type <span>▼</span>										
<pre>#Create the Result DataFrame result = pd.DataFrame(result).T result = result.reset_index().rename(columns={'index': 'ID'}) result.count() sentiment.count() result = result.merge(sentiment, how='left')</pre>										
Slide Type <span>▼</span>										
result.head(10)										
ID	neg	neu	pos	compound	geo_enabled/conectados	status.source/dispositivo	Fecha	user.name	user.location	user.following/sigui
0	0	0.636	0.364	0.000	-0.5423	True	False	2022-08-23 15:20:50+00:00	Daniel*	Guayaquil, Ecuador
1	1	0.355	0.444	0.202	-0.4215	False	False	2022-08-23 15:18:55+00:00	inigualablesrecord	Guayaquil
2	2	0.000	0.769	0.231	0.3400	False	False	2022-08-23 15:18:11+00:00	Valery Brumel Suárez	Valle de los Chillos

**Ilustración 45.** Dataset con los resultados de la polaridad en cada Tweet. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Para visualizar los datos, se realizó un gráfico de barras y diagrama pastel con un porcentaje general de los Tweets positivos, negativos, neutros de un equipo. En la siguiente grafico se visualiza las funciones utilizadas para generar el grafico de barras y el diagrama pastel.



**Gráfico 1.** Gráfico de barras con la polaridad de los tweets. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.



los resultados para obtener valor del análisis y posteriormente tener la capacidad de proporcionar retroalimentación. Los resultados de la etapa de visualización de datos brindan a los usuarios la capacidad de realizar un análisis visual, lo que permite el descubrimiento de respuestas a las preguntas que los usuarios tienen aún ni siquiera formuladas. (Erl, Khattak, & Buhler, 2016)

### **3.7 Resultados**

Para presentar los resultados de los datos generados por tweets de las cuentas oficiales de los equipos; BarcelonaSC, CSEmelec, IDV\_EC, LDU\_Oficial. Aplicando la metodología distribuida en 6 fases establecida anteriormente.

Estas consisten en realizar y asegurar la validez de los datos, y seguir con la secuencia del proceso del análisis con la finalidad de obtener los datos listos para el uso y la implementación del análisis de sentimientos y poder mostrar los resultados de cómo piensan y que emoción comparten sobre los equipos de fútbol mencionados.

Esto se realiza haciendo uso de un análisis exploratorio con la técnica del análisis de polaridad, se ha desarrollado en un Jupyter Notebook ya que es de uso factible y se puede trabajar de forma ordenada, mostrando como resultado la clasificación de polaridad en cada tweet escrito, con valores para los comentarios “positivos >1”, “negativos<0” y “neutros =0”

Se mostrará una nube de palabras de cada equipo con las palabras que se repitan con frecuencia en el texto y a su vez los gráficos de barras y diagramas pastel del porcentaje de la polaridad en cada uno de los equipos.

#### **3.7.1 @BarcelonaSC**

Para realizar este análisis de sentimientos de este equipo fue mediante 2000 tweets que fueron extraídos en lapso de tiempo debido a que siempre los usuarios están activos opinando a cerca de su equipo. En la ilustración 48 se puede visualizar un data set con los tweets a analizar.

Total interacciones

AR@River...

RT @FunerariaAlach: @BarcelonaSC Amigos de @BarcelonaSC, nos ponemos a disposición de todos los directivos, jugadores, cuerpo técnico, soci...

@MiroslavJd @ardazad @BarcelonaSC @LDU\_Oficial Todos los partidos son como cualquier otro para ustedes jajajajajaja

@carolarr2 @LassoGuillermo @BarcelonaSC @CONMEBOL Usted si es peliona, que nostalgia no estar de acuerdo en temas p... <https://t.co/9xmXfp8jg7>

RT @BarcelonaSC: ¡Uffffff!!! Paco Rodríguez 🐼 <https://t.co/FoI4X6z265>

RT @degolladismo: @Barcerick18\_1 @BarcelonaSC <https://t.co/3RI6MAwKRL>

RT @BarcelonaSC: 🐼 #GualaceoBSC #LigaProBetcris

#VamosÍdolo 🐼 <https://t.co/mqgiq7NMB9>

@ardazad @BarcelonaSC @LDU\_Oficial Esto de polemizar si es o no clásico fue una cortina de humo en caso de perder... <https://t.co/p1bRlaoV9H>

RT @DeporFinanzas: 🌟 ¡Los clubes @CONMEBOL de cada país más populares en redes sociales durante julio 2022! 🌟

Total interacciones

**Ilustración 48.** Tweets extraídos del equipo de BarcelonaSC. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Luego se procede a utilizar los datos ya procesados en cada una de las fases de la metodología que se implementó con la debida corrección para una mejor comprensión de datos al momento de realizar el análisis de polaridad, como muestra la ilustración 49, clasificando cada uno de los comentarios con la numeración de polaridad ya asignada durante el proceso con valores 1 y 0.

```
0% | 0/1948 [00:00<?, ?it/s]

[0: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 1: {'neg': 0.0, 'neu': 0.714, 'pos': 0.286, 'compound': 0.4215},
2: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 3: {'neg': 0.0, 'neu': 0.648, 'pos': 0.352, 'compound': 0.5859}, 4:
{'neg': 0.141, 'neu': 0.489, 'pos': 0.37, 'compound': 0.6369}, 5: {'neg': 0.162, 'neu': 0.541, 'pos': 0.297, 'compound': 0.36
12}, 6: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 7: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 8:
{'neg': 0.537, 'neu': 0.463, 'pos': 0.0, 'compound': -0.7003}, 9: {'neg': 0.0, 'neu': 0.575, 'pos': 0.425, 'compound': 0.571
9}, 10: {'neg': 0.77, 'neu': 0.23, 'pos': 0.0, 'compound': -0.7717}, 11: {'neg': 0.0, 'neu': 0.775, 'pos': 0.225, 'compound':
0.4404}, 13: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 14: {'neg': 0.301, 'neu': 0.515, 'pos': 0.184, 'compoun
d': -0.3818}, 15: {'neg': 0.0, 'neu': 0.345, 'pos': 0.655, 'compound': 0.5859}, 16: {'neg': 0.688, 'neu': 0.312, 'pos': 0.0,
'compound': -0.765}, 17: {'neg': 0.0, 'neu': 0.648, 'pos': 0.352, 'compound': 0.5859}, 18: {'neg': 0.394, 'neu': 0.367, 'po
s': 0.239, 'compound': -0.1779}, 19: {'neg': 0.0, 'neu': 0.714, 'pos': 0.286, 'compound': 0.4215}, 20: {'neg': 0.0, 'neu': 0.
714, 'pos': 0.286, 'compound': 0.4939}, 21: {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}, 22: {'neg': 0.0, 'neu': 0.
357, 'pos': 0.643, 'compound': 0.5574}, 23: {'neg': 0.0, 'neu': 0.714, 'pos': 0.286, 'compound': 0.4215}, 24: {'neg': 0.759,
'neu': 0.241, 'pos': 0.0, 'compound': -0.743}, 25: {'neg': 0.0, 'neu': 0.593, 'pos': 0.407, 'compound': 0.6705}, 26: {'neg':
```

**Ilustración 49.** Análisis de polaridad. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En la ilustración 50 se muestra el dataset con los datos y numeración de polaridad.

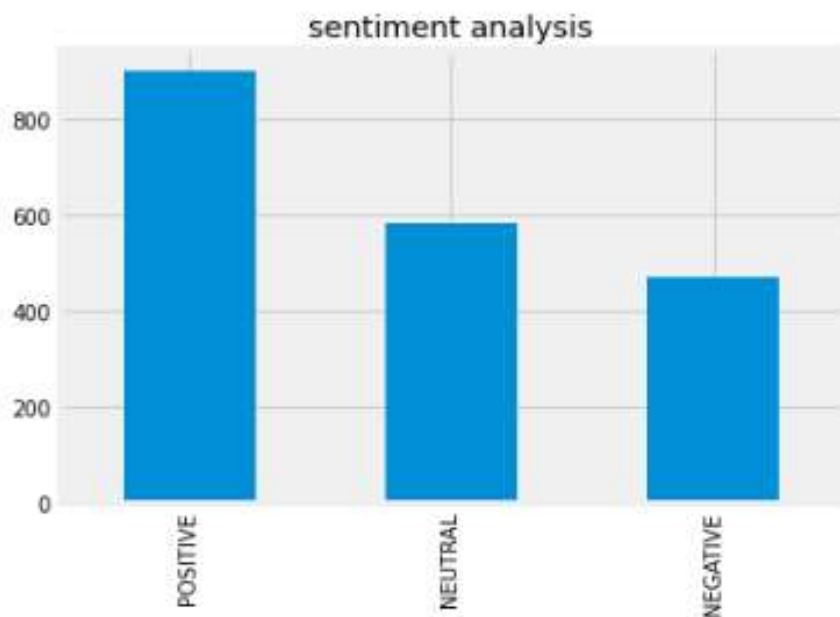
Slide Type ▼

```
barcelona_result.head(10)
```

