



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

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

**ÁREA
TECNOLOGÍA DE LOS ORDENADORES**

**TEMA
ANÁLISIS DE REPUTACIÓN EN LÍNEA DE
PRECANDIDATOS ELECTORALES EN TWITTER
MEDIANTE TÉCNICA DE MINERÍA DE DATOS.**

**AUTORA
SEGURA TORRES MIXI JOSELYNE**

**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 INGENIERÍA EN TELEINFORMÁTICA**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA			
FICHA DE REGISTRO DE TRABAJO DE TITULACIÓN			
TÍTULO Y SUBTÍTULO:	Análisis de reputación en línea de precandidatos electorales en Twitter mediante técnica de minería de datos.		
AUTOR(ES) (apellidos/nombres):	Segura Torres Mixi Joselyne		
TUTOR(ES)/ REVISOR(ES) (apellidos/nombres):	Ing. Comp. Vargas Plaza Angel Marcel, Mg / Ing. Sist. Pincay Bohórquez Freddy Steve, Mg.		
INSTITUCIÓN:	Universidad de Guayaquil		
UNIDAD/FACULTAD:	Facultad De Ingeniería Industrial.		
MAESTRÍA/ESPECIALIDAD:			
GRADO OBTENIDO:	Ingeniera en Teleinformática.		
FECHA DE PUBLICACIÓN:	28 de septiembre del 2022	No. DE PÁGINAS:	112
ÁREAS TEMÁTICAS:	Tecnología de los ordenadores		
PALABRAS CLAVES/ KEYWORDS:	Twitter, Api de Twitter, Reputación en línea, Minería de datos, Precandidatos, Python, Jupyter Notebook, Dashboard.		
<p>RESUMEN:</p> <p>Debido al confinamiento por la pandemia del covid-19, se puso en tendencia el uso masivo de las redes sociales para tener conocimiento e información inmediata. En este tiempo de preparación de los partidos políticos en vista del próximo proceso electoral, se desarrolla una actividad abundante de comunicaciones por medio de las redes sociales Twitter, la cual nos permitirá evaluar la percepción de la reputación en línea de los precandidatos y así identificar los niveles de aceptación de los diferentes nombres propuestos. Mediante este proceso se aplicará una técnica de minería de datos con el respectivo análisis de la reputación en línea de los precandidatos electorales en Twitter mediante técnica de minería de datos, utilizando la metodología que consta de 5 fases: Selección, Análisis y extracción de datos, Limpieza y validación de datos, Minería de datos e Interpretación, para la visualización de los resultados obtenidos se trabajará con un Dashboard dinámico y de fácil comprensión, donde se podrá analizar la aceptación del perfil de cada precandidato a la alcaldía de Guayaquil.</p> <p>ABSTRACT:</p> <p>Due to the confinement that caused the covid-19 pandemic, the massive use of social networks for immediate knowledge and information has become a trend. In this time of</p>			

preparation of the political parties in view of the next electoral process, an abundant activity of communications is developed through social networks Twitter, which will allow us to evaluate the perception of the online reputation of the pre-candidates and thus identify the levels of acceptance of the different proposed names.

Through this process a data mining technique will be applied with the respective analysis of the online reputation of the electoral pre-candidates on Twitter by means of a data mining technique, using the methodology that consists of 5 phases: Selection, Data analysis and extraction, Data cleaning and validation, Data mining and Interpretation.

For the visualization of the results obtained, we will work with a dynamic and easy to understand Dashboard, where it will be possible to analyze the acceptance of the profile of each pre-candidate for the mayor of Guayaquil.

ADJUNTO PDF:	SI X	NO
CONTACTO CON AUTOR/ES:	Teléfono: 0988327102	E-mail: mixi.segurat@ug.edu.ec
CONTACTO CON LA INSTITUCIÓN:	Nombre: Ing. Ramón Maquilón Nicola	
	Teléfono: 593-2658128	
	E-mail: direccionTi@ug.edu.ec	

**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, **SEGURA TORRES MIXI JOSELYNE** con C.C. No. 0952913457, certifico que los contenidos desarrollados en este trabajo de titulación, cuyo título es “**ANÁLISIS DE REPUTACIÓN EN LÍNEA DE PRECANDIDATOS ELECTORALES EN TWITTER MEDIANTE TÉCNICA DE MINERÍA DE DATOS.**” 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.

Mixi Segura Torres

SEGURA TORRES MIXI JOSELYNE

C.C.No. 0952913457



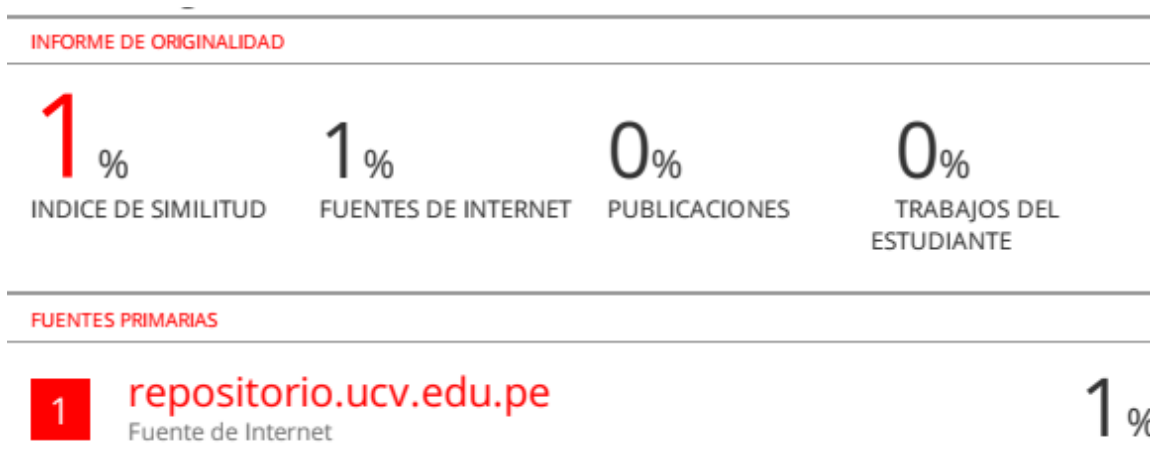
**ANEXO VII.- CERTIFICADO PORCENTAJE
DE SIMILITUD
FACULTAD DE INGENIERÍA INDUSTRIAL
CARRERA INGENIERÍA EN TELEINFORMÁTICA**



Habiendo sido nombrado **ING. COMPL. PLAZA VARGAS ANGEL MARCEL, MG.**, tutor del trabajo de titulación certifico que el presente trabajo de titulación ha sido elaborado por **SEGURA TORRES MIXI JOSELYNE**, con mi respectiva supervisión como requerimiento parcial para la obtención del título de **INGENIERO EN TELEINFORMÁTICA**.

Se informa que el trabajo de titulación: **ANÁLISIS DE REPUTACIÓN EN LÍNEA DE PRECANDIDATOS ELECTORALES EN TWITTER MEDIANTE TÉCNICA DE MINERÍA DE DATOS**, ha sido orientado durante todo el periodo de ejecución en el promedio Antiplagio **TURNITIN** quedando el 1% de coincidencia.

<https://ev.turnitin.com/app/carta/es/?s=1&lang=es&o=1897463079&u=1133714257>



Firmado electrónicamente por:

**ANGEL MARCEL
PLAZA VARGAS**

ING. COMPL. PLAZA VARGAS ANGEL MARCEL, MG
DOCENTE TUTOR
C.C. 0915953665
FECHA: 11 DE SEPTIEMBRE DEL 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:
**“ANÁLISIS DE REPUTACIÓN EN LÍNEA DE PRECANDIDATOS
ELECTORALES EN TWITTER MEDIANTE TÉCNICA DE MINERÍA DE
DATOS”** del estudiante **SEGURA TORRES MIXI JOSELYNE**, 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. COMPL. PLAZA VARGAS ANGEL MARCEL, MG
DOCENTE TUTOR**

C.C. 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 **“ANÁLISIS DE REPUTACIÓN EN LÍNEA DE PRECANDIDATOS ELECTORALES EN TWITTER MEDIANTE TÉCNICA DE MINERÍA DE DATOS”** del estudiante **SEGURA TORRES MIXI JOSELYNE**. 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 16 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
PINCAY
BOHORQUEZ**

ING. SIST. PINCAY BOHÓRQUEZ FREDDY STEVE, MG

C.C.: 0919786285

FECHA: 22 DE SEPTIEMBRE DE 2022

Agradecimiento

Agradezco en primer lugar a Dios, por haberme permitido culminar este nivel académico, mis padres por guiarme a que siga adelante con mis estudios, a mi novio por ayudarme en este proceso tan importante a nivel profesional, mis profesores por brindarme sus conocimientos en cada una de las materias impartidas, especialmente a mi tutor que me ayudo con sus aportes profesionales, con paciencia y constancia en este proyecto de tesis.

Índice General del Contenido

N°	Descripción	Pág.
	Introducción	2

Capítulo I

El problema

N°	Descripción	Pág.
1	El problema	3
1.1.	Planteamiento del problema.	3
1.2.	Formulación del problema	4
1.3.	Sistematización del problema	4
1.4.	Justificación	5
1.5.	Objetivos	6
1.5.1.	Objetivo general	6
1.5.2.	Objetivos específicos	6
1.6.	Hipótesis	6
1.7.	Variables	6
1.7.1.	Variable	6
1.7.2.	Variable independiente	6
1.8.	Conceptualización y operacionalización de las variables	7
1.9.	Indicadores	7
1.10.	Alcance	8

Capítulo II

El problema

N°	Descripción	Pág.
2	Marco teórico	
53		
2.1.	Antecedentes de estudio	
53		
2.1.1.	Antecedentes internacionales	10
2.1.2.	Antecedentes nacionales	11
2.2.	Marco teórico	14
2.2.1.	Reputación en línea	14
2.2.2.	Ciudadano digital	14
2.2.3.	Publicidad digital	15
2.2.4.	Redes sociales	16

2.2.4.1.	Twitter	16
2.2.4.2.	Facebook	17
2.2.4.3.	Instagram	18
2.2.5.	Api	18
2.2.5.1.	Api de twitter	19
2.2.5.2.	Api de facebook	20
2.2.5.3.	Api de instagram	21
2.2.6.	Comportamiento online	22
2.2.7	Minería de datos	22
2.2.7.1.	Técnicas para el minado de datos	23
2.2.7.2.	Análisis de minería de datos	25
2.2.7.3.	Tipos de datos que pueden ser minados	26
2.2.8.	Minería de datos en redes sociales	27
2.2.8.1.	Función de la minería de datos sociales	29
2.2.8.2.	Recopilación de la minería de datos en redes sociales	30
2.2.9.	Metodologías de la minería de datos	31
2.2.9.1.	Metodología KDD	31
2.2.9.2.	Metodología SEMMA	33
2.2.9.3.	Metodología CRISP-DM	34
2.2.10.	Herramientas y lenguajes empleados en la minería de datos	36
2.3.	Marco conceptual	39
2.3.1.	Dashboard	39
2.3.2.	Funcionamiento del dashboard	40
2.3.3.	Ventajas de un dashboard	41
2.3.3.1.	Supervisar varios kpis y métricas a la vez	41
2.3.3.2.	Los dashboards son fáciles de leer	42
2.3.3.3.	Accesibilidad en la nube: comparte tu dashboard con quien quieras	43
2.3.3.4.	Los dashboards hacen que los informes sean más eficientes	43
2.3.4.	Metodología de un dashboard	44
2.3.5.	Power bi	44
2.3.5.1.	¿Por qué power bi?	45
2.3.5.2.	Ventajas power bi	45
2.3.5.3.	Tipos de licencias: free, pro y premium	46
2.3.6.	Developer twitter	47
2.3.6.1.	Políticas de uso	48
2.3.7.	Anaconda	49
2.3.8.	Python	49

2.4.	Marco legal	50
Capítulo III		
La Propuesta		
N°	Descripción	Pág.
3	La propuesta	53
3.1.	Metodología del proyecto	53
3.1.1.	Diseño de investigación	53
3.1.1.1.	Diseño cuasi experimental	53
3.1.2.	Tipo de investigación	53
3.1.2.1.	Investigación exploratoria	53
3.1.2.2.	Investigación descriptiva	53
3.1.3.	Técnica de investigación	54
3.1.3.1.	Muestreo	54
3.1.3.1.1.	No probabilístico	54
3.2.	Descripción del proyecto	54
3.2.1.	Factibilidad técnica	55
3.2.2.	Factibilidad legal	55
3.3	Esquema general del análisis de datos	56
3.4.	Recursos para el análisis de datos	56
3.4.1.	Anaconda	57
3.4.2.	Api twitter	57
3.4.3.	Tweets	57
3.4.4.	Entorno de desarrollo: jupyter notebook	57
3.4.5.	Lenguaje de programación: python	58
3.4.6.	Formato csv	60
3.5.	Procedimiento: Aplicación de la Metodología escogida	60
3.5.1.	Selección	60
3.5.2.	Análisis y extracción de datos	61
3.5.2.1.	Análisis de datos	62
3.5.3.	Limpieza y validación de datos	65
3.5.4.	Minería de datos	67
3.5.5.	Visualización de datos	72
3.6.	Presentación de resultados	74
3.7.	Conclusiones	79
3.8.	Recomendaciones	81
	Anexos	82
	<u>Bibliografía</u>	94

Índice de Tablas

N°	Descripción	Pág.
1.	Definiciones de las variables	7
2.	Indicadores de las variables	7
3.	Antecedentes Internacionales	10
4.	Antecedentes Nacionales	12
5.	Característica de Clementine de SPSS	36
6.	Característica de WEKA	37
7.	Característica de RAMSES	38
8.	Característica de SAS	39
9.	Softwares utilizados en el desarrollo del proyecto	55
10.	Definiciones de las librerías utilizadas	58

Índice de Figuras

N°	Descripción	Pág.
1.	Encuesta de Omar Saluk.	53
2.	Publicidad digital.	15
3.	Redes sociales.	16
4.	Twitter.	17
5.	Facebook.	17
6.	Instagram.	18
7.	Funcionamiento API.	19
8.	Twitter API.	20
9.	Facebook API.	21
10.	Instagram API.	21
11.	Comportamiento Online	22
12.	Permiso de la Minería de datos.	23
13.	Minería de datos	24
14.	Tipos de datos.	27
15.	Minería de datos en redes sociales	27
16.	Minería de datos en rede sociales	29
17.	Metodología KDD.	32
18.	Metodología SEMMA.	34
19.	Metodología CRISP- DM.	35
20.	Tipos de Dashboard.	40
21.	Dashboard.	41
22.	Fácil uso de dashboard.	42
23.	Metodologías para Dashboard.	44
24.	Componentes Power BI.	45
25.	Ventajas Power BI	46
26.	Tipos de Licencia de Power BI.	46
27.	Developer.	47
28.	Anaconda.	49
29.	Python.	50
30.	Metodología.	53

31.	Esquema general del análisis de datos	56
32.	Selección .	60
33.	Tweets.	61
34.	Código de programación	62
35.	Código de programación.	64
36.	Código de programación.	64
37.	Código de programación.	65
38.	Código de programación.	65
39.	Código de programación.	66
40.	Código de programación.	66
41.	Código de programación.	67
42.	Código de programación.	68
43.	Código de programación.	68
44.	Código de programación.	69
45.	Código de programación.	69
46.	Código de programación.	70
47.	Código de programación.	70
48.	Código de programación.	71
49.	Código de programación.	72
50.	Dashboard.	73
51.	Dashboard.	73
52.	Dashboard.	75
53.	Dashboard.	75
54.	Dashboard.	76
55.	Dashboard.	76
56.	Dashboard.	77
57.	Dashboard.	77
58.	Dashboard.	78
59.	Dashboard.	78



ANEXO XIII.- RESUMEN DE TRABAJO DE TITULACIÓN(ESPAÑOL)



FACULTAD DE INGENIERÍA INDUSTRIAL CARRERA INGENIERÍA EN TELEINFORMÁTICA

ANÁLISIS DE REPUTACIÓN EN LÍNEA DE PRECANDIDATOS ELECTORALES EN TWITTER MEDIANTE TÉCNICA DE MINERÍA DE DATOS.