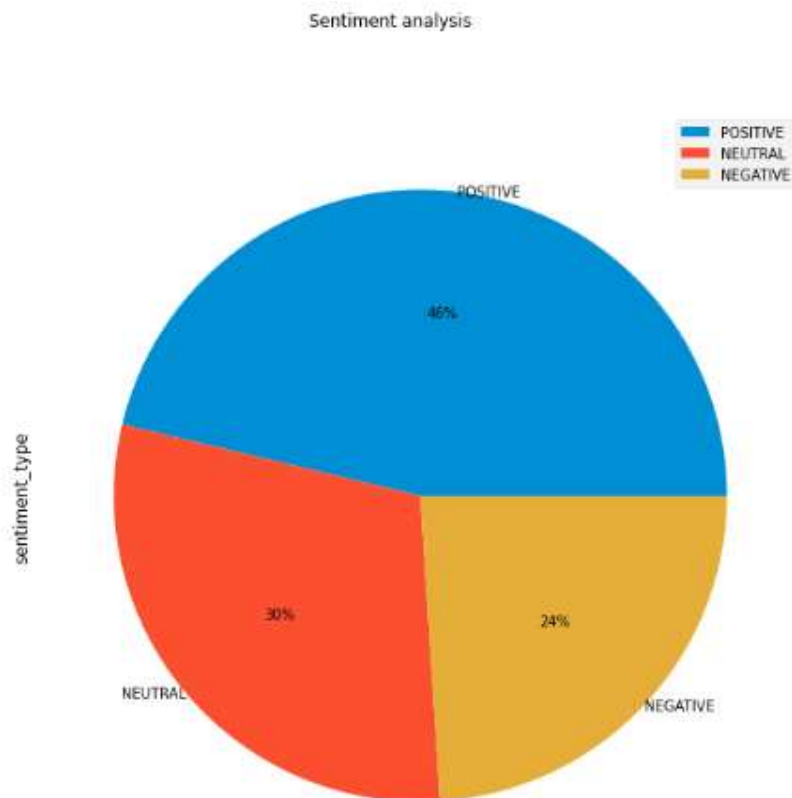
	ID	neg	neu	pos	compound	geo_enabled	conectados	status.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo
0	0	0.000	1.000	0.000	0.0000		False	False	2022-08-29 22:17:32+00:00	Odiseo, o sea nadie	NaN	NaN
1	1	0.000	0.714	0.286	0.4215		False	False	2022-08-29 22:17:13+00:00	Matias Montanaro Garozzo	NaN	NaN
2	2	0.000	1.000	0.000	0.0000		True	False	2022-08-29 22:16:49+00:00	Ferlopez_BSC	guayaquil - ecuador	NaN
3	3	0.000	0.648	0.352	0.5859		False	False	2022-08-29 22:16:39+00:00	Fabrizio Carvajal	NaN	NaN
4	4	0.141	0.489	0.370	0.6369		True	False	2022-08-29 22:16:28+00:00	Oscar	Quito, Ecuador	NaN

**Ilustración 50.** Dataset con resultados de polaridad tweets BarcelonaSC. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Se puede observar cómo cada comentario tiene un valor, clasificando de esta manera los comentarios positivos, negativos y neutros del equipo analizado. Para la visualización de los gráficos de barra y diagrama pastel se hizo una concatenación de los datos emitidos. Dejando un porcentaje total de los tweets según la valoración positiva, negativa y neutra que se hayan generado. Ver ilustración # del diagrama de barras y la # del diagrama pastel.



**Gráfico 3.** Gráfico de Barras equipo de BarcelonaSC. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.



**Gráfico 4.** Diagrama Pastel equipo BarcelonaSC. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En el gráfico se puede observar que el equipo de Barcelona el 46% de los tweets son positivos, el 24% dicen palabras negativas y el 30% opiniones neutrales.

### 3.7.2 @CSEmelec

Para realizar el análisis de sentimiento de este equipo al igual que el equipo anterior se trabajó con 2000 tweets, extraídos en una semana del mes de agosto como se puede ver en la ilustración 51.

```

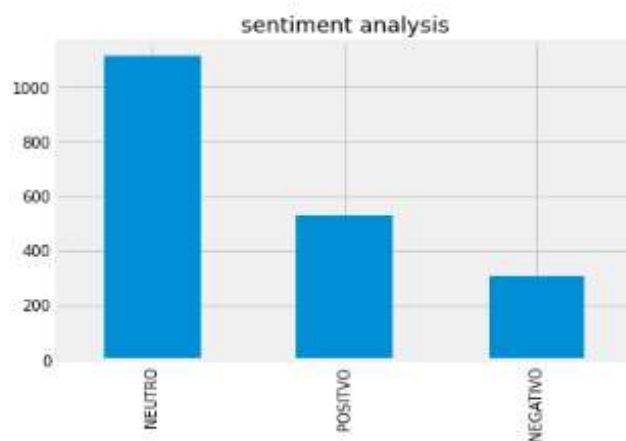
similar y mejor
RT @BarcelonaSC: T Alejandro Domínguez, Presidente @COMIBOL

"La Final única de la #Libertadores será una verdadera fiesta, un hecho h...
RT @ccarrasco758: La directiva de @BarcelonaSC muestra su inconformidad con la Comisión Nacional de Arbitraje, luego de que al club no se le...
@David8807188253 @pedropeladinesm @llobgunner @BarcelonaSC Liga en la Sierra tiene más que Cmelec.
RT @devilla75: @LassoGuillermo @BarcelonaSC @COMIBOL Otra obra magistral de tu gobierno !!!!! Te felicito. ANNN , p...
on cierto , hoy...
RT @ccarrasco758: La directiva de @BarcelonaSC muestra su inconformidad con la Comisión Nacional de Arbitraje, luego de que al club no se le...
@presidencia_ec @COMIBOL @Libertadores @Comunicacionec @Deportec @BarcelonaSC @Turismec @alcaldiaec q Alguien L. http://t.co/tbsh3dVtNT
RT @ccarrasco758: La directiva de @BarcelonaSC muestra su inconformidad con la Comisión Nacional de Arbitraje, luego de que al club no se le...
@LassoGuillermo @BarcelonaSC @COMIBOL Ascoooo renuncia eres el peor del mundo, ni todo el dinero del mundo te deja dormir @Ppkas04 @BOBILLO82 @workpage4 @ElJojoBenavides @LDU_Oficial @UDV_EC @BarcelonaSC @elnacionalec @CSEmelec @SDQUITO. http://t.co/hiyta2M537
RT @utpl: ¡¡@utpl firmó convenio con @BarcelonaSC en #Guayaquil como parte de su compromiso con la #EducaciónSuperior del país.

```

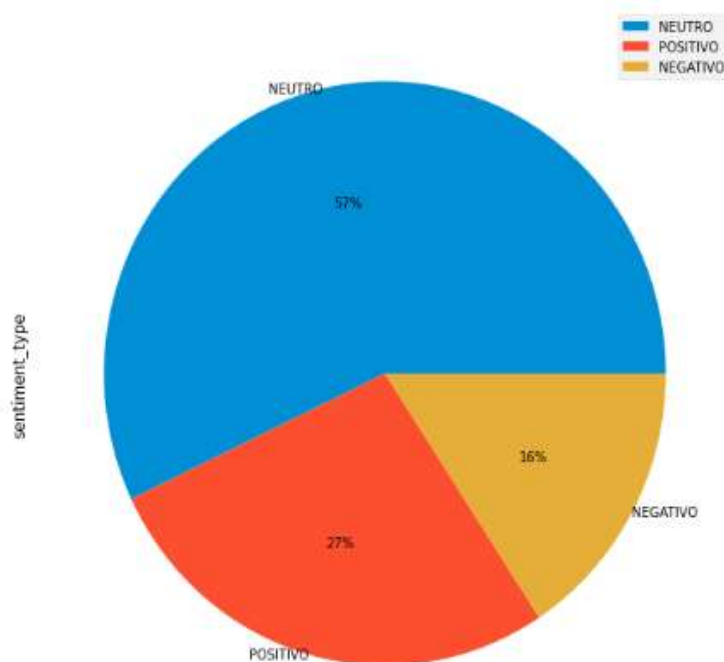
**Ilustración 51.** Tweets del Equipo de CSEmelec. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En el un cuaderno de Jupyter, codificando de forma ordenada bloque por bloque. Se realizó la traducción de los datos procesados ya que las librerías a utilizar para tener mejor predicción de información reciben los datos en inglés. Los resultados obtenidos se presentan mediante grafico pastel y diagrama de barras. Ver gráfico 5



**Gráfico 5.** Gráfico de barras del CSEmelec. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En el diagrama pastel se visualiza el porcentaje y como resultado se obtuvo que en el equipo de Emelec el 57% de los comentarios tienen una polaridad neutra, el 27% son positivos y el 16% de opiniones son negativas.



**Gráfico 6.** Diagrama pastel del CSEmelec. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.



En la ilustración 52 se muestra el dataset con la polaridad en cada tweet en columnas en el dataset está a columna “compound” es la que muestra el valor concatenado de la polaridad.