Autor: Segura Torres Mixi Joselyne

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

Resumen

Debido al confinamiento por la pandemia del covid-19, se puso en tendencia el uso masivo de las redes sociales para tener conocimiento e información inmediata. En este tiempo de preparación de los partidos políticos en vista del próximo proceso electoral, se desarrolla una actividad abundante de comunicaciones por medio de las redes sociales Twitter, la cual nos permitirá evaluar la percepción de la reputación en línea de los precandidatos y así identificar los niveles de aceptación de los diferentes nombres propuestos.

Mediante este proceso se aplicará una técnica de minería de datos con el respectivo análisis de la reputación en línea de los precandidatos electorales en Twitter mediante técnica de minería de datos, utilizando la metodología que consta de 5 fases: Selección, Análisis y extracción de datos, Limpieza y validación de datos, Minería de datos e Interpretación.

Para la visualización de los resultados obtenidos se trabajará con un Dashboard dinámico y de fácil comprensión, donde se podrá analizar la aceptación del perfil de cada precandidato a la alcaldía de Guayaquil.

Palabras Claves: Twitter, Api de Twitter, Reputación en línea, Minería de datos, Precandidatos, Python, Jupyter Notebook, Dashboard.



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



FACULTAD DE INGENIERÍA INDUSTRIAL CARRERA INGENIERÍA EN TELEINFORMÁTICA

ONLINE REPUTATION ANALYSIS OF ELECTORAL PRECANDIDATES ON TWITTER USING A DATA MINING TECHNIQUE.

Autor: Segura Torres Mixi Joselyne

Advisor: Engr. Plaza Vargas Angel Marcel, Mg.

Abstract

Due to the confinement that caused the covid-19 pandemic, the massive use of social networks for immediate knowledge and information has become a trend. In this time of preparation of the political parties in view of the next electoral process, an abundant activity of communications is developed through social networks Twitter, which will allow us to evaluate the perception of the online reputation of the pre-candidates and thus identify the levels of acceptance of the different proposed names.

Through this process a data mining technique will be applied with the respective analysis of the online reputation of the electoral pre-candidates on Twitter by means of a data mining technique, using the methodology that consists of 5 phases: Selection, Data analysis and extraction, Data cleaning and validation, Data mining and Interpretation.

For the visualization of the results obtained, we will work with a dynamic and easy to understand Dashboard, where it will be possible to analyze the acceptance of the profile of each pre-candidate for the mayor of Guayaquil.

Keywords: Twitter, Twitter Api, Online Reputation, Data Mining, Candidates, Python, Jupyter Notebook, Dashboard.

Introducción

La red social Twitter fue creada por Jack Dorsey, es una herramienta que limita sus mensajes. Twitter es una plataforma digital para postear y recibir mensajes de texto. Las redes sociales son sitios en Internet donde las personas publican y comparten todo tipo de información, noticias ya sea personal o profesional, lo cierto es que es un espacio virtual para facilitar la interacción entre personas.

Por otro lado, la reputación en línea se trata de como el elector digital se expresa por medio de comentarios o publicaciones en un tema en específico, creando una estrategia mediante los tweets para generar un cierto impacto, tratando de mantener una interacción entre usuarios de la comunidad digital. Sin embargo, es importante tener en cuenta que la reputación de los precandidatos va a estar inmersas en un variado tipo de percepción que se tiene acerca de ellos visibilizando la atención que genere sus propuestas de campaña.

Mientras tanto la minería de datos estudia métodos y algoritmos que permiten la extracción automática, visualización y análisis de información recopilada; también facilita el análisis de los datos de forma ordenada y correcta. En la minería de datos se puede trabajar con diversas técnicas estadísticas y con Machine Learning.

Este proyecto tiene como objetivo general desarrollar un análisis de reputación en línea de precandidatos electorales en Twitter mediante técnica de minería de datos, para visualizar la información extraída mediante tweets de sus propuestas planteadas en campaña. Los comentarios sobre los precandidatos a la alcaldía de Guayaquil en Twitter en la actual situación van incrementando la inseguridad referente al voto mediante el análisis de los usuarios ya que esto trae como relación la crisis política en la cual está inmerso nuestro país.

Mediante este proceso se aplicará una técnica de minería de datos con el respectivo análisis de la reputación en línea de los precandidatos electorales en Twitter mediante técnica de minería de datos, utilizando la metodología que consta de 5 fases: Selección, Análisis y extracción de datos, Limpieza y validación de datos, Minería de datos e Interpretación. Para la visualización de los resultados obtenidos se trabajará con un Dashboard dinámico y de fácil comprensión, donde se podrá analizar la aceptación del perfil de cada precandidato a la alcaldía de Guayaquil.

Capítulo I, desarrollamos la investigación en relación con los siguientes temas: El planteamiento del problema, formulación del problema, sistematización del problema, justificación, objetivos, hipótesis, variables, conceptualización y operacionalización de las variables, indicadores y culminamos el capítulo con el alcance del proyecto.

Capítulo II, se procedió a realizar los antecedentes del estudio, marco teórico, marco conceptual y justificó en base al marco legal.

Capítulo III, culminando el proyecto se presentará la propuesta de investigación a lineada a la metodología, descripción del proyecto, esquema general del análisis de datos, recursos para el análisis de datos, procedimientos: aplicación de la metodología escogida y se culmina con la presentación de resultados, detallando sus respectivas conclusiones y recomendaciones.

Capítulo I

El problema

1.1. Planteamiento del problema.

Cada vez que se realizan elecciones es un problema apremiante la limitada información socializada por las redes, lo cual denota una insatisfacción popular en los diversos factores que involucran la baja calidad de la información enmarcada en el proceso, creemos necesario que esta información propicie argumentos válidos para los precandidatos, los cuales concentraran todas sus propuestas con el fin de sostener el buen manejo de sus campañas, disminuyendo la incertidumbre para el votante.

Debido al confinamiento por la pandemia del covid-19, se puso en tendencia el uso masivo de las redes sociales para tener conocimiento e información inmediata. En este tiempo de preparación de los partidos políticos en vista del próximo proceso electoral, se desarrolla una actividad abundante de comunicaciones por medio de la red social Twitter, la cual nos permitirá evaluar la percepción de la reputación en línea de los precandidatos y así identificar los niveles de aceptación de los diferentes nombres propuestos.

Actualmente, el elector digital está conectado e informado en función de las tendencias otorgadas por las redes sociales ya que mediante estas amplias características de los precandidatos podrán definir su mejores propuestas, demostrando su participación activa en la investigación para poder modificar su percepción hacia los candidatos, ya que tradicionalmente se basaba en el populismo, es así que los electores han incrementado su capacidad de análisis implementando nuevas exigencias para satisfacer las necesidades de la ciudad de Guayaquil.

El cambio del paradigma en las redes sociales ha implementado nuevas exigencias en función de las nuevas tecnologías que tiene como objetivo recaudar una mayor información en un menor tiempo, desarrollando un análisis autocrítico del ofrecimiento de los precandidatos permitiendo una mejor visión referentes a las propuestas ofrecidas en campaña.

La transformación de la investigación mediante la información política nos genera una mejor adaptación a los procesos cambiantes que impone una visión actualizada de los perfiles de los candidatos para no llegar a una comunicación populista que forcé a los cibernautas a tener una buena reputación o un rechazo a estas elite políticas mediante el mal uso de la información que pueden ser posteadas, ocultando las carencias de sus propuestas, emergiendo como reacción a una falsa percepción hacia los partidos tradicionales.

Realizando este análisis, conoceremos la reputación en cada uno de los precandidatos por parte de los electores digitales. Sino más bien llegar a tener un abanico de información en función a su capacidad de exigencia llegando hacer una autoevaluación para el beneficio de la ciudad, aplicando técnica de minería de datos, definiremos una serie de procesos para obtener el resultado del análisis de tendencias electorales en Twitter.

1.2. FORMULACIÓN DEL PROBLEMA

Teniendo como referencia la problemática planteada se deriva las siguientes interrogantes:

1. ¿Es posible medir la reputación en línea de los precandidatos a la alcaldía de Guayaquil mediante la Api de Twitter?
2. ¿Es posible definir características y reputación de los precandidatos electorales mediante el desenvolvimiento en línea de sus propuestas de campaña?

1.3. SISTEMATIZACIÓN DEL PROBLEMA

- ¿Cuáles son las herramientas optimas que se debería considerar para la recopilación de datos en la red social Twitter?
- ¿Cuáles serían los programas viables que se deberían utilizar para el análisis de la reputación en línea de los precandidatos electorales?

1.4. JUSTIFICACIÓN

Inmerso en el análisis de datos de una era moderna digital constantemente se crea muchas interrogantes con la finalidad de interpretar el comportamiento de los internautas digitales en las redes sociales. Proyectando un crecimiento acelerado de diferentes tipo de información a través del internet mediante una red de servidores que nos conectan con cualquier parte del mundo, creyendo conveniente la manera de comunicar en función a radioenlace desde un punto principal con otros nodos dispersos en el mundo, generando un modelo de distribución de información basado en la velocidad de interpretación de gran cantidad de datos, distribuyendo dicha información para poder evaluar y percibir la aceptación por parte de los consumidores digitales.

El buen uso de métodos, parámetros y herramientas nos permiten seleccionar y diferenciar grandes volúmenes de información mediante la extracción de datos en redes sociales como Twitter, las cuales marcaran una distancia en función de la percepción, sea esta mediante satisfacción o desagrado manifestadas en publicaciones o comentarios en las redes sociales Twitter, esperando ser aprovechada dicha información con los distintos segmentos de consumidores digitales con el fin de mejorar la decisión a tomar. Hoy en día el internet es el principal medio de acceso a la información, mediante el cual optimizamos el tiempo en función de la captación de información, mejorando procesos y el avance en función a las nuevas tecnologías en investigación.

Mediante este análisis denotaremos la información extraída que nos permitirá proyectar las diferentes características de los precandidatos mediante la interacción de los consumidores online, aplicando la minería de datos la cual nos ayudará con la percepción de la reputación en línea de cada uno de ellos, a partir de palabras repetitivas mostrando el avance del proceso de la búsqueda de datos como parte primordial.

Esperando que los resultados obtenidos tengan como base la información de los precandidatos y deslumbre a los electores de una manera más acertada en la toma de decisiones, sobre el fracaso o éxito en una elección que cada vez se torna más exigentes y analíticos en la evaluación de la información los cuales esperan una preferencia más acertada en función de un objetivo específico.

1.5. OBJETIVOS

1.5.1. OBJETIVO GENERAL

Desarrollar un análisis de reputación en línea de precandidatos electorales en Twitter mediante técnica de minería de datos

1.5.2. OBJETIVOS ESPECÍFICOS

1. Determinar contenido teórico con respecto a la reputación en línea en Twitter.
2. Definir la metodología para recopilación de datos en redes sociales.
3. Recopilar información mediante el análisis de minería de datos.

1.6. HIPÓTESIS

Mediante el uso de técnicas de minería de datos en Twitter es posible realizar un análisis de reputación en línea de precandidatos electorales.

1.7. VARIABLES

1.7.1. VARIABLE DEPENDIENTE

- Análisis de reputación en línea de precandidatos electorales

1.7.2. VARIABLE INDEPENDIENTE

- Uso de técnica de minería de datos en Twitter

1.8. CONCEPTUALIZACIÓN Y OPERACIONALIZACIÓN DE LAS VARIABLES

Tabla 1. Definiciones de las variables

	DEPENDIENTE	INDEPENDIENTE
VARIABLES	Reputación en línea	Uso de técnica de minería de datos
DEFINICIONES CONCEPTUALES	Se basa en los comentarios que realizan los consumidores digitales de una persona, basándose en las publicaciones en redes sociales en las cuales se crea una imagen sobre quienes son o como nos ven los demás	Se trata de la extracción de datos, en los cuales realizamos estadísticas por medio de la inteligencia artificial en la que se basa en algoritmos con el objetivo de que la información sea la correcta.

Información obtenida de la investigación elaborado por Mixi Segura

1.9. INDICADORES

Tabla 2. Cuadro comparativo de variables mediante indicadores

VARIABLES	INDICADORES	DESCRIPCIÓN	TIPO DE INDICADOR
REPUTACIÓN EN LÍNEA	Usuario	Persona que utiliza redes sociales para navegar, publicar e informarse.	Cuantitativo
	Post	Publicaciones que se hace mediante fotos, videos y textos.	Cualitativo

	Activo	Tiempo que una persona pasa en redes sociales.	Cuantitativo
	Plataformas digitales	Monitoreo de las redes sociales	Cualitativo
USO DE TÉCNICA DE MINERÍA DE DATOS	Datos	Información obtenida para un análisis.	Cualitativo
	Algoritmos	Procesos en los cuales podemos resolver un problema.	Cualitativo

Información obtenida de la investigación elaborado por Mixi Segura

1.10. ALCANCE

- Mediante la red social Twitter se abracará datos recopilados en un periodo corto de tiempo, utilizando la Api de Twitter, con la técnica de minería de datos.
- La investigación se basará en las características, percepción y reputación de los precandidatos, fundamentados en los comentarios y publicaciones dentro de los espacios de interacciones.
- La información procesada permitirá establecer una catalogación de los precandidatos a la alcaldía de Guayaquil.
- Por medio de un Dashboard dinámico y de fácil comprensión, se podrá visualizar la aceptación del perfil de cada precandidato a la alcaldía de Guayaquil.

Capítulo II

Marco Teórico

2.1. Antecedentes de Estudio

Se ha comprobado que la utilización de las redes sociales colabora de gran manera en la velocidad mediante la cual se distribuye la información, nos lleva a consultar diferentes referencias bibliográficas para impulsar el incremento y agilización de nuevas tendencias en productividad de datos, convencido de esto la información arrojará resultados muy parecidos, pero con enfoques diferentes aplicados al desenvolvimiento del elector digital, con esta convicción se utilizara una API de Twitter que permita minar datos centralizados disponibles de esta red social, empleando metodologías que colaborara al desarrollo de la técnica de minería de datos.

Posteriormente, se proyectará la búsqueda científica de investigaciones realizadas que realzará el conocimiento del proyecto del cual se está investigando, con la finalidad de obtener mejoras en nuestra metodología y herramientas a utilizar para el beneficio del tema propuesto.

Basado en la encuesta del Ing. Omar Maluk Salem, el cual es uno de los que tienen autorización por parte del CNE en brindar este tipo de información, seleccionamos a los cuatros precandidatos para este proyecto de tesis, los cuales podemos observar en la ilustración 1.

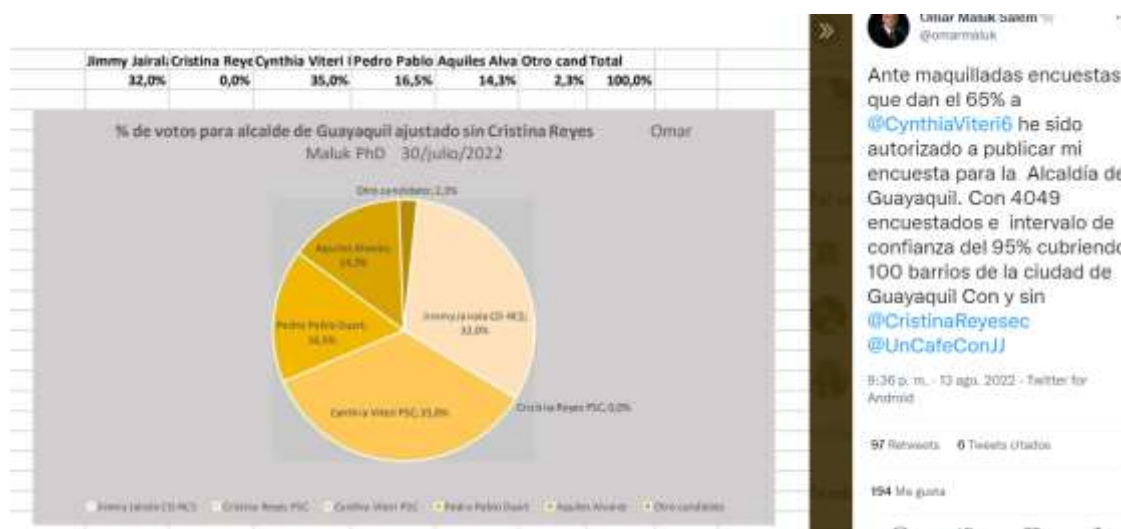


Ilustración 1. Encuesta. Información tomada de la cuenta de Twitter de Omar Maluk. Elaborado por Mixi Segura.

2.1.1. ANTECEDENTES INTERNACIONALES

(Roales, 2019) realizó una tesis que tiene como objetivo el uso de técnicas de minería de datos para analizar y utilizar la información disponible en Twitter procedente de esta fuente de datos.

Según (Herran, 2021) su tesis hace referencia a la Investigación y análisis del discurso de odio sobre la población flotante en las redes sociales desde la perspectiva del trabajo social.

En el artículo realizado por (Olarte, Panizzi, & Bertone, 2018) implica utilizar la minería de datos para descubrir patrones en los datos de las redes sociales.

En el presente estudio de (Lobillo & Gallart, 2020) la investigación reveló un análisis del contenido de las comunicaciones de la Universidad Rey Juan Carlos durante la crisis provocada por el denominado "caso principal".

Tabla 3. Antecedentes Internacionales

REFERENCIA	TEMA	METODOLOGÍA	TÉCNICA	HERRAMIENTA
(Roales, 2019)	Detección de tendencias en Twitter utilizando minería de datos adaptativa	KDD “Descubrimiento de conocimiento en bases de datos”	Clasificación, clusterización, asociación y/o análisis de tendencias	Apache Lucene, una API de código abierto originalmente implementada en Java
(Herran, 2021)	Análisis de los discursos de odio debido a la población migrante en las redes.	Metodología aplicada por el Autor. Investigación cualitativa y cuantitativa	Análisis documental, Big Data.	Facebook, extracción de datos.

(Olarte, Panizzi, & Bertone, 2018)	Segmentación de Mercado Usando Técnicas de Minería de Datos en Redes Sociales	Aplicación de minería de datos.	de Clasificación, de clusterización, análisis de tendencias.	Python, redes sociales, Api de Twitter
(Lobillo & Gallart, 2020)	Análisis de la crisis reputacional universitaria mediante de las redes sociales Twitter y Facebook.	Metodología aplicada por el Autor. Investigación cualitativa y cuantitativa	Clasificación y descripción	Facebook, Twitter.

Información obtenida de la investigación elaborado por Mixi Segura

2.1.2. ANTECEDENTES NACIONALES

Según (Ordoñez, 2018) el propósito del modelo de minería de datos que utiliza el enfoque CRISM-DM, que es el mismo que uno de los métodos de generación de proyectos de minería de datos más utilizados en la actualidad, es para el modelo de análisis de datos.

En el trabajo de investigación de (Villacis, 2021) explora la construcción de la reputación del expresidente Lenin Moreno a partir de las percepciones públicas en la plataforma de microblogging Twitter entre 2019 y 2021, en el panorama político general, incluyendo temas como la comunicación política digital, los momentos de crisis y el uso de las redes sociales.

En la tesis encontrada de (Cañarte, 2021) hace referencia al análisis de la reputación online del presidente chileno Sebastián Piñera a partir de sus opiniones publicadas en la plataforma de microblogging Twitter entre 2019 y 2021.

(Yorly, 2020) nos comenta en su tesis de grado que el uso de las telecomunicaciones y redes sociales de los consumidores ecuatorianos, así como los cambios en los modelos de negocio de las empresas por la crisis sanitaria, obligan a las empresas a desarrollar estrategias de marketing digital, dar a conocer o mejorar los productos y servicios ofrecidos para satisfacer las necesidades.

Tabla 4. Antecedentes Nacionales

REFERENCIA	TEMA	METODOLOGÍA	TÉCNICA	HERRAMIENTA
(Ortega, 2018)	Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes de primer ciclo de la modalidad abierta y a distancia de la UTPL	CRISP-DM (Cross Industry Standard Process for Data Mining)	Clasificación, clusterización, asociación y/o análisis de tendencias	WEKA (Waikato Environment for Knowledge Analysis)
(Villacis, 2021)	Reputación online de Lenin Moreno en la plataforma microblogging Twitter durante el periodo 2019 - 2021	Metodología aplicada por el autor, cualitativa.	Extracción de datos.	Microblogging Twitter, Api de Twitter.
(Cañarte, 2021)	Construcción de la reputación online de	Metodología aplicada por el autor,	Extracción de contenido digital	Microblogging Twitter, Api de Twitter.

	Sebastián piñera durante el periodo 2019-2021 en la plataforma de microblogging Twitter	cualitativa.			
--	---	--------------	--	--	--

(Yorly, 2020)	Análisis del comportamiento del consumidor digital en redes sociales mediante técnica de minería de texto	Identificación, adquisición, filtrado, extracción, validación, limpieza de datos, análisis y visualización de datos.	Minería de datos mediante Twitter.	Jupyter Notebook con Python 3, Api de twitter	
--------------------------	--	--	--	--	--

Información obtenida de la investigación elaborado por Mixi Segura

2.2. MARCO TEÓRICO

2.2.1. REPUTACIÓN EN LÍNEA

En internet creamos una imagen sobre quienes somos y como nos ven los demás. Esta imagen o personalidad se conforma a partir de la información que publicamos en foros, blogs, redes sociales, videos o en los enlaces que compartimos. Además, otras personas pueden contribuir a la contribución de nuestra imagen a través de los comentarios que generan en nuestras paginas o por medio de las imágenes que ellos publican. Una imagen positiva permite a las personas obtener grandes oportunidades en su desarrollo personal y profesional, es una actividad que se lleva a cabo en nuestra vida cotidiana, pero que debe extenderse a las actividades en línea.

“El uso de las tecnologías de información y comunicación requiere responsabilidad y respeto en cada una de las actividades que llevemos a cabo. Esto involucra mantenerse informado sobre las distintas tecnologías, sus ventajas y posibles riesgos para que, como usuarios migrantes o nativos digitales, podemos tomar decisiones basadas en el conocimiento obtenido del uso de la tecnología, construyendo una nueva ciudadanía digital que permita a las personas interactuar en ambientes seguros y confiables “. (Ortega, 2018)

2.2.2. CIUDADANO DIGITAL

El ciudadano digital ha encontrado en la sociedad de la información una infinidad de posibilidades que la informática y las autopistas de la información aportan a su vida, aunque con ellas no logre completamente satisfacer los sentimientos de incertidumbre que lo rondan. (Irene, 2021)

La COVID-19 llegó a poner fin a las dudas sobre cómo se vería la educación y el trabajo en un entorno en línea. Aunque muchas escuelas y trabajos buscaban maneras de digitalizarse o ya lo estaban haciendo, nadie estaba preparado para la cuarentena y los cambios que trajo la pandemia. Sin embargo, esta es una situación que sólo beneficia a los verdaderos ciudadanos digitales, aquellos que no sólo tienen acceso a internet, sino también a las herramientas y el conocimiento necesarios para continuar con sus trabajos, su educación y sus vidas a pesar de todo. Ojalá esta época sirva de aprendizaje sobre la urgencia de garantizar el acceso a internet como un derecho básico universal y de darle

a los ciudadanos las herramientas necesarias para convertirse en verdaderos ciudadanos digitales. El libro “Digital Citizenship. The Internet, Society, and Participation” define la ciudadanía digital con una simple frase, “la capacidad de participar en la sociedad en línea”. (Delgado, 2020)

(Alcázar, 2017) expresa que el comportamiento del ecuatoriano en medios digitales continúa estando centrado en la búsqueda, consumo de video, noticias, transacciones y redes sociales, evolucionando día a día hacia la generación de contenido y consumo de alternativas locales que complementan la oferta internacional, siendo cada vez más adepto a la generación de transacciones en comercio electrónico y servicios. Una sociedad que evoluciona en cuanto al comportamiento y tipo de consumo digital que día a día investiga, prueba y utiliza de forma más madura los recursos que presentan medios, plataformas y formatos digitales centrados en la generación de valor para usuarios.

2.2.3. PUBLICIDAD DIGITAL

Las campañas electorales que realizan los diferentes partidos políticos, en el marco de los procesos democráticos de elecciones que se llevan a cabo en la actualidad ofrecen la oportunidad, de que se pueden utilizar tecnologías de información y comunicación y de esta manera dar a conocer y publicitar electoralmente a sus candidatos y sus ideologías políticas y de esa manera, los electores pueden elegir al mejor candidato, no sin antes tener muy en claro los pormenores de su elección; es aquí, que las tecnologías de información y comunicación tendrán un protagonismo importante en la transmisión de los mensajes publicitarios. (Guerrero, 2018)



Ilustración 2. Publicidad digital. Información tomada de la página web. Elaborado por Mixi Segura

2.2.4. REDES SOCIALES

“Las redes sociales son lugares en Internet donde las personas publican y comparten todo tipo de información, personal y profesional, con terceras personas, conocidos y absolutos desconocidos”, afirma Celaya (2018). Por su parte, Wikipedia la define como: “una estructura social que se puede representar en forma de uno o varios grafos donde nodos representan individuos y las aristas las relaciones entre ellos”.

(Alcázar, 2017) comenta que Facebook mantiene el liderazgo absoluto con 11 millones de usuarios registrados en Ecuador, de los cuales, un promedio de 6,4 millones es activo mensualmente; esto quiere decir que acceden a la plataforma, seguido por Instagram, LinkedIn y Twitter. Todas estas redes, utilizadas principalmente desde dispositivos móviles.



Ilustración 3. Redes sociales. Información tomada de la web. Elaborado por Mixi Segura

2.2.4.1. TWITTER

Twitter es una red social gratuita que te permite escribir mensajes de forma rápida y sencilla. Cualquier persona puede crear una cuenta o perfil para compartir sus opiniones, gustos e intereses, pero también, para estar informada sobre todo lo que acontece en el mundo, ya que la principal característica de Twitter es que allí encuentras información en tiempo real. Una vez creas tu cuenta podrás seguir a otros usuarios para ver lo que publican y así mismo, otras personas podrán seguirte, ver tus tuits y comentarte. Esa es la forma como se interactúa en esta plataforma. Los mensajes que circulan en Twitter son llamados tuits o trinos que, en su mayoría, son fragmentos de texto que no superan los 280 caracteres. Una vez creas tu cuenta podrás seguir a otros

usuarios para ver lo que publican y así mismo, otras personas podrán seguirte, ver tus tuits y comentarte. Esa es la forma como se interactúa en esta plataforma. (GCF, 2022)



Ilustración 4. Twitter. Información tomada de la web. Elaborado por Mixi Segura

2.2.4.2. FACEBOOK

Es la red social que más usuarios tienen en el mundo. Una red de vínculos virtuales, cuyo principal objetivo es dar un soporte para producir y compartir contenidos. Llegó para ampliar las posibilidades de relación social y causó una revolución sensible en el mundo de las comunicaciones, son cerca de 2,2 millones de usuarios y en él puedes encontrar personas conocidas, participar en grupos de interés, compartir contenido, enviar y recibir mensajes, hacer contactos, realizar búsquedas, anunciar, etc.

Entre las redes sociales, es la que los usuarios escogen más seguido. Facebook es, actualmente, el sitio web más accedido del mundo. Gracias a su éxito y gran adhesión, es considerada como la red que influencia directamente en la política, en la cultura y en la opinión pública de los usuarios. (Gonzalez, 2021)



Ilustración 5. Facebook. Información tomada de la web. Elaborado por Mixi Segura

2.2.4.3. INSTAGRAM

(Blas, 2018) comenta que Instagram es una red social y una aplicación móvil al mismo tiempo, que permite a sus usuarios subir imágenes y vídeos con múltiples efectos fotográficos como filtros, marcos, colores retro, etc., para posteriormente compartir esas imágenes en la misma plataforma o en otras redes sociales. Esta aplicación o red, es actualmente una de las que mayor crecimiento tiene en el mundo, cuando con más de 2.000 millones de usuarios activos.



Ilustración 6. Instagram. Información tomada de la web. Elaborado por Mixi Segura

2.2.5. API

Una API o Interfaz de Programación de Aplicaciones son una serie de protocolos que permiten integrar el software de diferentes aplicaciones. Por tanto, la información puede comunicarse o “traspasarse” entre una plataforma y otra. En un comercio conversacional, las API permiten a las empresas tener más opciones en las conversaciones con los clientes. Podrás convertir una conversación en un ticket de soporte, integrarse en tu CRM, etc. (Centribal, 2022)

En un nivel superior, las API son la forma en que los programas informáticos "hablan" entre sí para solicitarse y enviarse información. Esto se hace permitiendo que un software llame a lo que se conoce como punto de conexión: una dirección que se corresponde con un tipo específico de información que proporcionamos los puntos de conexión son generalmente únicos, como los números de teléfono. (Centribal, 2022)

¿Cómo funciona una API?