Slide Type

result.head()

ID	neg	neu	pos	compound	geo_enabled/conectados	status.source/dispositivo	Fecha	user.name	user.location	use
0	0	0.636	0.364	0.000	-0.5423	True	False	2022-08-23 15:20:50+00:00	Daniel*	Guayaquil, Ecuador
1	1	0.355	0.444	0.202	-0.4215	False	False	2022-08-23 15:18:55+00:00	inigualablesrecord	Guayaquil
2	2	0.000	0.789	0.231	0.3400	False	False	2022-08-23 15:18:11+00:00	Valery Brumel Suárez	Valle de los Chillos
3	3	0.000	1.000	0.000	0.0000	True	False	2022-08-23 15:13:55+00:00	CSE*593	NaN
4	4	0.000	0.000	0.000	0.0000	True	False	2022-08-23 15:08:06+00:00	Diego Jalil	Bahia de Caraquez,Ecuador

**Ilustración 52.** Dataset con la polaridad de los tweets del equipo de CSEmelec. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

### 3.7.3 @LDU\_Oficial.

Se hizo el análisis con 2000 tweets, así mismo se realizó la extracción durante una semana del mes de agosto, se pudo extraer esta cantidad debido a que los usuarios comparte su opinión con mucha frecuencia en esta red social. En la ilustración 53 se muestran los tweets que se pudieron extraer.

```
@LassoGuillermo @BarcelonaSC @CONMEBOL Por sia acaso sabes lo que significa la palabra turismo???
RT @BaqueP: @2marco_polo @LassoGuillermo @BarcelonaSC @CONMEBOL Es un acto que le compete a la Presidencia de la Republica, para n
ada, ni al Mini.
RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la C
opa @CONMEBOL L.
@robertoarturocj @CisPardo @Lucas_Matteo_ @ElJojoBenavides @Luisd_galarza @LDU_Oficial @IDV_EC @BarcelonaSC. https://t.co/I4
gIDbwe21
@RetroAucas @xavi8083 @CisPardo @Lucas_Matteo_ @ElJojoBenavides @Luisd_galarza @LDU_Oficial @IDV_EC @BarcelonaSC. https://t.
co/9Ll5wSZa6X
RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la C
opa @CONMEBOL L.
RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la C
opa @CONMEBOL L.
RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la C
opa @CONMEBOL L.
@BarcelonaSC @goltvecu Que seria de Lino si se hubiera quedado 2 años más al mando de Almada.
RT @BarcelonaSC: 🇪🇦 Alejandro Domínguez, Presidente @CONMEBOL

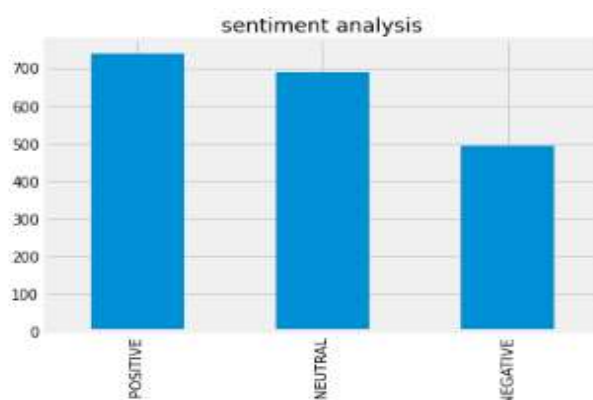
🇪🇦 "La Final Única de la #Libertadores será una verdadera fiesta, un hecho h...
```

**Ilustración 53.** Tweets del equipo de LDU, Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.



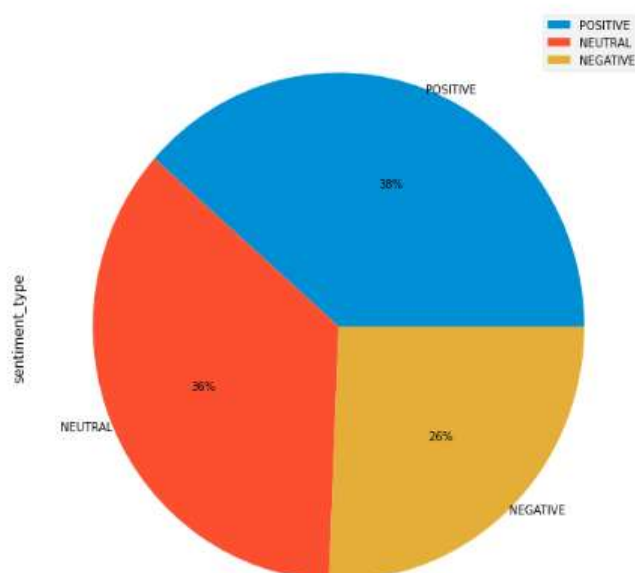
Al igual que el análisis en los equipos anteriores se ejecutó en Jupyter Notebook, bloque por bloque, traducir los datos para mejor lectura de datos al usar las librerías que permitan clasificar la información se muestran los resultados obtenidos de forma ordenada a través de gráficos de barra, el porcentaje total de la polaridad en un diagrama pastel y el dataset con las polaridades distribuidas en columnas con nombre pos, neg, neu y la columna llamada compount que muestra los resultados de la polaridad de cada tweets en forma concatenada.

En las siguientes ilustraciones se muestran los resultados.



**Gráfico 7.** Diagrama de barras del equipo LDU. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

Como resultado de polaridad en este equipo, el diagrama pastel muestra un porcentaje de 38% de comentarios son positivos, el 26% negativos y el 36% neutrales.



**Gráfico 8.** Diagrama pastel con la polaridad del equipo LDU Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En la ilustración 54 se mostrará como ya antes mencionado la polaridad en cada columna y el resultado de la concatenación de estas.

Slide Type ▼

result.head()

	ID	neg	neu	pos	compound	geo_enabled/conectados	status.source/dispositivo	Fecha	user.name	user.location	user.following/siguiend
0	0	0.000	0.575	0.425	0.5719	True	False	2022-08-25 04:49:32+00:00	Cesar Procel N.	UT: -2.151748,-79.893159	Na
1	1	0.000	0.400	0.600	0.8402	False	False	2022-08-25 04:49:16+00:00	WHITE TIGER	En algún lugar de la costa.	Na
2	2	0.000	1.000	0.000	0.0000	True	False	2022-08-25 04:32:29+00:00	alejandra rullova	Buenos Aires, Argentina	Na
3	3	0.000	0.745	0.255	0.3400	True	False	2022-08-25 04:30:10+00:00	Fevernova	Peru	Na
4	4	0.224	0.694	0.082	-0.5423	False	False	2022-08-25 04:20:58+00:00	J. Jara	Quito, Ecuador	Na

**Ilustración 54.** Dataset con los resultados de polaridad del equipo LDU. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

### 3.7.4 @IDV\_EC

Al igual que los equipos anteriores en este se realizó el análisis con 2000 Tweets extraídos durante el mes de agosto realizando el mismo proceso en un cuaderno de Jupyter como se puede ver en la ilustración 55

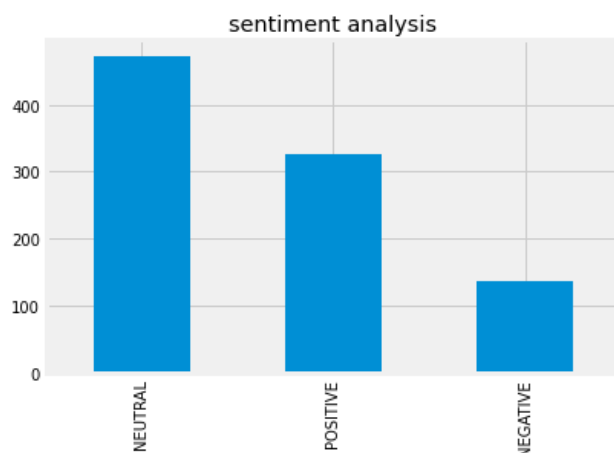
```
@LassoGuillermo @BarcelonaSC @CONMEBOL En qué consiste un "destino deportivo"?
Por qué GYE sería un destino deporti... https://t.co/Suu2MOM8iK
@robertoarturocj @CisPardo @Lucas_Matteo_ @ElJojoBenavides @Luisd_galarza @LDU_Oficial @IDV_EC @BarcelonaSC... https://t.co/3CkJKTP1S7

RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la Co
pa @CONMEBOL L...
@BarcelonaSC @agdw @franciscoegas Egas conoció a Dios
RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la Co
pa @CONMEBOL L...
@BarcelonaSC @AlfaroMoreno @franciscoegas @agdw @LassoGuillermo @CynthiaViteri6 @SebasPalaciosSP @nielsolsenp Que... https://t.co/FQrMUM2fI4
@cheong_wilson @emartinez666 @BarcelonaSC Ok, muchas gracias por la corrección?
@BarcelonaSC @AlfaroMoreno @franciscoegas @agdw @LassoGuillermo @CynthiaViteri6 @SebasPalaciosSP @nielsolsenp Como... https://t.co/jO4XYzszuj
RT @LassoGuillermo: Desde la casa de @BarcelonaSC, celebramos que el Ecuador será el anfitrión de la #FinalÚnica2022 de la Co
pa @CONMEBOL L...
@RetroAucas @CisPardo @Lucas_Matteo_ @ElJojoBenavides @Luisd_galarza @LDU_Oficial @IDV_EC @BarcelonaSC... https://t.co/YkiIT29Ih9
RT @BarcelonaSC: Estadio Banco Pichincha 🏟️
```

**Ilustración 55.** Tweets del equipo Independiente del Valle. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

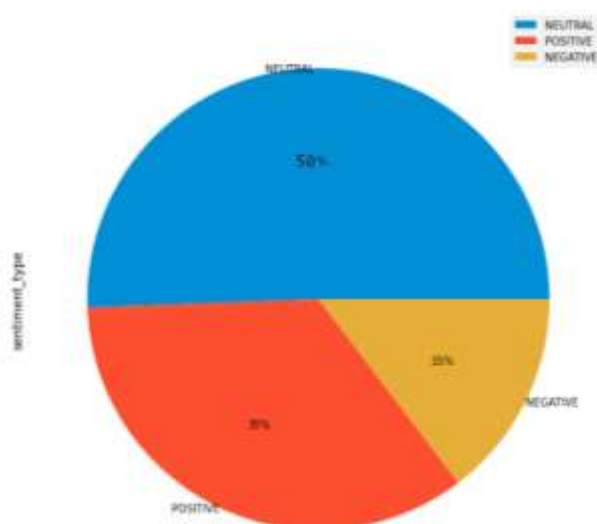
Se obtuvieron los resultados mediante gráficos de barras y diagramas pastel, mostrando a su vez el dataset con la polaridad por columnas y está la columna de la información concatenada.

En el gráfico de barras se muestra la respuesta de los datos en forma longitudinal y la comparación de porcentajes de cada categoría.