Ilustración 7. Funcionamiento API. Información tomada de la web. Elaborado por Mixi Segura

2.2.5.1. API DE TWITTER

En la página oficial de (Twitter, 2022) se comunica lo qué está pasando en el mundo y sobre lo que las personas están hablando en este momento. Puedes acceder a Twitter en la Web o desde tu dispositivo móvil. Para compartir información en Twitter de la forma más amplia posible, también les proporcionamos a las empresas, los desarrolladores y los usuarios acceso programático a los datos de Twitter mediante nuestras API (interfaces de programación de aplicaciones).

Twitter permite acceder a partes de nuestro servicio mediante las API para permitirles a las personas crear software que se integre con Twitter, como una solución que ayuda a una empresa a responder a la opinión del cliente en Twitter.

Los datos de Twitter son únicos y se extraen a partir de datos de la mayoría de las otras plataformas sociales porque reflejan información que los usuarios deciden compartir de forma pública. Nuestra plataforma de API ofrece acceso amplio a los datos de Twitter que los usuarios han decidido compartir con el mundo. Twitter también es compatible con API que permiten a los usuarios administrar su propia información de Twitter que no es pública (como los Mensajes directos) y brindarla a los desarrolladores que ellos han autorizado a administrarla. (Twitter, 2022)



Ilustración 8. Twitter API. Información tomada de la web. Elaborado por Mixi Segura

2.2.5.2.API DE FACEBOOK

La API de Facebook no es solo una aplicación, sino un conjunto de interfaces de programación de aplicaciones, una cosa que distingue a Facebook, aparte de su competencia en el campo de las redes sociales es su amplio uso de la construcción de la plataforma del sitio, características y aplicaciones con software de código abierto. Tomando la filosofía de código abierto para el corazón, Facebook comparte un conjunto de interfaces de programación de aplicaciones que los desarrolladores pueden usar para aprovechar la red social y extender las aplicaciones y sitios web del desarrollador. Las principales interfaces de programación de aplicaciones que ofrece Facebook incluyen la Ads API, Credits API, Chat API y Graph API. (Mitchell, 2022)

La API de Facebook más usada es la Graph API, que es una característica fundamental que traza el gráfico social para todas las conexiones entre las personas, páginas, eventos, mensajes, fotos, comentarios y "me gusta". Tocar el gráfico social Facebook le permite a tus aplicaciones Web interactuar con los datos en múltiples formas. Puedes recibir los mensajes más recientes de un usuario, acceder y descargar un álbum de fotos y descubrir el más popular "me gusta" en toda la red social. El Graph API incluye una función de búsqueda que te permite recuperar fácilmente y analizar estos datos, siempre y cuando los datos del usuario sean públicos y no privados. Si es privado, entonces tu aplicación necesitará un acceso de Facebook simbólico para obtener el permiso del usuario. (Mitchell, 2022)

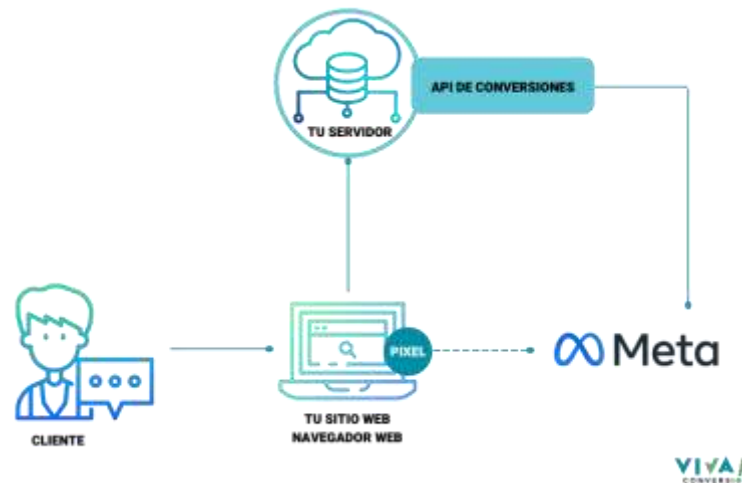


Ilustración 9. Facebook API. Información tomada de la web. Elaborado por Mixi Segura

2.2.5.3. API DE INSTAGRAM

La API Graph de Instagram permite a los profesionales de Instagram utilizar tu aplicación para administrar su presencia en Instagram. La API se puede utilizar para obtener y publicar contenido multimedia, administrar los comentarios y responder a ellos en dicho contenido, identificar el contenido multimedia con @menciones realizadas por otros usuarios de Instagram, buscar contenido multimedia con hashtags y obtener los resultados y metadatos básicos de otras empresas y creadores de Instagram. (Meta, 2022)

Está diseñada para los creadores y empresas en Instagram que necesitan insights y control completo de todas sus interacciones en los medios sociales. Si estás creando una aplicación para consumidores o solo necesitas obtener información básica del perfil de un usuario, fotos y vídeos, puedes utilizar la API de visualización básica de Instagram. (Meta, 2022)



Ilustración 10. Instagram API. Información tomada de la web. Elaborado por Mixi Segura.

2.2.6. COMPORTAMIENTO ONLINE

El comportamiento online del usuario son los actos o acciones que son manifestados por los consumidores dentro de una web o App que se pueden observar a través de herramientas de medición o análisis de comportamiento. Es un factor de relevancia para evaluar por qué un usuario actúa de una determinada manera, anticipar comportamientos y detectar posibles brechas en el proceso de compra. Debido a la cantidad de datos que se pueden obtener de observar la navegación de un usuario y su comportamiento online, es aconsejable haber definido muy bien los objetivos antes de comenzar el trabajo de investigación y determinar qué nos interesa medir realmente. Un buen análisis del comportamiento de usuario debe comenzar cuando éste empieza a buscar información sobre el producto y mantenerse después de haber concretado la venta. (arimetrics, 2022)



Ilustración 11. Comportamiento Online. Información tomada de la web. Elaborado por Mixi Segura

2.2.7 MINERÍA DE DATOS

Es un proceso de hurgar en los datos para descubrir conexiones ocultas y predecir tendencias futuras tiene una larga historia. Conocido algunas veces como "descubrimiento de conocimientos en bases de datos", el término "minería de datos" no se acuñó sino hasta la década de 1990. Pero su base comprende tres disciplinas científicas entrelazadas: estadística “el estudio numérico de relaciones de datos, inteligencia artificial inteligencia similar a la humana exhibida por software y/o máquinas y machine learning “algoritmos que pueden aprender de datos para hacer predicciones. Lo que era antiguo es nuevo otra vez, ya que la minería de datos continúa

evolucionando para igualar el ritmo del potencial sin límites del big data y poder de cómputo asequible. (SAS, 2022)



Ilustración 12. Permiso de la Minería de datos. Información tomada del artículo. Elaborado por Mixi Segura

Aprenda más acerca de técnicas de minería de datos en Minería de datos de la A hasta la Z, documento que muestra cómo pueden las organizaciones utilizar la analítica predictiva y la minería de datos para revelar nuevos insights partiendo de datos. La minería de datos, como una disciplina compuesta, representa diversos métodos o técnicas que se utilizan en diferentes capacidades analíticas que abordan una gama de necesidades organizacionales, hacen diferentes tipos de preguntas y utilizan diferentes niveles de aportación humana o reglas para llegar a una decisión. (SAS, 2022)

2.2.7.1. TÉCNICAS PARA EL MINADO DE DATOS

Las técnicas que necesitas conocer para llevar a cabo un minado de datos:

- **Asociación:** Se trata de una de las técnicas más utilizadas. En esta técnica, una transacción y la relación entre los elementos se utilizan para identificar un patrón. Esta es la razón por la que también se conoce como «técnica de relación». Se utiliza para realizar un análisis de la cesta de la compra, que se hace para conocer todos aquellos productos que los clientes compran juntos habitualmente, por ejemplo.

2.2.7.2. ANÁLISIS DE MINERÍA DE DATOS

(IEBS, 2022) los mineros o exploradores de datos a la hora de llevar a cabo un análisis de minería de datos deberán realizar los siguientes pasos:

- **Investigación comercial**

Antes de empezar, deberás tener una idea completa de los objetivos de tu empresa, de los recursos disponibles y de los diversos escenarios actuales en consonancia con los requisitos. Esto sería muy útil de cara a crear un plan detallado que alcance los objetivos de la organización.

- **Análisis de calidad**

A medida que vamos recopilando los datos de distintas fuentes, necesitaremos verificarlos y compararlos para garantizar que no haya cuellos de botella en el proceso de integración de datos. La garantía de calidad ayuda a detectar cualquier anomalía en los datos, como la interpolación de datos faltantes, manteniendo los datos en plena forma antes de que se sometan a una extracción.

- **Limpieza de datos**

Se trata de la selección, limpieza, enriquecimiento, reducción y transformación de la base de datos. Se calcula que el 90% del tiempo en este tipo de procesos se gasta en este paso.

- **Transformación de datos**

Este paso consta de cinco subetapas. Los procesos involucrados hacen que los datos estén listos en conjuntos de datos finales.

- **Suavizado de datos:** Se elimina el ruido de los datos
- **Resumen de datos:** La agregación de conjuntos de datos se aplica en este proceso
- **Generalización de datos:** Los datos se generalizan reemplazando cualquier dato de bajo nivel con conceptualizaciones de nivel superior
- **Normalización de datos:** Los datos se definen en rangos establecidos
- **Construcción de atributos de datos:** Los conjuntos de datos deben estar en el conjunto de atributos antes de la minería de datos

- **Modelado de datos**

Por último, para una mejor identificación de los patrones de datos, se implementan varios modelos matemáticos en el conjunto de datos, basados en varias condiciones. Actualmente este tipo de trabajos se están realizando en seguridad de datos, finanzas, salud, marketing, detección de fraude, búsquedas online, procesamiento del lenguaje natural, coches inteligentes, entre otros. Es por este motivo, que la minería de datos se está convirtiendo en uno de los trabajos con mayor proyección para el futuro. (IEBS, 2022)

2.2.7.3. TIPOS DE DATOS QUE PUEDEN SER MINADOS

Como nos podemos imaginar, no todos los modelos de datos pueden ser minados. A continuación, te indicamos cuáles sí:

- **Datos almacenados en una base de datos**

Una base de datos también puede denominarse sistema de gestión de bases de datos o DBMS. Cada DBMS almacena datos que están relacionados entre sí de una forma u otra. También tiene un conjunto de programas de software que se utilizan para administrar datos y proporcionar un fácil acceso a ellos. Estos programas de software sirven para muchas cosas, incluida la definición de la estructura de la base de datos o asegurarse de que la información almacenada permanezca segura y consistente.

- **Data Warehouse**

Un almacén de datos o data Warehouse es una única ubicación de almacenamiento de datos que recopila datos de varias fuentes y luego los almacena en forma de plan unificado. Cuando los datos se almacenan en estos sistemas se someten a una limpieza, integración, carga y actualización.

- **Data transaccional**

La base de datos transaccional almacena registros que se capturan como transacciones. Por ejemplo, reservas de vuelos, compras, clics en un sitio web, etc. Cada registro de transacciones tiene una identidad única. También engloba todos los elementos que la han convertido en una transacción.

Otros tipos de datos

Por último, también existen muchos otros tipos de datos que son conocidos por su estructura, significados semánticos y versatilidad. (IEBS, 2022)



Ilustración 14. Tipos de datos. Información tomada del artículo. Elaborado por Mixi Segura

2.2.8. MINERÍA DE DATOS EN REDES SOCIALES

La extracción de datos de redes sociales se refiere al proceso de extracción de datos sociales. A diferencia de la minería de datos regular, la minería de datos de redes sociales explora más que solo las bases de datos internas y los sistemas de una determinada empresa o firma de investigación.

Por lo general, implica recopilar, procesar y analizar datos sin procesar obtenidos de plataformas de redes sociales como Facebook, Instagram, Twitter, TikTok, LinkedIn, YouTube, etc. para descubrir patrones y tendencias significativas, sacar conclusiones y proporcionar insights procesables. (ATRIBUS, 2022)



Ilustración 15. Minería de datos en redes sociales. Información tomada del artículo. Elaborado por Mixi Segura

Estos son algunos ejemplos de quién y cómo se utiliza la minería de datos de redes sociales:

Algunos de sus principales usos en las empresas incluyen campañas de marketing dirigidas, investigación de mercado, habilitación de ventas, análisis predictivo, marketing de influencer y monitoreo de reputación de marca.

- **Análisis de tendencias:** las empresas utilizan la minería de datos de redes sociales para obtener información valiosa sobre las palabras clave, las menciones y los temas de tendencias actuales en las plataformas de redes sociales.
- **Detección de eventos (mapa de calor social):** esta métrica es importante para las instituciones e investigadores que utilizan el monitoreo de redes sociales.
- **Detección de spam social:** la minería de datos de redes sociales facilita la detección de spammers y bots en plataformas de redes sociales como Instagram y Twitter.
- **Comercio electrónico:** la extracción de datos de las redes sociales se utiliza para analizar cómo habla la gente sobre los productos.
- **Medios digitales:** la extracción de datos de medios sociales también se utiliza en el espacio de medios digitales. Por ejemplo, el contenido que se mostrará en una cartelera digital en particular se puede determinar realizando un proceso de extracción de datos de redes sociales para satisfacer las preferencias o necesidades de la audiencia.
- **Blogueros e influencer:** los blogueros y las personas influyentes en las redes sociales a menudo utilizan la extracción de datos de las redes sociales para ayudarlos a analizar las actitudes y los sentimientos de sus seguidores, de qué están hablando y cómo se sienten acerca de ciertos temas de discusión.
- **Marcas:** la extracción de datos de redes sociales ayuda a las marcas a tomar decisiones importantes, por ejemplo, al decidir sobre posibles mercados futuros.
- **Propósito del estudio:** los investigadores han descubierto que el uso de datos de redes sociales en su investigación es un activo valioso para su trabajo debido al gran tamaño y la facilidad de acceso a los datos. La minería de datos de redes sociales se puede aplicar en diferentes campos de investigación, incluidas las ciencias sociales, la investigación, la investigación en salud y la investigación tecnológica. Algunos de sus usos en el campo de la investigación incluyen la recopilación de opiniones, la realización de investigaciones, el reclutamiento de

participantes en la investigación, la realización de «ciencia ciudadana» participativa o la facilitación de la participación de las partes interesadas.