**Gráfico 9.** Diagrama de barras del equipo IDV. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

En el diagrama pastel se muestra que este equipo el 38% de comentarios son positivos, el 26% son negativos y el 36% son comentarios neutrales como muestra el gráfico 10.



**Gráfico 10.** Diagrama Pastel con la polaridad del equipo IDV. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

El dataset de este equipo al igual que los 3 anteriores muestra la polaridad en columnas con el resultado de la concatenación de cada item de polaridad.

ID	neg	neu	pos	compound	geo_enabled/conectados	status.source/dispositivo	Fecha	user.name	user.location	user.foll
0	0	0.00	0.598	0.402	0.5719	False	False	2022-09-02 03:55:30+00:00	Juanito	Quito
1	1	0.00	1.000	0.000	0.0000	True	False	2022-09-02 03:54:05+00:00	Oscar	Quito, Ecuador
2	3	0.00	1.000	0.000	0.0000	False	False	2022-09-02 03:48:13+00:00	Luis Noboa	Paysand
3	4	0.00	1.000	0.000	0.0000	False	False	2022-09-02 03:47:32+00:00	Marcelo Chila	NaN
4	5	0.00	0.330	0.670	0.7269	False	False	2022-09-02 03:43:14+00:00	DANIEL	NaN
5	6	0.09	0.711	0.199	0.3182	False	False	2022-09-02 03:38:34+00:00	josÃ	landÃn ecuador
6	7	0.00	0.513	0.487	0.5859	False	False	2022-09-02 03:38:23+00:00	Esteban	LDU
7	8	0.00	1.000	0.000	0.0000	False	False	2022-09-02 03:37:56+00:00	Matti	Quito Ecuador

**Ilustración 56** Dataset con los resultados de polaridad de los tweets del equipo IDV. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

### 3.7.4 Comparación de los resultados de los cuatro equipos y la nube de palabras.

En la siguiente tabla se muestran la unión de los porcentajes de cada uno de los equipos y uso solo grafico de barras comparando los resultados de cada uno de ellos.

**Tabla 13.** Porcentajes concatenados de los cuatro equipos

POLARIDAD	BARCELONA	EMELEC	LDU	IDV
POSITIVO	47%	27%	38%	35%
NEUTRO	30%	57%	36%	50%
NEGATIVO	24%	16%	26%	15%

Información tomada de la presente investigación. Elaborado por Merly Avila.

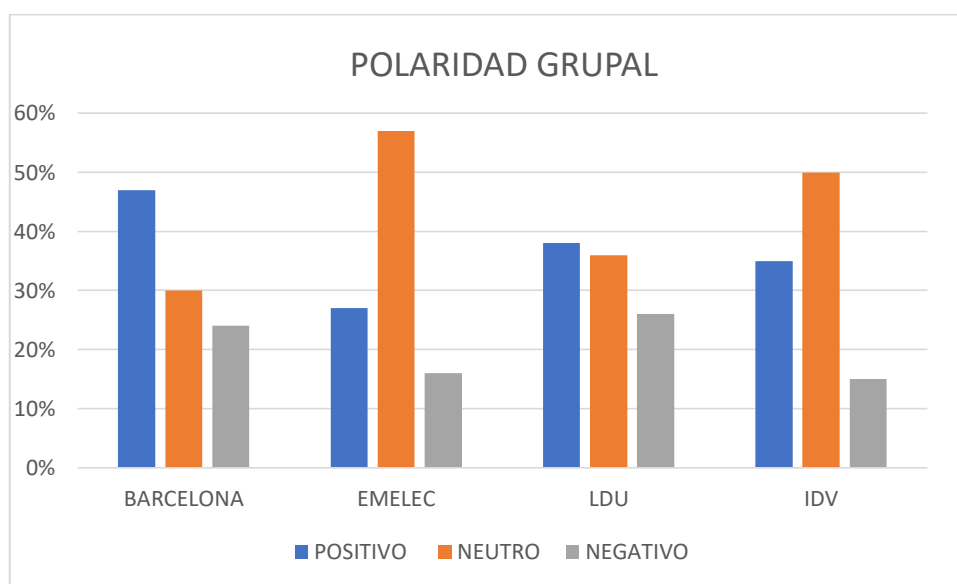
Se puede observar que en el equipo de Barcelona el mayor porcentaje lo tienen los comentarios positivos al contrario del equipo de Emelec que muestra un porcentaje de comentarios positivos más bajos que los demás equipos con un 27%, IDV con 35% y LDU con un 38% de positivos.

En los tweets neutros el equipo con mayor porcentaje de todos es el de Emelec con un porcentaje de 57% no son ni buenos ni malos en su mayoría, siguiéndole los equipos de IDV con un 50% y LDU con el porcentaje en neutralidad de 36% y por último Barcelona con un 30% de Tweets neutros.

Los comentarios negativos los equipos con un poco el porcentaje más alto que los otros equipos lo tienen el IDV Y LDU con el 26%, le sigue Barcelona con el 24% de polaridad negativa y para finalidad el equipo de Emelec con 16% de comentarios negativos.

Según los porcentajes mostrados en la tabla indican que el equipo con mayor aceptación o buenos comentarios es el de Barcelona, seguido de Independiente del Valle y Liga de Quito dejando por último al equipo de Emelec.

El siguiente grafico de barras, muestra las respuestas de cada uno de los equipos unificadas.



**Gráfico 11.** Diagrama de barras con los datos de los 4 equipos. Información tomada de la presente investigación. Elaborado por Merly Avila.

La nube de palabras fue creada para los comentarios tanto positivos como negativos, teniendo en cuenta que se mostraran las palabras que con más frecuencia se repetían en grande y las de menos frecuencia serían las pequeñas. En las siguientes ilustraciones se muestran las nubes de palabras de cada uno de los equipos de un lado de palabras positivas y del otro las palabras negativas.



**Ilustración 57.** Nube de palabras positivas y negativas equipo Barcelona. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.





**Ilustración 58.** Nube de palabras positivas y negativas del equipo de Emelec. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.



**Ilustración 59.** Nube de palabras positivas y negativas del equipo LDU. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.



**Ilustración 60.** Nube de palabras positivas y negativas del equipo de IDV. Información tomada de Jupyter con datos de la presente investigación. Elaborado por Merly Avila.

### 3.7.8 Conclusión

En esta investigación se llegó a la conclusión que el trabajar mediante la red social Twitter fue el entorno más factible para la realización de este análisis de sentimientos. Para tener acceso a los datos públicos de Twitter se necesitaron credenciales que permitan la extracción de estos, creando una cuenta en el developer de la aplicación, la Api permitió la extracción de 8.000 tweets entre los 4 equipos de la serie A del fútbol ecuatoriano. Esto se realizó en el lapso de una semana del mes de agosto.

La herramienta que se utilizó para este proceso de minado de datos fue Jupyter Notebook siendo de gran utilidad para el desarrollo del código y de esta manera realizar el análisis de los datos obtenidos de la red social haciendo uso de las librerías mencionadas en el capítulo 3 que fueron de gran utilidad para la realización de cada uno de los procesos permitiendo desarrollar la extracción, clasificación, limpieza, validación y análisis de sentimientos de cada uno de los datos.

El cuaderno de Jupyter permitió trabajar de forma ordenada, sin cometer errores en cada proceso cada vez que se ejecutaban los códigos.

La aplicación de un análisis de sentimientos en este trabajo dio a saber las emociones a través de opiniones emiten los usuarios en cada tweet del equipo de su elección y a su vez permitió por medio de estos datos la clasificación de polaridad, mostrando como resultados los datos positivos, negativos y neutros.

El análisis aplicado a cada equipo se realizó en cuadernos separados teniendo como respuesta a cada equipo con porcentajes que varían según el nivel de aceptación o de interacción de sus hinchas en Twitter.

Para finalizar se puede decir que en este trabajo se cumplió paso a paso cada uno de los objetivos, indicando la metodologías y herramientas que se utilizaron para el desarrollo del mismo, siguiendo paso a paso cada una de las fases planteadas y mostrándolo los resultados mediante gráficos y nubes de palabras.

### 3.7.9 Recomendaciones

Visualizando los resultados obtenidos del análisis de sentimiento en tweets del fútbol ecuatoriano, se recomienda usar una cantidad de tweets considerables de un tema en específico, ya que el uso de poca información puede no mostrar los resultados precisos o poco favorables.

Como recomendación se debe realizar este tipo de análisis con cuentas que estén en constante interacción, para poder trabajar de manera factible y con obtener los resultados deseados, trabajar de forma ordenada en el Notebook, estos análisis de datos requieren seguir una secuencia de pasos para la obtención de resultados precisos, crear carpetas que faciliten la ruta al momento de ejecutar el código y así cada que se realice la llamada en el Jupyter Notebook y muestre los dataset del código, de esta manera evitar confusiones cuando el programa se esté ejecutando importar las librerías que se utilizan para el análisis correspondiente.

Antes de iniciar con el análisis se debe de crear una cuenta en el developer para la obtención de credenciales y así tener acceso a los datos públicos de la red social.

Como ultima recomendación, se debe investigar en otras fuentes, diferentes atributos que ayuden a la obtención de los tweets, más información precisa sobre la implementación de este tipo de análisis en el fútbol.

Revisar el anexo 3 los pasos a seguir para crear la cuenta en el developer de Twitter y así obtener las credenciales necesarias para la realización del análisis.



## **Glosario de palabras.**

**Api** “Interfaz de programación de aplicaciones” conjunto de funciones que permiten la integración de sistemas, permitiendo reutilizar sus funcionalidades en otro software

**Lenguaje Utf-8** formato estándar que sirve para almacenar caracteres Unicode asignándoles una cadena de bits determinada.

**Tweets** mensaje publicado en la red social Twitter que contiene textos, imágenes o videos.

**Atributos** características especiales que se le da a un objeto para su identificación, determinando su estado o cualidades.