- **Agencias gubernamentales:** las agencias gubernamentales también utilizan cada vez más la extracción de datos de las redes sociales para intervenciones centradas en el bienestar. Un enfoque para la minería de datos de redes sociales es rastrear el movimiento de los residentes a lo largo del día mientras registran sus actividades en lugares marcados. Claramente, la minería de redes sociales puede ser una herramienta poderosa que puede ayudar a mejorar la vida de los residentes y la seguridad de las comunidades. (ATRIBUS, 2022)

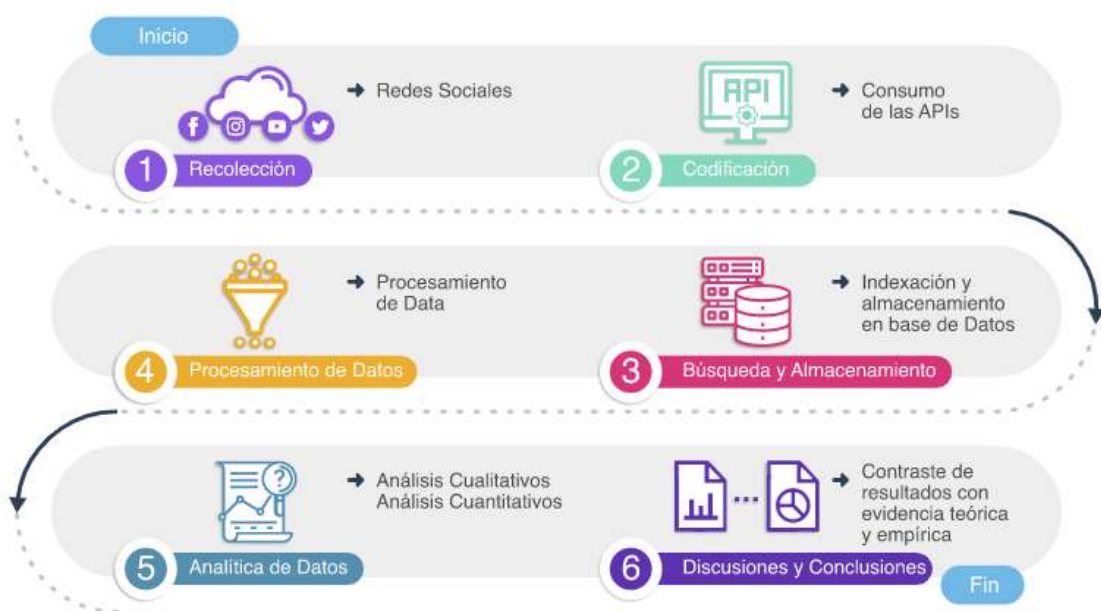


Ilustración 16. Minería de datos en red social. Información tomada del artículo. Elaborado por Mixi Segura

2.2.8.1. FUNCIÓN DE LA MINERÍA DE DATOS SOCIALES

(Atribus, 2022) nos comenta que a menudo, el proceso de extracción de datos sociales implica una combinación de técnicas estadísticas, matemáticas y aprendizaje automático.

El primer paso es recopilar y procesar datos sociales de diferentes fuentes de redes sociales. Además de las plataformas de redes sociales como Facebook, Twitter o YouTube, los mineros de datos también extraen datos de blogs, sitios de noticias, foros

o cualquier otra página pública donde los usuarios interactúan y publican comentarios. Toda esta información debe procesarse antes de continuar con el siguiente paso.

Después de recopilar y procesar los datos, lo siguiente es la aplicación de varias técnicas de minería de datos que facilitan la identificación de patrones y correlaciones comunes de varios puntos de datos en grandes conjuntos de datos. Algunas de las técnicas de minería de datos de redes sociales más utilizadas incluyen clasificación, asociación, patrones de seguimiento, análisis predictivo, extracción de palabras clave, análisis de sentimiento y análisis de mercado/tendencia.

Además, la minería de datos de redes sociales emplea muchas soluciones de software de minería de datos de redes sociales para optimizar el proceso de minería. Algunas de las soluciones de software de minería de datos más conocidas incluyen: Microsoft SharePoint, Sísense, IBM Cognos, RapidMiner y Dundas BI. Los mineros de datos también pueden decidir utilizar el aprendizaje automático en el proceso si se requiere un examen más profundo de los datos.

El paso final en el proceso de minería es crear una representación visual de los conocimientos adquiridos a lo largo del proceso para entregar la información al público objetivo. (Atribus, 2022)

2.2.8.2. RECOPIACIÓN DE LA MINERÍA DE DATOS EN REDES SOCIALES

Los datos representan las actitudes, conexiones, comportamientos y sentimientos de las personas sobre un tema, producto o servicio. Dependiendo de la plataforma de redes sociales en cuestión, si los datos de las redes sociales de destino provienen de retweets o impresiones de Facebook, Twitter o tasas de participación y uso de hashtags de Instagram, estos datos pueden incluir la cantidad de seguidores, comentarios, me gusta o acciones. (ATRIBUS, 2022)

Cuando intentes optimizar tus contenidos en redes sociales, promocionar tu negocio online, descubrir clientes influyentes o mejorar tus estrategias de marketing, siempre es importante concentrarse en recopilar los tipos de datos anteriores.

2.2.9. METODOLOGÍAS DE LA MINERÍA DE DATOS

La minería de datos forma parte del Big Data, ya que trata de gestionar y analizar un conjunto enorme de datos. Se puede aplicar a numerosos ámbitos de estudio y sectores profesionales. En función del objetivo que persigan, los algoritmos del Data Mining pueden realizar tres tipos de analítica: descriptivos, predictivos o prescriptivos.

- **Analítica descriptiva.** Los algoritmos descriptivos buscan nexos con la finalidad de describir datos. Realizan la toma de una imagen de una realidad concreta para conocer con mayor profundidad la situación de una empresa, un mercado o un sector, entre otros.
- **Analítica predictiva.** Este tipo de algoritmos utilizan, además de la foto de la situación, datos reales y modelos de comportamiento. Estos permiten predecir cómo evolucionará y cuáles serán las próximas tendencias.
- **Analítica prescriptiva.** La tecnología empleada en este caso utiliza datos históricos y predicciones extraídas de la analítica predictiva para hacer recomendaciones sobre la toma de decisiones. Estas decisiones se toman de forma automatizada, ya que están basadas en cálculos fiables que combinan la optimización matemática con los sistemas de gestión empresarial. Esto permite reducir los costes e incrementar las ventas. (Veigler, 2022) Las tres metodologías dominantes para el proceso de la minería de datos son: KDD, CRISP-DM y SEMMA.

2.2.9.1. METODOLOGÍA KDD

Es una metodología propuesta por Fayyad en 1996, propone 5 fases: Selección, preprocesamiento, transformación, minería de datos y evaluación e implantación. Es un proceso iterativo e interactivo. (Leon, 2021)

1. Desarrollar un entendimiento de la aplicación de dominio y los conocimientos previos y la identificación de la meta del proceso de KDD desde el punto de vista del cliente.
2. Crear un conjunto de datos objetivo: la selección de un conjunto de datos, o que se centren en un subconjunto de variables o datos de muestras, el descubrimiento que se llevará a cabo.

3. Limpieza y preprocesamiento de datos. Operaciones básicas incluyen la eliminación de ruido campos de datos vacíos, etc.
4. reducción de datos y la proyección: la búsqueda de características útiles para representar los datos en función del objetivo de la tarea. (Reducción de la dimensionalidad)
5. Es colocar el objetivo del KDD (paso 1) a un método de minería de datos.
6. Es el análisis exploratorio, hipótesis y el modelo de selección: la elección del algoritmo de minería de datos que se utilizará para la búsqueda de patrones de datos.
7. Séptimo es la minería de datos: la búsqueda de patrones de interés en una determinada forma de representación o de un conjunto de tales representaciones.
8. Interpretación de los patrones minadas, posiblemente se puede regresar a cualquiera de los pasos 1 a 7 para más iteración. Este paso puede implicar también la visualización de los patrones y modelos extraídos o visualización de los datos que figuran extraídos modelos.
9. Está actuando sobre el conocimiento descubierto: el uso del conocimiento directamente, incorporando el conocimiento en otro sistema para la adopción de nuevas medidas o, simplemente, documentación y presentación de informes a las partes interesadas. Este proceso también incluye la comprobación y la solución de posibles conflictos con creían (o extrae) los conocimientos.

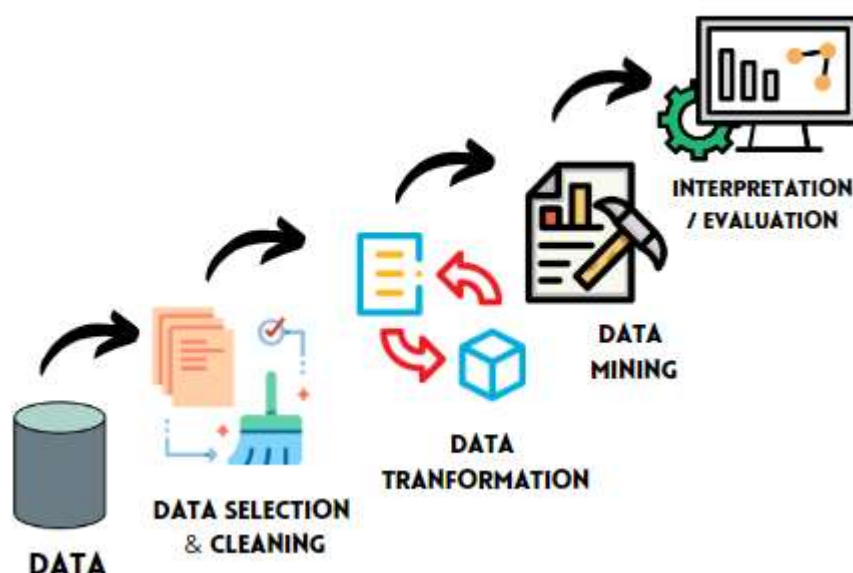


Ilustración 17. Metodología KDD. Información tomada del artículo. Elaborado por Mixi Segura.

2.2.9.2. METODOLOGÍA SEMMA

Esta metodología se trata de un proceso de selección, exploración y modelamiento de grandes cantidades de datos para descubrir patrones de negocios desconocidos. (Leon, 2021)

- **Fase muestral:** Este proceso se inicia con la extracción de la población muestral sobre la que se va a aplicar el análisis. El objetivo de esta fase consiste en seleccionar una muestra representativa del problema en estudio. La representatividad de la muestra es indispensable ya que de no cumplirse invalida todo el modelo y los resultados dejan de ser admisibles. La forma más común de obtener una muestra es la selección al azar, es decir, cada uno de los individuos de una población tiene la misma posibilidad de ser elegido. Este método de muestreo se denomina muestreo aleatorio simple.
- **Fase Exploración:** Se procede a la exploración de la información disponible con el fin de simplificar en lo posible el problema para optimizar la eficiencia del modelo. Para lograr este objetivo, se propone la utilización de herramientas de visualización o de técnicas estadísticas que ayuden a poner de manifiesto relaciones entre variables. De esta forma, se pretende determinar cuáles son las variables explicativas que van a servir como entradas al modelo.
- **Fase manipulación:** La tercera fase de la metodología consiste en la manipulación de los datos, en base a la exploración realizada, de forma que se definan y tengan el formato adecuado los datos que serán introducidos en el modelo.
- **Fase Modelado:** El objetivo de esta fase consiste en establecer una relación entre las variables explicativas y las variables objeto del estudio, que posibiliten inferir el valor de estas con un nivel de confianza determinado. Las técnicas utilizadas para el modelado de los datos incluyen métodos estadísticos tradicionales (tales como análisis discriminante, métodos de agrupamiento, y análisis de regresión), así como técnicas basadas en datos tales como redes

neuronales, técnicas adaptativas, lógica fuzzy (difusa), árboles de decisión, reglas de asociación y computación evolutiva.

- **Fase Valoración:** Esta fase del proceso consiste en la valoración de los resultados mediante el análisis de bondad del modelo o modelos contrastados con otros métodos estadísticos o con nuevas poblaciones muestrales. (Leon, 2021)

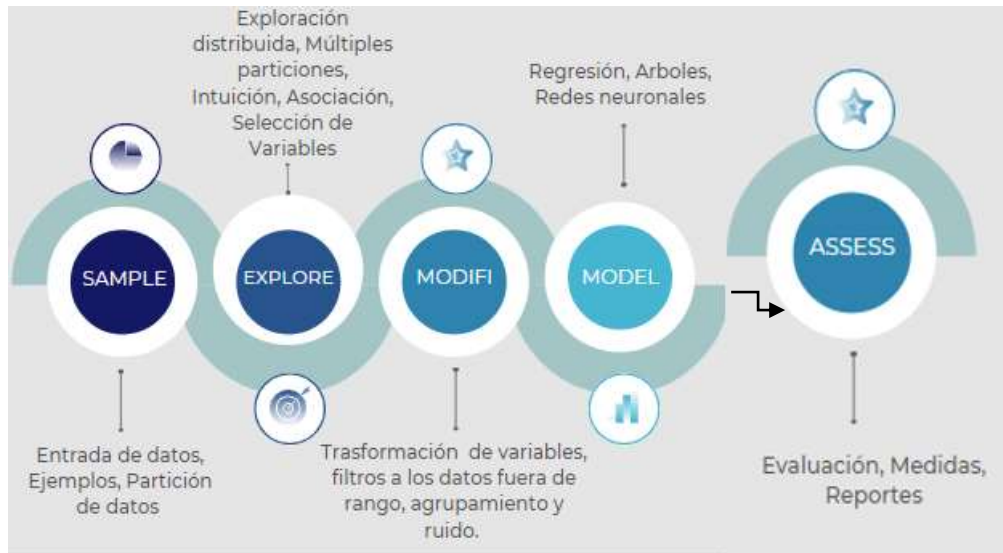


Ilustración 18. Metodología SEMMA. Información tomada del artículo. Elaborado por Mixi Segura

2.2.9.3. METODOLOGÍA CRISP-DM

La metodología de referencia CRISP-DM contiene las fases de un proyecto, sus respectivas tareas y sus relaciones entre ellas. El ciclo de vida de un proyecto de Minería de Datos consiste esencialmente en seis fases. La secuencia de las fases no es rígida, se puede regresar o adelantar a alguna de ellas siempre que se necesario. Todo depende de los resultados de cada fase. (IBM, 2021)

- **Análisis del problema:** Primera fase, se centra en el análisis de los objetivos y requisitos desde una perspectiva comercial, posiblemente también académica, seguido este análisis se transcribe como un problema de Minería de Datos y se diseña un plan preliminar para el logro de los objetivos.
- **Análisis de los datos:** Esta fase comienza con la obtención de los datos para familiarizarse con ellos, identificar problemas de la calidad de estos, para

descubrir relaciones entre ellos. O detectar subconjuntos interesantes de datos para formular hipótesis de información posiblemente desconocida.

- **Preparación de los datos:** Esta fase cubre todas las actividades para construir el conjunto de datos final en el cual se aplicarán los modelos. Entre las tareas se incluye la selección de cuadros de datos, registros y atributo, así como su transformación, en caso sea necesario, para aplicar los modelos.
- **Modelado:** En esta fase varias técnicas de modelado son seleccionadas y aplicadas y sus parámetros son calibrados en valores óptimos. Generalmente existen diversas técnicas aplicables al mismo tipo de problema. Algunas técnicas necesitan que los datos estén preparados de cierta forma para poderse aplicar. Si los datos no están en condiciones es necesario regresar a la fase preparación de los datos
- **Evaluación:** Una vez creado un buen modelo se debe evaluar el rendimiento de este y la integridad de todos los pasos, es importante verificar que se han incluido todos los objetivos del negocio o investigación.
- **Desarrollo:** Frecuentemente los proyectos de Minería de Datos no terminan con la implementación del modelo sino en la obtención de conocimiento a partir de los datos. Para ello es importante documentar y presentar los resultados de manera comprensible. Se debe confirmar el mantenimiento de la aplicación y la difusión de los resultados, en caso de ser posible. (IEBS, 2022)

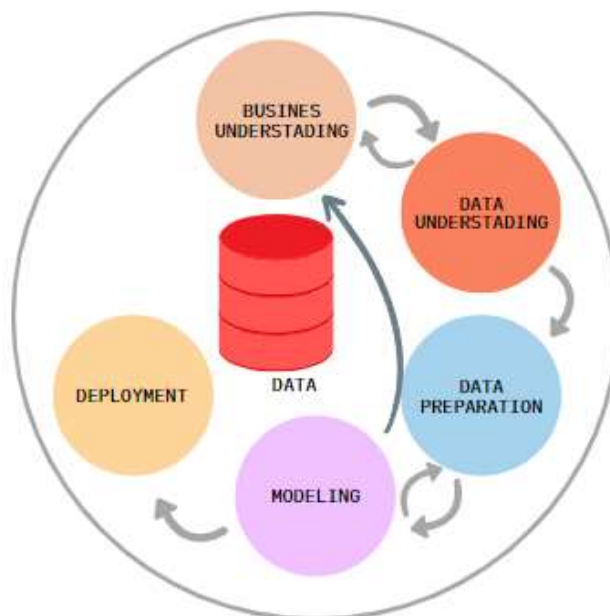



Ilustración 19. Metodología CRISP- DM. Información tomada del artículo. Elaborado por Mixi Segura.

2.2.10. HERRAMIENTAS Y LENGUAJES EMPLEADOS EN LA MINERÍA DE DATOS

En el artículo de (Rodriguez & Diaz, 2019) nos explican que algunas herramientas diseñadas para extraer conocimientos desde bases de datos que contienen grandes cantidades de información. Las más populares de estas herramientas son SPSS, Oracle Data Miner y Weka.

Clementine de SPSS: se centra en la integración de data mining con otros procesos y sistemas de negocio que ayuden a entregar inteligencia predictiva en un tiempo eficiente durante las operaciones de negocio diarias. La funcionalidad abierta en bases de datos que permite que muchos de los procesos se realicen en entornos que mejoran tanto el rendimiento como el despliegue de los resultados de data mining. La última versión extiende la funcionalidad al incluir un conjunto de reglas de scoring y modelos de árboles de decisión y carga de resultados de data mining en la base de datos. Sistema integrado de minería de datos que permite encontrar patrones en la información para facilitar la toma de decisiones a los usuarios. Utilizando Clementine se podrá:

Tabla 5. Característica de Clementine de SPSS

	<ul style="list-style-type: none"> • Acceder, preparar e integrar fácilmente datos numéricos, de texto, datos provenientes de páginas web y de encuestas. • Construir y validar modelos rápidamente, utilizando las técnicas estadísticas y de aprendizaje automáticos disponibles más avanzadas. • Implantar eficientemente los modelos predictivos, en tiempo real o según una programación establecida. • Obtener rápidamente un mejor retorno de la inversión y mejores tiempos de respuesta aprovechando las características de rendimiento y escalabilidad. • Transmitir de forma segura los datos confidenciales a las aplicaciones de data mining en los casos donde la seguridad es crítica.
---	--

Esta herramienta permite seleccionar campos o filtrar los datos, permite mostrar propiedades de los datos, encontrar relaciones, ambiente integrado de minería de datos para usuarios finales y desarrolladores.

YALE: Es una herramienta creada en la universidad de Dortmund bastante flexible para el descubrimiento del conocimiento y la minería de datos. Puesto que YALE está escrito enteramente en Java, funciona en las plataformas o sistemas operativos más conocidos. Es un software de código abierto GNU y con licencia GPL. Recientemente fue lanzada la última versión, la cual incluye características como las de implicar nuevos formatos de entrada de datos con operadores para Microsoft Excel y SPSS. Desde la perspectiva de la visualización YALE ofrece representaciones de datos en dispersión en 2D y 3D; representaciones de datos en formato SOM; coordenadas paralelas y grandes posibilidades de transformar las visualizaciones de los datos.

WEKA: Es de libre distribución y destacada por la cantidad de algoritmos que presenta, así como por la eficiencia de estos, ella proporciona gran cantidad de herramientas para la realización de tareas propias de minería de datos, la visualización y permite la programación en JAVA de algoritmos más sofisticados para análisis de datos y modelado predicativo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. En ella se implementan las técnicas de clasificación, asociación, agrupamiento, y predicciones existentes en la actualidad. Su sistema operativo es multiplataforma. Los puntos fuertes de Weka son:

Tabla 6. Característica de WEKA



- Está disponible libremente bajo la licencia publica general de GNU
- Es muy portable porque completamente implementado en Java y puede correr en casi cualquier plataforma.
- Contiene una extrema colección de técnicas para reprocesamiento de datos y modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario.

Weka soporta varias tareas estándar de minería de datos, especialmente, reprocesamiento de datos, clustering, clasificación, regresión, visualización, y selección. Todas las técnicas de Weka se fundamentan en la asunción de que los datos están disponibles en un fichero plano o una relación, en la que cada registro de datos está descrito por un número fijo de atributos. Weka también proporciona acceso a bases de datos vía SQL gracias a la conexión JDBC y puede procesar el resultado devuelto por una consulta hecha a la base de datos. No puede realizar minería de datos multi-relacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una única tabla que ya puede ser procesada con Weka.

RAMSES: Sistema de Gestión de Selección y Evaluación de Análisis de Riesgo es un programa de gestión de riesgos integrado en el sistema de proceso de datos de Bureau Veritas. El programa recopila todos los datos correspondientes a las operaciones de comercio internacional y está interconectado con la aplicación de minería de datos. Este software es aceptado como uno de los líderes del mercado en minería de datos y cumple las recomendaciones de la Convención de Kyoto de la OMA de 1999 y del Marco de Normas de la OMC. Es utilizado por organismos gubernamentales en el mundo entero. RAMSES ofrece a las autoridades gubernamentales una forma de identificar los embarques de mayor riesgo, facilitando por otro lado la circulación y el despacho de las mercancías de menor riesgo. Interconectado con las bases de datos de Bureau Veritas, RAMSES proporciona una gestión automatizada y digna de confianza de los riesgos inherentes al comercio internacional. Beneficios:

Tabla 7. Característica de RAMSES



- Analizar todos los datos del programa de inspección de importaciones
- Evaluar los niveles de riesgo de las diferentes expediciones de mercancías
- Favorecer los controles mejor orientados.
- Indicar las medidas a tomar para agilizar el despacho aduanero
- Se puede aplicar a diferentes tipos de bases de datos.
- Optimizar la asignación de recursos humanos.

SAS Enterprise Miner: Su compañía es SAS, es una solución de minería de datos que permite incorporar patrones inteligentes a los procesos de marketing, tanto operativos como estratégicos. El software de SAS es un sistema de entrega de información que provee acceso transparente a cualquier fuente de datos, incluyendo archivos planos, archivos jerárquicos, y los más importantes manejadores de bases de datos relacionales. También incluye su propia base de datos de información para almacenar y manejar los datos, es decir, un "data warehouse". También soporta los principales protocolos de comunicación, cubre los cinco modelos de procesamiento cliente/ servidor de acuerdo con Gartner Group y cumple con las 12 reglas de OLAP.

Tabla 8. Característica de SAS



El sistema soporta un amplio rango de aplicaciones, destacándose el análisis estadístico, análisis gráfico de datos, análisis de datos guiado, mejoramiento de la calidad, diseño experimental, administración de proyectos, programación lineal y no lineal, generación de reportes y gráficas, manipulación y despliegue de imágenes, sistemas de información geográfica, visualización multidimensional de datos, aplicaciones de multimedia, así como los sistemas de información ejecutiva.

Información obtenida de la investigación elaborado por Mixi Segura

2.3. MARCO CONCEPTUAL

2.3.1. DASHBOARD

Es una representación gráfica de los principales indicadores (KPI) que intervienen en la consecución de los objetivos de negocio, y que está orientada a la toma de decisiones para optimizar la estrategia de la empresa. (Tristán, 2021)

(Kyocera, 2021) nos comenta en su página que para tener un dashboard bien confeccionado podremos disponer de una gran herramienta que nos ayudará con el análisis estratégico. Un dashboard puede indicarnos que tenemos un problema o que todo o algo marcha muy bien. Nos presentará un hecho, que nos llevará a un origen

global, no particular, que nos pondrá en alerta para, posteriormente, poder llegar al origen último y tomar medidas correctoras, si fuera el caso, o potenciarlo, si el hecho que se nos ha presentado fuera positivo.

Por tanto, el dashboard nos indica un hecho, nos lleva a un origen general y nos muestra unas consecuencias. Con todo ello, deberemos poder dar una primera recomendación rápida y, a posterior, entrar en un análisis del detalle, con información complementaria, para encontrar el error concreto y poner solución. Como vemos, el disponer de un buen dashboard nos servirá para saber qué está pasando, reconocer rápidamente que es lo que necesita nuestra atención, tanto si es un problema como una oportunidad y, por tanto, no podemos obviar su verdadera función y significado estratégico para la empresa o cualquiera de sus departamento o áreas.



Ilustración 20. Tipos de Dashboard. Información tomada del artículo. Elaborado por Mixi Segura

2.3.2. FUNCIONAMIENTO DEL DASHBOARD

No todos los dashboards sirven para el mismo propósito, por lo que es importante que los usuarios entiendan qué KPIs deben seguir y por qué. Los dashboards responden a preguntas importantes sobre tu negocio. A diferencia de las herramientas avanzadas de inteligencia empresarial, los dashboards están diseñados para el análisis rápido y el conocimiento de la información. El enfoque más común para diseñar un dashboard es construirlo utilizando un formato de pregunta-respuesta. (TuDashboard, 2022)

2.3.3. VENTAJAS DE UN DASHBOARD

Son muchas las ventajas del uso de dashboards, entre ellas que nos permiten utilizar los datos brutos de diversas fuentes como hojas de cálculo y bases de datos para crear tablas, gráficos de líneas, gráficos de barras y medidores en un tablero central que los usuarios pueden consultar y comprender inmediatamente.

Al conocer qué es un dashboard sabrás que simplifican la elaboración de informes a final de mes, ya que permiten a los usuarios comunicar la información en cualquier momento sin necesidad de pasar horas preparando y analizando. (TuDashboard, 2022)

¿Para qué sirve un dashboard? Estos son algunos otros beneficios de contar con un dashboard dentro de tu organización:



Ilustración 21. Dashboard. Información tomada de la web. Elaborado por Mixi Segura

2.3.3.1. SUPERVISAR VARIOS KPIS Y MÉTRICAS A LA VEZ

Los cambios en cualquier aspecto de un negocio, ya sea en marketing, ventas, soporte o finanzas, tienen un impacto en el negocio en su conjunto. La gente ha estado monitoreando sus negocios sin un tablero de control de datos durante años, los tableros de control de datos hacen que sea mucho más fácil.

Con un dashboard los usuarios son capaces de profundizar en el panorama general para correlacionar este impacto con los KPIs y métricas específicas para

entender lo que funciona y lo que no. Tanto si los datos de tu empresa se almacenan en un servicio web, un archivo adjunto o una API, un dashboard extrae esta información y te permite supervisar todos tus datos en una ubicación central.

Otra de las ventajas de un dashboard es que son capaces de correlacionar datos de diferentes fuentes en una única visualización si el usuario así lo desea. Al supervisar múltiples KPIs y métricas en un tablero central, los usuarios pueden realizar ajustes en sus prácticas empresariales en tiempo real. (TuDashboard, 2022)

2.3.3.2. LOS DASHBOARDS SON FÁCILES DE LEER

En un Dashboard puedes utilizar colores, símbolos y visualizaciones para resaltar los puntos de datos importantes. Esto permite a los usuarios escanear rápidamente el Dashboard y obtener la información que necesitas sin tener que ir a buscar en hojas de cálculo, correos electrónicos o iniciar sesión en un servicio web.

Los Dashboard son útiles porque visualizan la información de forma accesible para todos. Incluso si no trabajas en marketing, por ejemplo, puedes entender sus números. Esto es clave: no es necesario ser un analista para utilizar y entender un tablero de mando. (TuDashboard, 2022)



Ilustración 22. Fácil uso de dashboard. Información tomada de la web. Elaborado por Mixi Segura

2.3.3.3. ACCESIBILIDAD EN LA NUBE: COMPARTE TU DASHBOARD CON QUIEN QUIERAS

Los dashboards llevan los datos de una empresa a la nube, haciendo que las métricas clave y los KPIs sean accesibles todo el tiempo a través de equipos de escritorio, el móvil y la tableta. Hay varias formas de compartir un dashboard: en televisores de pared, informes por correo electrónico, informes impresos o acceso directo.

Los usuarios pueden crear un enlace público a su dashboard que dé acceso a cualquiera, o a través de un enlace privado que sólo permita acceder a quienes tengan el enlace. Cada vez es más popular mostrar los dashboards en los televisores de las oficinas como forma de mantener a todo el mundo consciente del rendimiento y los objetivos. Los dashboards son cada vez más habituales porque permiten crear entornos de trabajo virtuales y facilitan la colaboración de los equipos. (TuDashboard, 2022)

2.3.3.4. LOS DASHBOARDS HACEN QUE LOS INFORMES SEAN MÁS EFICIENTES

Los dashboards ahorran tiempo. Los usuarios ya no tienen que acudir a múltiples fuentes desconectadas para hacer un seguimiento de sus datos. Obtener los datos, crear una hoja de cálculo, generar y diseñar el informe y compartirlo: los dashboards hacen todo esto automáticamente.

Hay varias razones por las que la elaboración de un informe de KPIs suele hacerse a final de mes, una de ellas es que requiere tiempo y recursos. Lo único que hay que hacer es invertir un poco de tiempo en la configuración del dashboard, lo que es más fácil que tener que hacer un informe manual. Los tableros de datos pueden generar automáticamente informes con sus datos, en cualquier momento y en cualquier lugar. Ya no es necesario que los usuarios reúnan, analicen y formateen los datos. Puedes crear informes en PDF, por correo electrónico y en directo mediante un cuadro de mando: elige el KPI que desees analizar, selecciona el formato del informe y preséntalo. Aquí algunos ejemplos de KPIs que puedes adoptar para monitorear en tu propio dashboard.