**Patrones** sucesos recurrentes, es una serie de variables constantes que se pueden identificar dentro de un conjunto de datos.

**Subjetividad** percepción personal acerca de un asunto, idea o alguna cultura. Está asociada a las emociones y sentimientos al expresar alguna idea sobre algo en específico.

**Polaridad** elementos de importancia de una opinión, hace referencia si la opinión es negativa o positiva.

**Dataset** representación de datos estructurados de uno o más archivos.

**ANEXOS**

## Anexo 1

## Ley de protección de datos personales.

Registro Oficial - Quinto Suplemento N° 459

Miércoles 25 de mayo de 2021



## CERTIFICACIÓN

En mi calidad de Secretario General de la Asamblea Nacional, me permito **CERTIFICAR** que los días 09 y 11 de febrero 2021, la Asamblea Nacional discutió en primer debate el **"PROYECTO LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES** y, en segundo debate el día 10 de mayo de 2021, siendo en esta última fecha finalmente aprobado.

Quito, 11 de mayo de 2021.



**DR. JAVIER RUBIO DUQUE**  
Secretario General

**LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES****CAPÍTULO I****ÁMBITO DE APLICACIÓN INTEGRAL**

**Artículo 1.- Objeto y finalidad.-** El objeto y finalidad de la presente ley es garantizar el ejercicio del derecho a la protección de datos personales, que incluye el acceso y decisión sobre información y datos de este carácter, así como su correspondiente protección. Para dicho efecto regula, prevé y desarrolla principios, derechos, obligaciones y mecanismos de tutela.

**Artículo 2.- Ámbito de aplicación material.-** La presente ley se aplicará al tratamiento de datos personales contenidos en cualquier tipo de soporte, automatizados o no, así como a toda modalidad de uso posterior. La ley no será aplicable a:

- a) Personas naturales que utilicen estos datos en la realización de actividades familiares o domésticas;
- b) Personas fallecidas, sin perjuicio de lo establecido en el artículo 28 de la presente Ley;
- c) Datos anonimizados, en tanto no sea posible identificar a su titular. Tan pronto los datos dejen de estar disociados o de ser anónimos, su tratamiento estará sujeto al cumplimiento de las obligaciones de esta ley, especialmente la de contar con una base de licitud para continuar tratando los datos de manera no anonimizada o disociada;
- d) Actividades periodísticas y otros contenidos editoriales;

## Anexo 2

## Decreto-Ejecutivo-N-1014

Nº 1014

RAFAEL CORREA DELGADO

PRESIDENTE CONSTITUCIONAL DE LA REPÚBLICA

## CONSIDERANDO:

Que en el apartado g) del numeral 6 de la Carta Iberoamericana de Gobierno Electrónico, aprobada por el IX Conferencia Iberoamericana de Ministros de Administración Pública y Reforma del Estado, realizada en Chile el 1 de Junio de 2007, se recomienda el uso de estándares abiertos y software libre, como herramientas informáticas;

Que es el interés del Gobierno alcanzar soberanía y autonomía tecnológica, así como un significativo ahorro de recursos públicos y que el Software Libre es en muchas instancias un instrumento para alcanzar estos objetivos;

Que el 18 de Julio del 2007 se creó e incorporó a la estructura orgánica de la Presidencia de la República la Subsecretaría de Informática, dependiente de la Secretaría General de la Administración, mediante Acuerdo Nº119 publicado en el Registro Oficial No. 139 de 1 de Agosto del 2007;

Que el numeral 1 del artículo 6 del Acuerdo Nº 119, faculta a la Subsecretaría de Informática a elaborar y ejecutar planes, programas, proyectos, estrategias, políticas, proyectos de leyes y reglamentos para el uso de Software Libre en las dependencias del gobierno central; y,

En ejercicio de la atribución que le confiere el numeral 9 del artículo 171 de la Constitución Política de la República;

## DECRETA:

Artículo 1.- Establecer como política pública para las Entidades de la Administración Pública Central la utilización de Software Libre en sus sistemas y equipamientos informáticos.

Artículo 2.- Se entiende por Software Libre, a los programas de computación que se pueden utilizar y distribuir sin restricción alguna, que permitan su acceso a los códigos fuentes y que sus aplicaciones puedan ser mejoradas.

Estos programas de computación tienen las siguientes libertades:

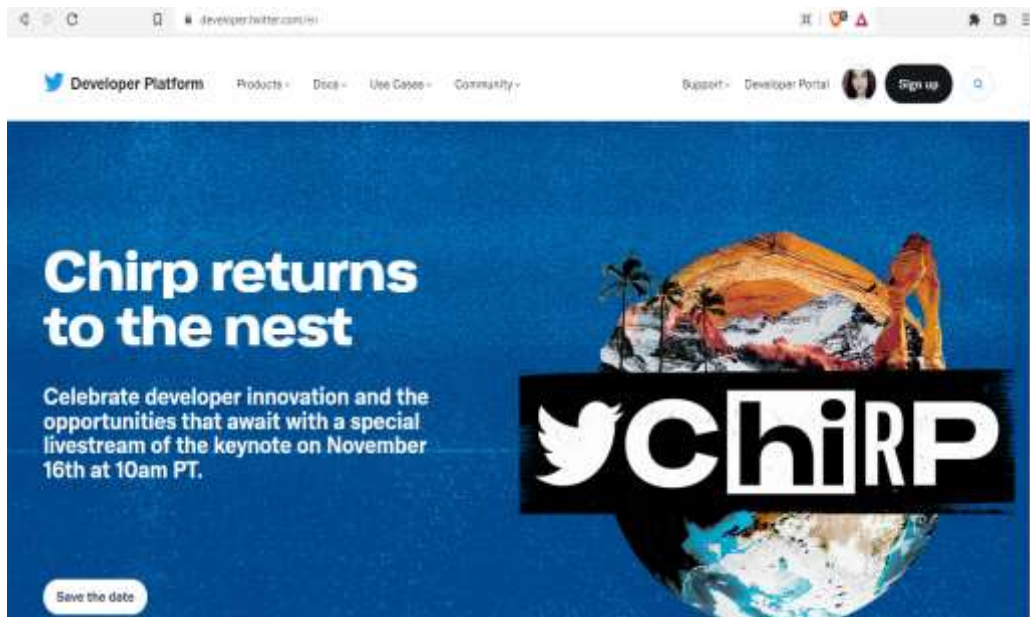
- a) Utilización del programa con cualquier propósito de uso común
- b) Distribución de copias sin restricción alguna.
- c) Estudio y modificación del programa (Requisito: código fuente disponible)
- d) Publicación del programa mejorado (Requisito: código fuente disponible).

Artículo 3.- Las entidades de la Administración Pública Central previa a la instalación del software libre en sus equipos, deberán verificar la existencia de capacidad técnica que brinde el soporte necesario para el uso de este tipo de software.

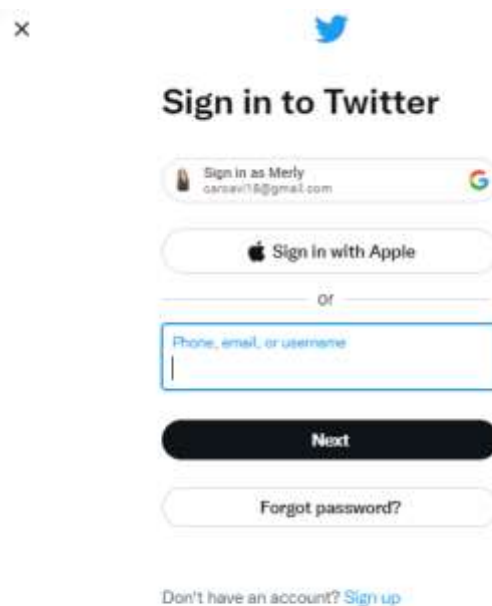
### Anexo 3

#### Pasos para crear cuenta en el Developer de Twitter.

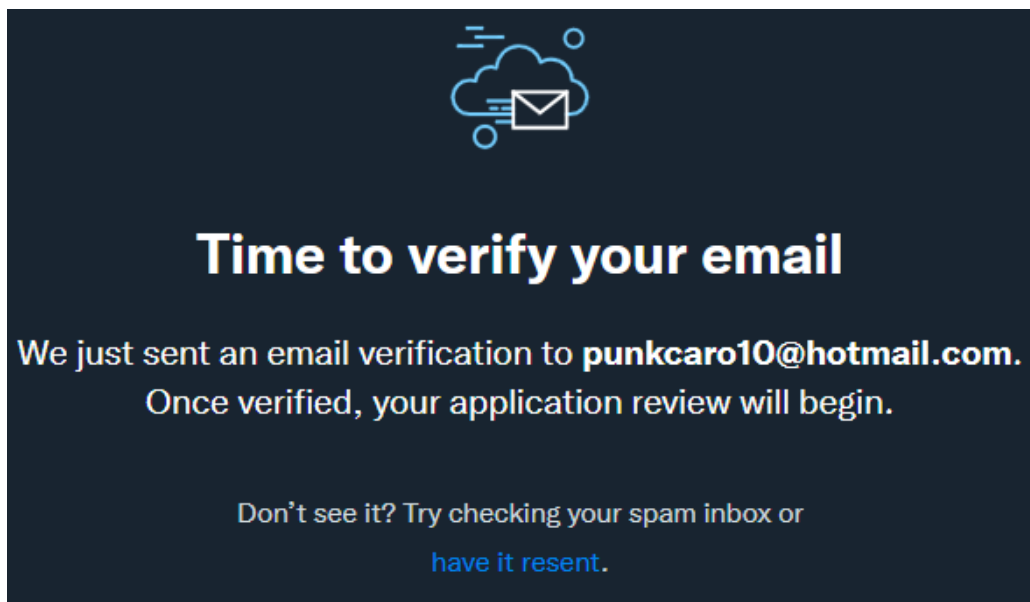
1. Ingresar
2. Ingresar a la pagina del developer. <https://developer.twitter.com/en>.



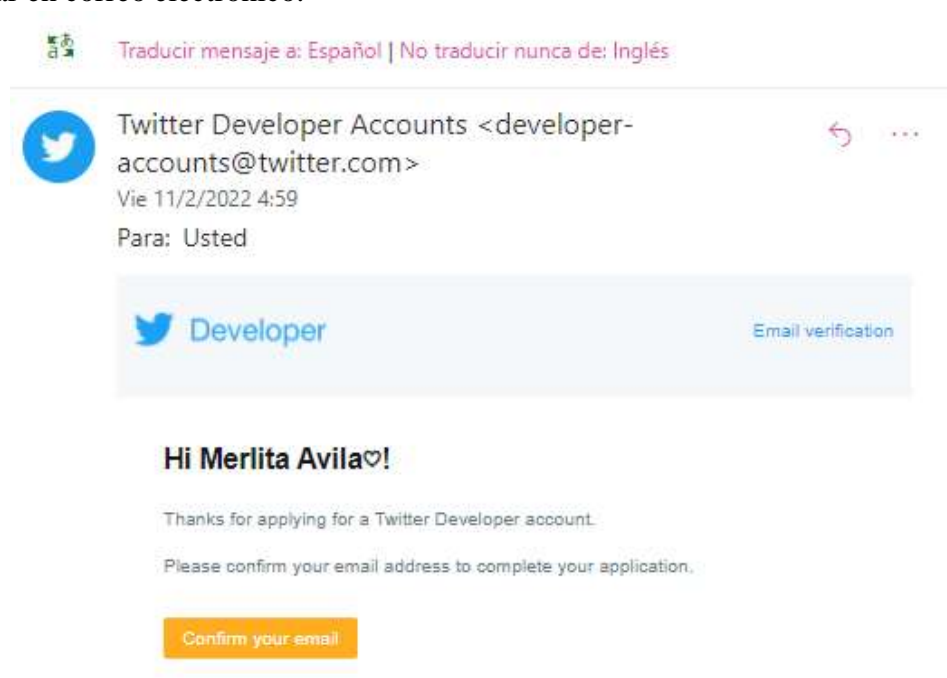
3. Se debe solicitar un permiso para crear una cuenta como desarrollador.
4. Al solicitar el permiso, en la siguiente página le pide el inicio de sesión de la cuenta que vaya a utilizar.



5. Una vez se haya iniciado sesión, la página siguiente muestra unas opciones a seleccionar, se debe dar clic en la opción “Academic for education of research” y dar clic en “next”.
6. En la siguiente página se deberá llenar un formulario con los datos requeridos. El ultimo formula preguntas acerca del uso que les dará a los datos de la red social. Una vez el formulario esté listo y acepte términos y condiciones se mostrará una página donde pide la verificación del correo electrónico.



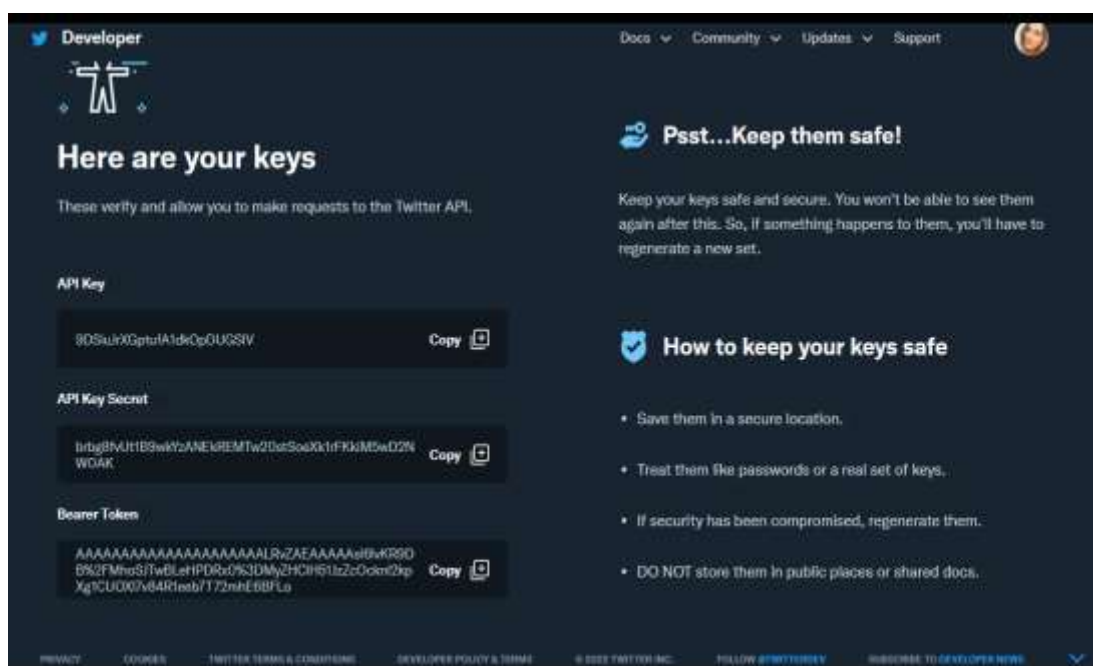
7. Confirmar en correo electrónico.



8. Una vez verificada la información se procede a crear la cuenta de desarrollador. Escribir nombre para la App en este caso será “MerlyAvila”.

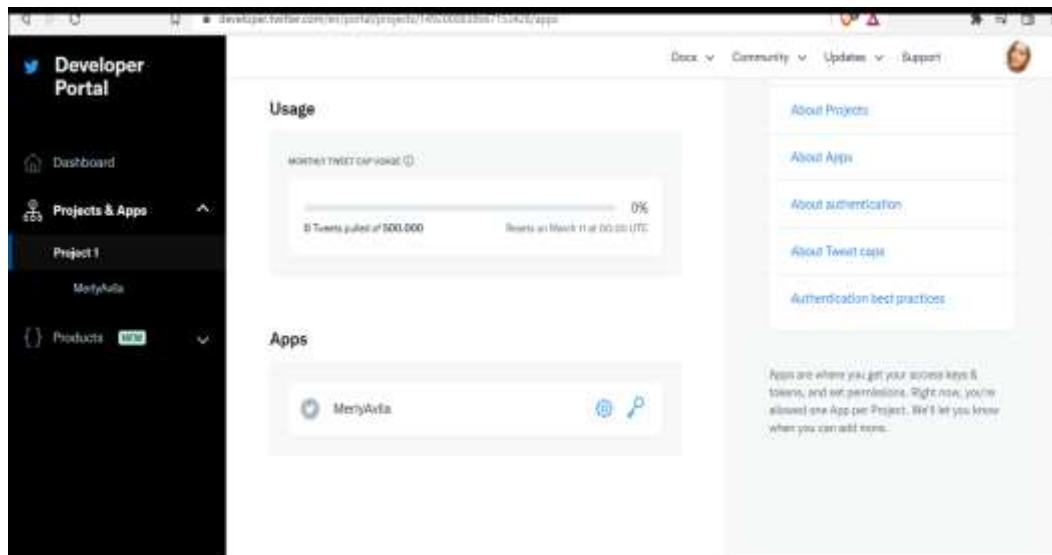


9. Se generan algunas claves, estas se deben almacenar en un documento a parte para su posterior uso.

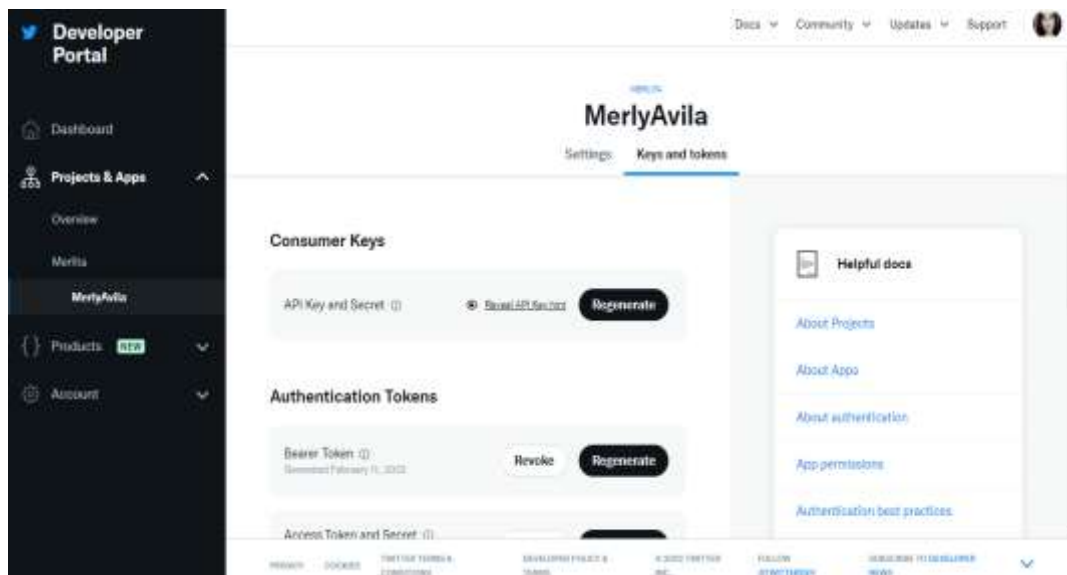




10. Luego de guardar las contraseñas, en la siguiente página se da clic en “let’s go”. Se observa el portal de la cuenta del developer dar clic en “dashboard”.