Además, los dashboards permiten una comunicación constante entre todos los niveles de una empresa. Pueden acceder tantos usuarios como la empresa decida, de modo que la información necesaria está al alcance de todos. Ya no es necesario que los ejecutivos y gerentes soliciten datos a los analistas. Los analistas pueden trabajar en segundo plano para garantizar un funcionamiento fluido, mientras que los usuarios acceden a esta información cuando lo desean. (TuDashboard, 2022)

2.3.4. METODOLOGÍA DE UN DASHBOARD

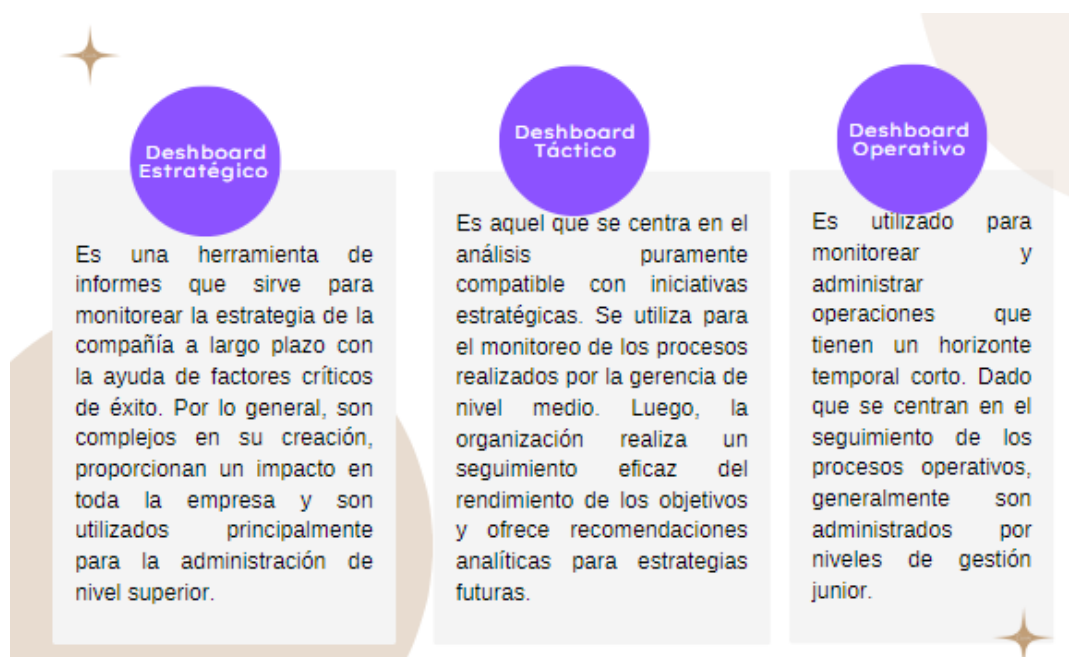


Ilustración 23. Metodologías para Dashboard. Información tomada del artículo. Elaborado por Mixi Segura

2.3.5. POWER BI

Power BI es un conjunto de herramientas que pone el conocimiento al alcance de todos y nos brinda acceder a nuestros datos de forma segura y rápida, generando grandes beneficios para nosotros y para nuestra empresa. Es un sistema predictivo, inteligente y de gran apoyo, capaz de traducir los datos (simples o complejos) en gráficas, paneles o informes por sus cualidades como la capacidad gráfica de presentación de la información, o la integración de Power Query: el motor de extracción, transformación y carga (ETL) incluido en Excel. (Deloitte, 2022)

2.3.5.1. ¿POR QUÉ POWER BI?

Power BI es una solución de análisis empresarial basado en la nube, que permite unir diferentes fuentes de datos, analizarlos y presentar un análisis de estos a través de informes y paneles. Con Power BI se tiene de manera fácil acceso a datos dentro y fuera de la organización casi en cualquier dispositivo. Estos análisis pueden ser compartidos por diferentes usuarios de la misma organización; por lo que directivos, financieros, comerciales, etc., pueden disponer de la información del negocio en tiempo real. (Deloitte, 2022)

Se conforma fundamentalmente de estos componentes:

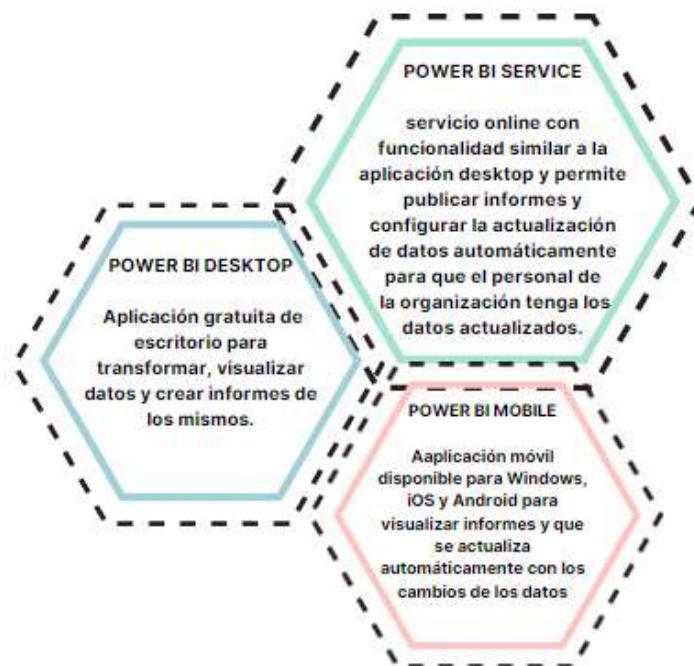


Ilustración 24. Componentes Power BI. Información tomada del artículo. Elaborado por Mixi Segura

2.3.5.2. VENTAJAS POWER BI

Las empresas generan cada vez más datos y tienen la necesidad de ordenarlos y procesarlos. Power BI ayuda a las organizaciones a interpretar lo que sucedió en el pasado y a predecir lo que podría suceder en el futuro. (Evotic, 2022). A continuación, procedemos a resumir algunas de las principales ventajas de este software:

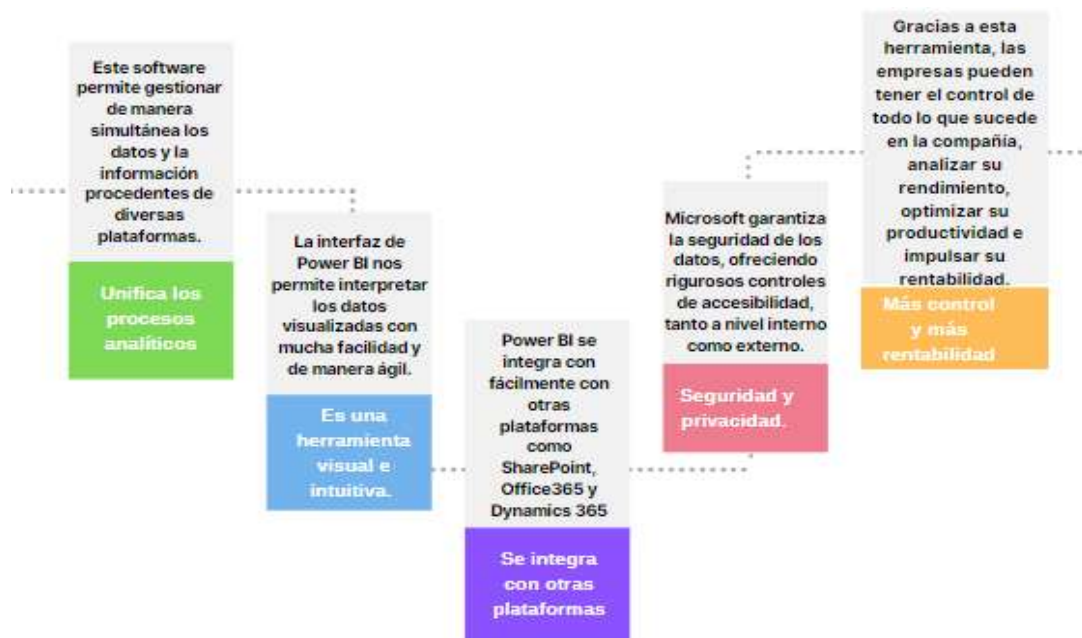


Ilustración 25. Ventajas Power BI. Información tomada del artículo. Elaborado por Mixi Segura

2.3.5.3. TIPOS DE LICENCIAS: FREE, PRO Y PREMIUM

Microsoft comercializa una amplia gama de versiones y planes disponibles para Power BI, con tarifas diferentes. Power BI Desktop y Power BI Mobile son gratuitos, mientras que los precios de Power BI Pro varían en función de cuántos usuarios tengan acceso a la herramienta y los de Power BI Premium responden a una tarifa mensual. (Evotic, 2022)



Ilustración 26. Tipos de Licencia de Power BI. Información tomada del artículo. Elaborado por Mixi Segura

2.3.6. DEVELOPER TWITTER

Twitter ama a los desarrolladores. Estamos encantados y sorprendidos por las herramientas y servicios que crea esta comunidad al aprovechar el poder de la plataforma del desarrollador de Twitter. Como parte de nuestro compromiso con esta comunidad, queremos ofrecer un acceso a los datos que sea abierto y justo para los desarrolladores, seguro para las personas en Twitter y beneficioso para la plataforma de Twitter en conjunto. Para lograr estos objetivos, hemos confeccionado la Política de Desarrolladores como una guía para ayudar a que las personas entiendan nuestras reglas y expectativas sobre el uso adecuado de la API y del Contenido de Twitter.

Esta Política de Desarrolladores, proporciona reglas y pautas para los desarrolladores que interactúan con el ecosistema de aplicaciones, servicios, sitios web, páginas web y contenido de Twitter. Es parte de su contrato con Twitter regular el acceso y el uso de la API de Twitter y del Contenido de Twitter. Los incumplimientos de la política se consideran incumplimientos de su contrato. Esta Política puede cambiar de un momento a otro sin previo aviso. Los términos en mayúsculas utilizados en esta Política, pero que no están definidos en ella, tendrán los significados respectivos que se les atribuyen en el Contrato de Desarrolladores o en el Contrato Marco de Licencia. (Developer, 2022)



Ilustración 27. Developer. Información tomada de la web. Elaborado por Mixi Segura

2.3.6.1. POLÍTICAS DE USO

Según en la página oficial de (Developer, 2022) se dice que han estructurado las políticas para que sea lo más fácil de seguir posible. Recuerde la información de las siguientes secciones con las políticas mientras usa la API de Twitter y el Contenido de Twitter:

Prepárese para el éxito: usted es responsable de cumplir con todas las políticas de Twitter. Es importante que revise y entienda esta Política, así como las políticas mencionadas en este documento, antes de comenzar a usar la API de Twitter y el Contenido de Twitter. El tiempo que pase revisando nuestras políticas puede ahorrarle horas de trabajo de revisión más adelante.

La Privacidad y el Control son Esenciales: la protección y defensa de la privacidad de las personas en Twitter están fijadas en el ADN principal de nuestra empresa. De esta forma, está prohibido el uso de los datos de Twitter de cualquier forma que no sea coherente con las expectativas razonables de privacidad de las personas. Al construir en la API de Twitter o acceder al Contenido de Twitter, usted tiene un papel especial para proteger este compromiso, sobre todo al respetar la privacidad de las personas y ofrecerles transparencia y control con respecto a cómo se usan sus datos.

Siga las Pautas de Uso de la Plataforma: tener la aprobación para acceder a la API de Twitter y al Contenido de Twitter es solo el primer paso. Nuestras Pautas de Uso de la Plataforma deben ser lo primero que consulte cuando tenga alguna pregunta sobre cómo garantizar el cumplimiento de la política para su uso planeado de la plataforma de Twitter.

Le hemos proporcionado más detalles sobre qué significa cada una de estas tres secciones clave a continuación. Revíselas con atención para asegurarse de que su uso de la API de Twitter y del Contenido de Twitter sea coherente con nuestras políticas. Si consideramos que incumple esta Política (o cualquier otra política de Twitter), podemos suspender o revocar de forma permanente el acceso a la API de Twitter y al Contenido de Twitter. Si le ocurre esto, no solicite ni registre claves de API adicionales. En lugar de eso, comuníquese con nosotros a través del Formulario de Política de Apoyo de API

Finalmente, observe que Twitter puede supervisar el uso de la API de Twitter para mejorar las Aplicaciones de Twitter, examinar el uso comercial y garantizar el cumplimiento de su caso de uso aprobado y de esta Política.

2.3.7. ANACONDA

Es un administrador de paquetes, una distribución de Python y una colección más de 1000 paquetes de código abierto los cuales son fáciles de instalar y tienen un soporte gratuito a través de su comunidad. (Valdez, 2018)

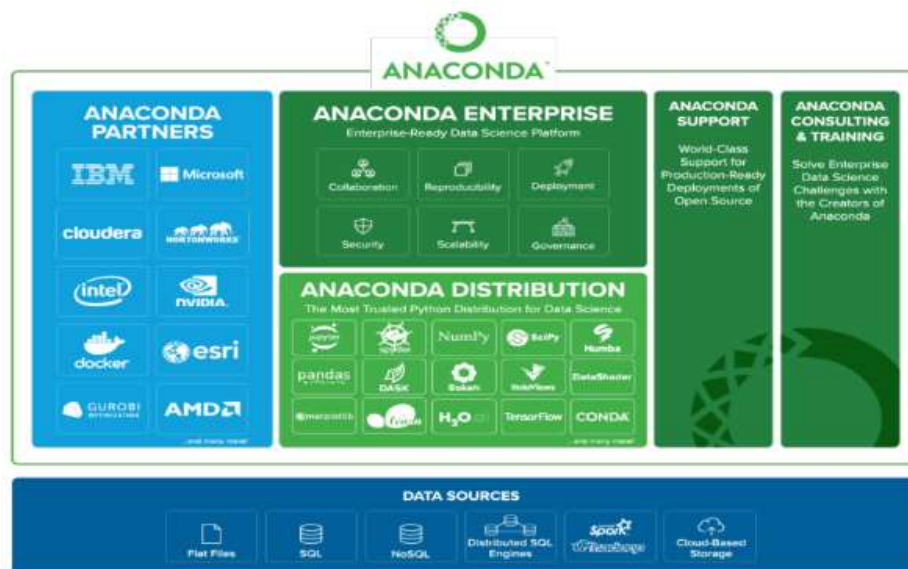


Ilustración 28. Anaconda. Información tomada de la web. Elaborado por Mixi Segura

2.3.8. PYTHON

Es un software libre y gratuito con licencia BSD (Este tipo de licencia es muy permisiva para las empresas), es una plataforma que funciona en diferentes sistemas operativos como: Windows, MacOS y Linux con un lenguaje de alto nivel conveniente para un desarrollo rápido de código. Python tiene un amplio espectro de programadores, tanto aquellos que son ocasionales, como aquellos que son desarrolladores, una comunidad grande y activa, podemos mencionar: PyCon, SciPy, EuroPython, PyData, entre otras. (Valdez, 2018)

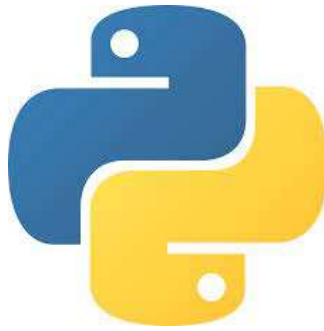


Ilustración 29. Python. Información tomada de la web. Elaborado por Mixi Segura

2.4. MARCO LEGAL

En la (ASAMBLEA, 2021) Esta nueva de forma de comunicación permite la interacción de los electores en la red, ya que genera una comunicación inmediata, de doble vía y una difusión del mensaje de manera más rápida y directa que por otras vías. Herramientas como los blogs, redes sociales o páginas web, han dado la posibilidad al elector de acceder a la información en el tiempo y lugar que este prefiera, cambiando así el canal del mensaje, la forma de resección e incluso la construcción de este, con relación a las formas clásicas de comunicación.