11. Se le asigna un nombre al proyecto.



## Anexo 4

### Código de programación aplicado para la realización del análisis.

#### Análisis de sentimientos en la red social Twitter.

```
# Twitter Sentiment analysis

# packages
import os
from pathlib import Path
import tweepy
import csv
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
from nltk.sentiment import SentimentIntensityAnalyzer
import nltk
from tqdm.notebook import tqdm
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
plt.style.use('fivethirtyeight')
```

```
#Reading Credentials
#Change Credentials for you user
credentials = pd.read_csv('./Credentials/CredentialsStudent.csv')
consumerKey = credentials['key'][0]
consumerSecret = credentials['key'][1]
accessToken = credentials['key'][2]
accessTokenSecret = credentials['key'][3]
```

```
# Create authentication
authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
# Add tokens for connection
authenticate.set_access_token(accessToken, accessTokenSecret)
# Create connection with Twitter - API
api_twitter = tweepy.API(authenticate)
```

#### Líneas para la adquisición de datos.

```
# Get Barcelona Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@BarcelonaSC").items(100):
    print(status.text)
```

```
#Create a file with Tweets

path = Path('./DataSets/BSC/BSC.csv')
#Remove file
if(path.is_file()):
    os.remove('./DataSets/BSC/BSC.csv')

#Fill rows csv
with open("./DataSets/BSC/BSC.csv", "w", newline= '\n', encoding="utf-8") as csvwriter:
    writer = csv.writer(csvwriter, delimiter= "|")
    for status in tweepy.Cursor(api_twitter.search_tweets, q="@BarcelonaSC", tweet_mode= "extended").items(2000):
        writer.writerow([status.user.geo_enabled, status.user.following, status.created_at,
                        status.user.name, status.user.location, status.place, status.retweet_count,
                        status.retweeted, status.full_text.replace("\n", "")])
```

## Extracción de datos.

```

#Create a file with Tweets

path = Path('./DataSets/BSC/BSC.csv')
#Remove file
if(path.is_file()):
    os.remove('./DataSets/BSC/BSC.csv')

#Fill rows csv
with open ('./DataSets/BSC/BSC.csv',"w",newline= '\n',encoding="utf-8") as csvwriter:
    writer = csv.writer(csvwriter,delimiter= "|")
    for status in tweepy.Cursor(api.twitter.search_tweets, q ="@BarcelonaSC", tweet_mode= "extended").items(2000):
        writer.writerow([status.user.geo_enabled,status.user.following, status.created_at,
                        status.user.name, status.user.location, status.place, status.retweet_count,
                        status.retweeted, status.full_text.replace("\n","")])

#Read CSV
barcelona_dataset = pd.read_csv('./DataSets/BSC/BSC.csv',names= ['geo_enabled/conectados','status.source/dispositivo',
                                                                'Fecha','user.name','user.location',
                                                                'user.following/siguiendo','place','retweet_count','full_text', ], delimiter="|")

barcelona_dataset.head()

```

## Limpieza, validación y traducción de datos.

```

# Stringify text
import cleantext
import unicode
import re

def cleanText(text):
    #Remove accent
    text = unicode.unidecode(text)
    #Remove Tags and hashtags
    text = re.sub(r'@[A-Za-z0-9]+','', text)
    text = re.sub(r'#[A-Za-z0-9]+','', text)
    #Remove others
    text = cleantext.clean(text, no_emoji=True, lower=True)
    text = cleantext.replace_currency_symbols(text, '')
    text = cleantext.replace_urls(text, '')
    text = cleantext.replace_emails(text, '')
    text = cleantext.replace_digits(text, '')
    text = cleantext.remove_punct(text)
    return text

#text = "RT @JohaReece: @LassoGuillermo @BarcelonaSC @CONMEBOL Así las prioridades de este menso superficial e irreflexivo."
#print(cleanTweet(text))

#Function StopWords
import unicode

#Link Stopwords https://countwordsfree.com/stopwords

def removeStopWords(text,lang='en'):
    stopwords_free = []
    language = 'english'

    if lang == 'es':
        language = 'spanish'
        stopwords_free = [cleanText(line.strip()) for line in open("./Resources/stopwords_es.txt", 'r',encoding='utf-8')]
    else:
        stopwords_free = [cleanText(line.strip()) for line in open("./Resources/stopwords_en.txt", 'r',encoding='utf-8')]

    filtered_list = []
    text = cleanText(text)
    stop_words = stopwords.words(language)
    stop_words.extend(stopwords_free)
    words = word_tokenize(text)
    for w in words:
        if w.lower() not in stop_words:
            filtered_list.append(w)

    return ' '.join(filtered_list)

#text = 'asi las prioridades de este menso superficial e irreflexivo'
#print(removeStopWords(text))

```

Slide Type

```

#Clean empty and NaN tweets
barcelona_dataset.dropna(subset=['full_text'],inplace=True)
indexes = barcelona_dataset[barcelona_dataset['full_text'] == ""].index
barcelona_dataset.drop(indexes, inplace=True)

#Remove StopWords of BSC.csv
barcelona_dataset['tweet'] = barcelona_dataset['full_text'].apply(lambda x : removeStopWords(x,'es'))

```

Slide Type

```

#Create CSV Clean
barcelona_dataset.to_csv('./DataSets/BSC/BSC-Clean.csv', sep='|')

```

Slide Type

```

barcelona_dataset.head()

```

Slide Type

```

# Function translate tweets
from deep_translator import GoogleTranslator

def translate_Text(text):
    return GoogleTranslator(source='es', target='en').translate(text)

def translate_Tweet(text):
    if not text :
        return ""
    else:
        return translate_Text(text)

```

Slide Type

```

#Tranlate full_text
barcelona_dataset["tweet_EN"] = barcelona_dataset['tweet'].apply(translate_Tweet)
barcelona_dataset.head()

```

Slide Type

```

#Create csv translate
barcelona_dataset.to_csv('./DataSets/BSC/BSC-EN.csv', sep="|")
barcelona_dataset.head()

```

## Análisis de sentimientos en los tweets.

Slide Type

```

#Read sentiment translate
barcelona_sentiment = pd.read_csv('./DataSets/BSC/BSC-EN.csv', delimiter="|")

#clean cvs
barcelona_sentiment.dropna(subset=['tweet','tweet_EN'],inplace=True)
#del barcelona_sentiment['Unnamed: 0']
barcelona_sentiment = barcelona_sentiment.rename(columns={'Unnamed: 0':'ID'})
#Remove StopWords of BSC.csv
barcelona_sentiment['tweet_EN'] = barcelona_sentiment['tweet_EN'].apply(removeStopWords)
barcelona_sentiment.head()

```

Slide Type

```

#Resources StopWords
nltk.download('stopwords')
nltk.download('punkt')

```

Slide Type

```

nltk.download('vader_lexicon')

```

```

Slide Type ▼
#Analysis Polarity
#nltk analysis
sia = SentimentIntensityAnalyzer()
result = {}
for i, row in tqdm(barcelona_sentiment.iterrows(), total = len(barcelona_sentiment)):

    text = row['tweet_EN']
    id = row["ID"]
    result[id] = sia.polarity_scores(text)

print(result)

```

```

Slide Type ▼
#Create the Result DataFrame
barcelona_result = pd.DataFrame(result).T
barcelona_result = barcelona_result.reset_index().rename(columns={'index': 'ID'})
barcelona_result.count()
barcelona_sentiment.count()
barcelona_result = barcelona_result.merge(barcelona_sentiment, how='left')

```

## Visualización de datos

### Diagrama de barras

```

Slide Type ▼
# BarGraph
barcelona_result['sentiment_type']=''
barcelona_result.loc[barcelona_result.compound>0,'sentiment_type']='POSITIVE'
barcelona_result.loc[barcelona_result.compound==0,'sentiment_type']='NEUTRAL'
barcelona_result.loc[barcelona_result.compound<0,'sentiment_type']='NEGATIVE'

barcelona_result.sentiment_type.value_counts().plot(kind='bar',title="sentiment analysis")

```

### Diagrama pastel.

```

Slide Type ▼
#Graph Pie
barcelona_result.sentiment_type.value_counts().plot(kind='pie',
                                                    subplots = True,
                                                    title="Sentiment analysis",
                                                    autopct='%1.0f%%',
                                                    y=None,
                                                    legend=True,
                                                    labeldistance=1,
                                                    figsize=(20, 10))

```

### Nube de palabras.

```

Slide Type ▼
positive_tweets = barcelona_result[barcelona_result['sentiment_type'] == 'POSITIVE']
negative_tweets = barcelona_result[barcelona_result['sentiment_type'] == 'NEGATIVE']

```

## Palabras positivas.

```
Slide Type ▼  
  
# Create a wordCloud Positive  
words = ''.join([word for word in positive_tweets['tweet']])  
#barcelona_sentiment['polarity'] = barcelona_sentiment['text_clean'].apply(is_empty)  
  
#print(words)  
wordCloud = WordCloud(width=1024, height=720, random_state=21, max_font_size=119).generate(words)  
  
plt.imshow(wordCloud, interpolation="bilinear")  
plt.axis('off')  
plt.show()
```

## Palabras negativas.

```
Sli  
  
# Create a wordCloud Negative  
words = ''.join([word for word in negative_tweets['tweet']])  
#barcelona_sentiment['polarity'] = barcelona_sentiment['text_clean'].apply(is_empty)  
  
#print(words)  
wordCloud = WordCloud(width=1024, height=720, random_state=21, max_font_size=119).generate(words)  
  
plt.imshow(wordCloud, interpolation="bilinear")  
plt.axis('off')  
plt.show()
```

## Bibliografía

- Ahlgren, M. (26 de Mayo de 2022). *Websiterating*. Obtenido de Websiterating: <https://www.websiterating.com/es/research/twitter-statistics/#chapter-1>
- Alfonso, I. (1995). *Tenicas de investigacion bibliografica. Tenicas de investigacion bibliografica*. Contexto Editores, Caracas.
- AliatUniversidades. (28 de Julio de 2020). *Universidad la Concordia*. Obtenido de Universidad la Concordia: <https://universidadlaconcordia.edu.mx/blog/index.php/tecnicas-de-investigacion/#:~:text=Las%20t%C3%A9cnicas%20de%20investigaci%C3%B3n%20son,conocimiento%20para%20resolver%20nuestras%20preguntas>.
- Alvardo, R. V. (2016). Analisis de sentimiento en Twitter de los socios de un club de futbol a través de la evaluacion de herramientas que manejan gran volumen de informacion. *Analisis de sentimiento en Twitter de los socios de un club de futbol a través de la evaluacion de herramientas que manejan gran volumen de informacion*. Universidad de Guayaquil, Guayaquil.
- Álvarez, C. M. (2011). Metodologia de la investigacion Cuantitativa y Cualitativa. *Metodologia de la investigacion Cuantitativa y Cualitativa*. Universidad SurColombiana, Neiva.
- Amores, M., Arco, L., & Artiles, M. (2015). Una herramienta para gestionar comentarios de la Web. *Revista Cubana de Ciencias Informaticas* , 20.
- Arancibia, J. G. (2009). Metodología para la definicon de requisitos en proyectos de Data Mining(ER-DM). *Metodología para la definicon de requisitos en proyectos de Data Mining(ER-DM)*. Departamento de lenguajes y sistemas informaticos e ingenieria de software, Madrid.
- Blasco, M., & Coenders, G. (2020). Análisis de sentimiento de la agenda de los partidos políticos españoles en Twitter durante la. *Análisis de sentimiento de la agenda de los partidos políticos españoles en Twitter durante la*. Universitat de Girona, Gerona, España.
- Blockchain. (23 de diciembre de 2021). *bbva*. Obtenido de bbva: <https://www.bbva.com/es/que-es-un-token-y-para-que-sirve/>
- Cabrera, E., & Díaz, E. (2018). Manual de uso de Jupyter Notebook. *Universidad Complutense de Madrid*. Universidad Complutense de Madrid, Madrid.
- Cendejas, J., Acuña, M., Cortes, G., & Bolaños, G. (2017). El uso de modelos y metodologías de minería de datos para la inteligencia de negocios. *Revista de sistemas computacionales y TIC'S*, 57.
- Chato970. (8 de Enero de 2015). *Chato970*. Obtenido de Chato970: <https://chato970.wordpress.com/2015/01/08/historia-de-la-seleccion-del-ecuador/>

- Córdova, C. (2019). Metodología basada en Minería de Datos para la detección de usuarios Influencers en Twitter. *Metodología basada en Minería de Datos para la detección de usuarios Influencers en Twitter*. Universidad Nacional de Trujillo, Trujillo, Perú.
- Decreto Ejecutivo N-1014. (2008). *Gestion Documental*. Quito.
- DeveloperTwitter. (2022). *DeveloperTwitter*. Obtenido de DeveloperTwitter: <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api>
- Díaz, N. C. (2006). Técnica de muestreo. Sesgos más frecuentes. *Técnica de muestreo. Sesgos más frecuentes*. Revistas Eden. Obtenido de <https://revistaseden.org/files/9-CAP%209.pdf>
- Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data Fundamentals Concepts, Drivers & Techniques*. Arcitura Education Inc.
- FEF. (2012). *Ecuador Mundial*. Guayaquil : Poligrafica.
- Fernandez, P., Vallejo, G., Livacic, P., & Tuerco, E. (2014). Validez Estructurada para una investigación cuasi-experimental de calidad. *Validez Estructurada para una investigación cuasi-experimental de calidad*. Universidad de Oviedo, España.
- Ferreira, M. (2020). Análisis de sentimiento en tweets de fútbol Argentino. *Análisis de sentimiento en tweets de fútbol Argentino*. Universidad Nacional de Córdoba, Córdoba.
- García, A. (13 de Noviembre de 2019). *El laberinto de Falken*. Obtenido de El laberinto de Falken : <https://www.ellaberintodefalken.com/2019/11/analisis-de-sentimiento-NLP.html>
- Gerardo Molina. (19 de Noviembre de 2020). *El Telegrafo*. Obtenido de El Telegrafo : <https://www.eltelegrafo.com.ec/noticias/fanatico/7/los-equipos-con-mas-simpatizantes-en-ecuador>
- Gutiérrez-Herrera, G. (2015). *Procesamiento del lenguaje natural aplicado al análisis de sentimiento de opiniones*. Universidad de Sevilla , Sevilla.
- IONOS. (05 de Noviembre de 2020). *Digital Guide Ionos*. Obtenido de Digital Guide Ionos : <https://www.ionos.es/digitalguide/online-marketing/analisis-web/analisis-de-sentimiento/>
- Jervis, T. M. (27 de Agosto de 2020). *Lifeder*. Obtenido de Lifeder: <https://www.lifeder.com/investigacion-descriptiva/>
- Ley Organica de Protección de datos Personales. (2021). *Quinto Suplemento*. Quito: Órgano de la República del Ecuador.
- LigaPro Betcris. (2022). *Ligapro.ec*. Obtenido de Ligapro.ec: <https://www.ligapro.ec/serie-a/equipos>
- Lillian, P. &. (2008). *Opinion mining and sentiment analysis*. The essence of knowledge.



- Loeber, R., & Farrington, D. (2015). Estudios longitudinales en la investigación de los problemas de conducta. *Estudios longitudinales en la investigación de los problemas de conducta*. Centre Lodres 94.
- López, E. A. (2011). POLITICA FISCAL Y ESTRATEGIA COMO FACTOR DE DESARROLLO DE LA MEDIANA EMPRESA COMERCIAL SINALOENSE. UN ESTUDIO DE CASO. *POLITICA FISCAL Y ESTRATEGIA COMO FACTOR DE DESARROLLO DE LA MEDIANA EMPRESA COMERCIAL SINALOENSE. UN ESTUDIO DE CASO*. Univversidad Autonoma de Sinaloa, Sinaloa.
- Martinez, E., Martín, M., & Ureña, L. A. (2011). *Analisis de Sentimiento*. Jaén: SINAI.
- Méndez, P. A. (2019). Metodología KDD para data mining. *Metodologia KDD para data mining*. Instituto Tecnologico Superior de los Rios.
- Moine, J. M. (2013). Metodologias para el descubrimiento de conocimientos en bases de datos: Un estudio comparativo. *Metodologias para el descubrimiento de conocimientos en bases de datos: Un estudio comparativo*. Universidad Nacional de la Plata, Buenos Aires.
- Montesinos, L. (2014). Analisis de sentimientos y prediccion de eventos en Twitter. *Analisis de sentimientos y prediccion de eventos en Twitter*. Universidad de Chile, Santiago de Chile.
- Narváez, M. (2017). APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA EL ANÁLISIS DE SENTIMIENTOS EN LAS REDES SOCIALES SOBRE PRODUCTOS DE LA MARCA FIDEOS CAYAMBE. *APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA EL ANÁLISIS DE SENTIMIENTOS EN LAS REDES SOCIALES SOBRE PRODUCTOS DE LA MARCA FIDEOS CAYAMBE*. Universidad de las Fuerzas Armadas, Sangolqui.
- Nectec. (27 de mao de 2019). *Nectec Global Knowledge*. Obtenido de Nectec Global Knowledge: <https://www.netec.com/post/mineria-de-datos-que-es-importancia-y-tecnicas-de-su-implementacion>
- Peiró, R. (2 de Julio de 2020 ). *Economipedia*. Obtenido de Economipedia: <https://economipedia.com/definiciones/base-de-datos.html>
- Perkins, J., Chopra, D., & Joshi, N. (2016). *Natural Language Processing: Python and NLTK*. Birmingham: Packt Publishing.
- Robledano, A. (23 de Septiembre de 2019). *OpenWebinars*. Obtenido de OpenWebinars: <https://openwebinars.net/blog/que-es-python/>
- Rodriguez, K., & Haber, Y. (2020). Análisis de sentimientos en Twitter aplicado al #impeachment de Donald Trump. *Análisis de sentimientos en Twitter aplicado al #impeachment de Donald Trump*. Universidad de Oriente, Cuba.
- Rondón, I. (4 de Febrero de 2022). *Eiposgrados.com*. Obtenido de Eiposgrados.com: <https://eiposgrados.com/blog-python/que-es-anaconda/>

- Sampieri, R. H., Collado, C. F., & LucioP, P. B. (1997). Proceso de investigacion. *Metodologia de la investigacion*. Mc Graw Hill.
- Sande, J. C. (s.f.). *Analisis de sentimiento en twitter*. Universidad Oberta de Catalunya, Barcelona.
- Sande, J. S. (2018). *Analisis de sentimientos en twitter*. Universidad Oberta de Catalunya, Barcelona.
- Twitter, D. (2022). *Developer platfor* . Obtenido de Developer platfor : <https://developer.twitter.com/en/products/twitter-api>
- Vallejo, H., Guevata, E., & Medina, S. (2018). Minería de datos. *Recimundo* , 343.
- Vega, R. R. (2020). Analisis de sentimientos en Twitter para descubrir contenido xenofobo hacia los inmigrantes Venezolanos en Ecuador. *Analisis de sentimientos en Twitter para descubrir contenido xenofobo hacia los inmigrantes Venezolanos en Ecuador*. Universidad Nacional de Loja, Loja.