Según la Ley Orgánica del Ecuador del uso responsable de redes sociales, en su artículo 1 su objeto es el uso responsable de las redes sociales con el fin de garantizar una nueva forma de comunicación responsable, ya que las redes no solo son fuentes de información si no medio de difusión de contenidos, y pueden contribuir al fortalecimiento de la institución y a la interacción con toda la ciudadanía del país y todo el mundo. En lo expresado en el artículo 2.- La presente Ley se aplica a todas personas naturales y jurídicas, su ámbito de actividad se extiende a todo el territorio nacional, sin perjuicio de lo dispuesto en los tratados internacionales suscritos en esta materia por el Ecuador. (ASAMBLEA, 2021)

Artículo 3.- Naturaleza y finalidad. La ley para el uso responsable de redes sociales es una ley que pretende comprometer a los usuarios de estas que sean responsables con la información que publiquen o compartan por este medio, ya que dicha información es de rápida difusión y puede llegar afectar a terceros e incluso a poner en riesgo la estabilidad Nacional.

La finalidad de dicha Ley es crear responsabilidad en los usuarios de las redes sociales, como al igual que quienes comunican a través de los medios de comunicación impresos y electrónicos, tienen una gran responsabilidad. Un comentario sin sustento puede implicar una violación a derechos de terceros, un atentado a la dignidad de las personas o afectar la reputación de la institución; las redes no solo son fuentes de información si no medio de difusión de contenido, y pueden contribuir al fortalecimiento de la institución y a la interacción con toda la ciudadanía del país y todo el mundo. (ASAMBLEA, 2021)

Artículo 4.- Prohibición. La presente Ley prohíbe la divulgación de información falsa, o que no se le haya obtenido por fuentes confiables que perjudiquen a un tercero, ya que sea persona natural, jurídica e incluso que ponga en peligro la seguridad estatal, y su sanción será la cual disponga la normativa pertinente de acuerdo con el daño causado. Según el artículo 5 de la presente Ley, la cual indica que toda persona natural o jurídica que será responsable de la información que comparta o divulgue en redes sociales. Y la misma tendrá la sanción respectiva si incumple lo indicado en esta Ley.

Referente al artículo 6 toda persona natural o jurídica que quiera hacer uso de información personal de algún determinado perfil, de alguna red social necesitará el permiso del propietario de esta, caso contrario será sancionado civil o penalmente dependiendo el caso. Según el artículo 7 será necesaria la autorización expresa de los usuarios para capturar y almacenar su información personal, así como la transparencia en la utilización de esas bases de datos. Caso contrario será sancionado civil o penalmente dependiendo el caso.

Capítulo III

La propuesta

3.1. Metodología del Proyecto

Según (Coelho, 2019) metodología se denomina a la serie de métodos y técnicas de rigor científico que se aplican sistemáticamente durante un proceso de investigación para alcanzar un resultado teóricamente válido.

La metodología que se utilizará es la KDD, la cual consta de 5 fases: Selección, Análisis y extracción de datos, Limpieza y validación de datos, Minería de datos e Interpretación.

En la primera fase “selección”, definiremos a la red social Twitter como nuestra fuente de información, la cual nos ayudará aportando mejor comprensión del dominio del estudio y establecimiento de objetivos para el actual proyecto de tesis, en la segunda fase “Análisis y extracción de datos”, nos tocará analizar todos los tweets, comentarios de los electores digitales y elegir de manera acertada la información que vamos a necesitar en el proceso de extracción, en la tercera fase “Limpieza y validación”, se tratará de mejorar la calidad de los datos, eliminando los datos erróneos y validar los resultados de la minería, en la cuarta fase “minería de datos”, nos referimos al algoritmo a utilizar mediante el cual llegaremos al objetivo de una manera correcta, finalizando el proceso aplicando la quinta y última fase presentando una oportuna interpretación de los datos minados. Posteriormente elegiremos un Dashboard – Power BI como herramienta viable para una mejor visualización, el cual nos ayudará a que los resultados sean presentados de una forma entendible.



Ilustración 30. Metodología. Información tomada de la web. Elaborado por Mixi Segura

3.1.1. DISEÑO DE INVESTIGACIÓN

El diseño de investigación en este proyecto de tesis tiene un enfoque mixto, cualitativo y cuantitativo, ya que se basará en las diferentes opiniones de los electores digitales sobre la reputación en línea de los precandidatos a la alcaldía de Guayaquil y aquellos resultados serán analizados, interpretados visualmente a través de diagramas, permitiendo la integración de información más completa de los datos recopilados cuantitativa y cualitativamente de manera separadas.

3.1.1.1. DISEÑO CUASI EXPERIMENTAL

Este proyecto es un diseño cuasi experimental ya que escoge los tweets de una manera aleatoria en la extracción de datos.

3.1.2. TIPO DE INVESTIGACIÓN

(Enrique, 2020) nos comenta que los tipos de investigación pueden agruparse según el objetivo que persiguen, el nivel de profundización, la forma de hacer inferencia estadística, la forma de manipular variables, el tipo de datos o el período de tiempo de estudio. Para llevar a cabo el presente estudio, se consideró varios tipos de investigación, los cuales son:

3.1.2.1. INVESTIGACIÓN EXPLORATORIA

Esta investigación nos permitirá realizar una indagación, la cual se llevará a cabo en la red social Twitter, recolectando información de los precandidatos y así tener conocimiento actual al momento de seleccionar los cuatros con más interacción con los usuarios.

3.1.2.2. INVESTIGACIÓN DESCRIPTIVA

Esta investigación nos ayudará a realizar un análisis del estudio propuesto, el cual tiene como objetivo describir las características para comprender de manera más exacta la cantidad de tweets, comentarios, interactuados por los usuarios en relación con

la reputación en línea que tiene cada uno de los precandidatos en función a la imagen que proyectan en la red social.

3.1.3. TÉCNICA DE INVESTIGACIÓN

(Ana, 2021) nos hace referencia que las técnicas de investigación son los recursos que permiten obtener de forma organizada y coherente los datos y/o la información para la búsqueda de conocimiento.

3.1.3.1. MUESTREO

Esta técnica nos ayudará a seleccionar una cierta cantidad de tweets, lo cual proporcionará una información precisa y detallada de lo que los usuarios piensan de los precandidatos, teniendo como resultado una aproximación general de la reputación en base a la población.

3.1.3.1.1. NO PROBABILÍSTICO

Nos fijamos en este método porque la selección de los tweets, no son basados en técnicas estadísticas si no en una extracción, que no todos los usuarios tienen probabilidad de ser elegidos para formar parte de la investigación, esto se basa por lo que es una red social, la cual tiene millones de usuarios que realizan tweets, entonces no todos pueden ser parte del muestreo.

3.2. DESCRIPCIÓN DEL PROYECTO

El presente Proyecto tiene como objetivo desarrollar un análisis de reputación en línea de precandidatos electorales en Twitter mediante técnica de minería de datos, el cual va a hacer realizado mediante una plataforma y una metodología que consta de 5 fases: Selección, preprocesamiento, transformación, minería de datos y evaluación e implantación.

Los resultados de la extracción de datos serán llevados a un Dashboard para una mejor visualización la cual nos ayudará al análisis de la reputación en línea de los precandidatos seleccionados.

3.2.1. FACTIBILIDAD TÉCNICA

Para efectuar el análisis de los datos, utilizaremos la API de Twitter la cual nos ayudará en la extracción de los tweets de los usuarios de la red social, para nosotros tener el acceso de los tokens generados debemos de registrarnos en developer. Para nosotros comenzar con la recopilación de los datos debemos de elegir en que programa vamos a trabajar en este caso escogimos Jupyter Notebook, ya que es una interfaz web, de fácil uso, también debemos de elegir en que lenguaje de programación nos vamos a basar, en este caso elegimos Python, ya que es un software libre y tiene una gran variedad de librería relacionadas con la minería de datos, una vez extraído los datos ya limpios y filtrados, los resultados serán visualizados de forma accesible en un dashboard, que será Power Bi.

Tabla 9. Softwares utilizados en el desarrollo del proyecto

SOFTWARE	VERSIÓN
Developer Twitter	-
Anaconda	3.2021.11.
Jupyter Notebook	6.4.12
Python	3
Formato CSV	Excel
Power Bi desktop	2.108.997.0 64-bit

Información obtenida de la investigación elaborado por Mixi Segura

3.2.2. FACTIBILIDAD LEGAL

El presente proyecto analiza la reputación en línea de precandidatos electorales en Twitter mediante técnica de minería de datos, el cual no pretende perjudicar las leyes involucradas para llevar a cabo su desarrollo; mediante la confidencialidad de los precandidatos, como los derechos que tienen los usuarios en opinar en las redes sociales.

Posteriormente, se detalla los artículos referenciados, los cuales son respaldados en la constitución de la república del Ecuador:

- Ley orgánica del uso de las redes sociales
- Constitución de la república del Ecuador

3.3 ESQUEMA GENERAL DEL ANÁLISIS DE DATOS

En el siguiente gráfico se detallará el procedimiento que se realizará para el desarrollo del análisis de la reputación en línea de los precandidatos electorales, con la intención de proyectar el pensamiento del elector digital.

En el desarrollo se aplicará las 5 fases, comenzando con la identificación de los datos y culminando con la visualización de los datos, los cuales serán representados mediante un dashboard, adicionalmente se podrán observar las herramientas y programas utilizados en el proceso.



Ilustración 31. Esquema general del análisis de datos. Información tomada de la web. Elaborado por Mixi Segura

3.4. RECURSOS PARA EL ANÁLISIS DE DATOS

Es muy útil actualmente, el contar con recursos que nos faciliten el análisis de datos entre los cuales podemos referenciar la técnica de minería de datos, que nos ayudará en primera instancia en la extracción de la información requerida,

posteriormente utilizaremos herramientas, programas, lenguajes de programación, que fueron los óptimos en el proceso de desarrollo de este proyecto de tesis y obtener buenos resultados al momento de la visualización.

3.4.1. ANACONDA

Es una suite de código abierto que engloba una serie de aplicaciones como Jupyter Notebook, librerías diseñadas para la extracción de datos mediante Python 3 de una manera muy práctica.

3.4.2. API TWITTER

Este api nos ayudará a obtener la información adecuada para una extracción de datos, mediante los tokens generados, los cuales nos permitirá el acceso como desarrollador en la red social y así cumplir con el objetivo propuesto.

3.4.3. TWEETS

Los tweets son publicaciones que realizan los electores digitales en la red social, que pueden contener textos, fotos y videos. Los cuales pueden ser expresados dependiendo del tema que desean exponer a los usuarios. Para el análisis de la reputación en línea se extraen tweets, comentarios de los cuatros precandidatos a la alcaldía de Guayaquil.

3.4.4. ENTORNO DE DESARROLLO: JUPYTER NOTEBOOK

En este punto podemos describir a Jupyter Notebook como un proceso, el cual nos servirá de soporte para la extracción de datos, filtración y limpieza de la información mediante el lenguaje de programación Python, el cual se ejecutará el código de una manera más fácil por medio de una interfaz más ordenada mediante cuadernillos y poder detectar los errores en las líneas del código y de esta manera conocer la reputación que tiene cada uno de los precandidatos a la alcaldía de Guayaquil.

3.4.5. LENGUAJE DE PROGRAMACIÓN: PYTHON

Para la ejecución del análisis de los datos utilizaremos Python como un lenguaje de programación de alto nivel ya que cuenta con un código abierto y es usado para obtener información de sitios web ya que dispone de diversas librerías referentes con data mining, la cual se enlazará como punto de conexión con la Api de Twitter para la recopilación de la información. Dentro de las librerías más utilizadas podemos mencionar:

Tabla 10. Definiciones de las librerías utilizadas

LIBRERÍAS	DEFINICIÓN
TWEEPY	Nos sirve de conexión para el servidor de la red social mediante las contraseñas generadas con developer, de una forma mucho más rápida y eficiente.
PATHLIB	Da acceso a rutas del sistema de archivos con semántica apropiada para diferentes sistemas operativos.
TEXTBLOB	Extrae etiquetado de parte de los tweets, frases sustantivas, análisis de sentimientos, y más.
WORDCLOUD	Define la salida final de nuestro algoritmo.

PANDAS

Los datos son fáciles de trabajar, ya que sus estructuras están conformadas por una serie de datos, así como el Data Frame para dos dimensiones.

NUMPY

Permite generar una estructura de datos universal, lo que se traduce por un mejor análisis de datos, ya que se da un intercambio de datos entre diferentes algoritmos.

MATPLOTLIB

Genera una gran variedad de gráficos de calidad necesaria para publicarlas tanto en papel como digitalmente.

NLTK

Es una de las librerías más antiguas en Python para procesamiento de lenguaje natural.

TQDM

Nos brinda una barra de progreso simple en nuestro código, se verá muy agradable a la vista y le dará la retroalimentación adecuada al usuario cada vez que ejecuta el código.

STOPWORDS

Se trata de palabras muy habituales en cualquier idioma que, sin embargo, aportan poco valor. Por ello, es interesante identificarlas y excluirlas. Es una forma de limpiar el texto.

3.4.6. FORMATO CSV

Este formato es exportado por Jupyter Notebook, cuando extraemos los tweets, al momento que los datos están ya limpios, es una forma de tabular toda la información por medio de columnas y filas con la extensión de .CSV.

3.5. PROCEDIMIENTO: APLICACIÓN DE LA METODOLOGÍA ESCOGIDA

Para realizar el actual proyecto que se trata del análisis de reputación en línea de precandidatos electorales en Twitter mediante técnica de minería de datos, elegimos la metodología Descubrimiento de Conocimiento en Base de Datos que se desarrolla mediante 5 fases, con el fin de cumplir con los objetivos planteados. Posteriormente se explicará cada fase:

3.5.1. SELECCIÓN

En esta primera fase que se trata de seleccionar la herramienta, la cual nos va a ayudar para la extracción de información, teniendo como objetivo el desarrollo de un análisis de reputación en línea de precandidatos electorales mediante técnica de minería de datos, el a ver escogido la red social Twitter nos brindará la información de las interacciones que los electores digitales tienen constantemente en las cuentas de los precandidatos, basado en aquella información podremos analizar la reputación que tienen cada uno de los precandidatos.

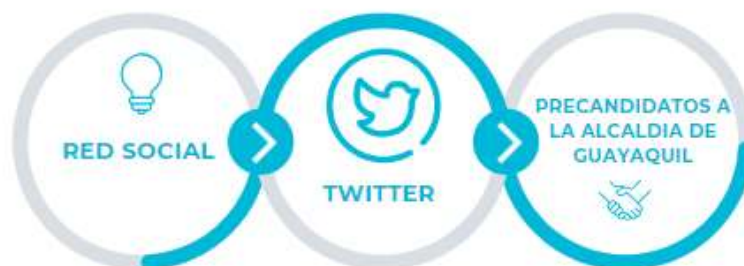


Ilustración 32. Selección de herramienta. Información obtenida del proyecto. Elaborado por Mixi Segura.

3.5.2. ANÁLISIS Y EXTRACCIÓN DE DATOS

En esta fase se basa en la extracción de los tweets, comentarios que hacen los electores digitales sobre cada uno de los precandidatos, dependiendo de la imagen que ellos proyecten en la red social Twitter.

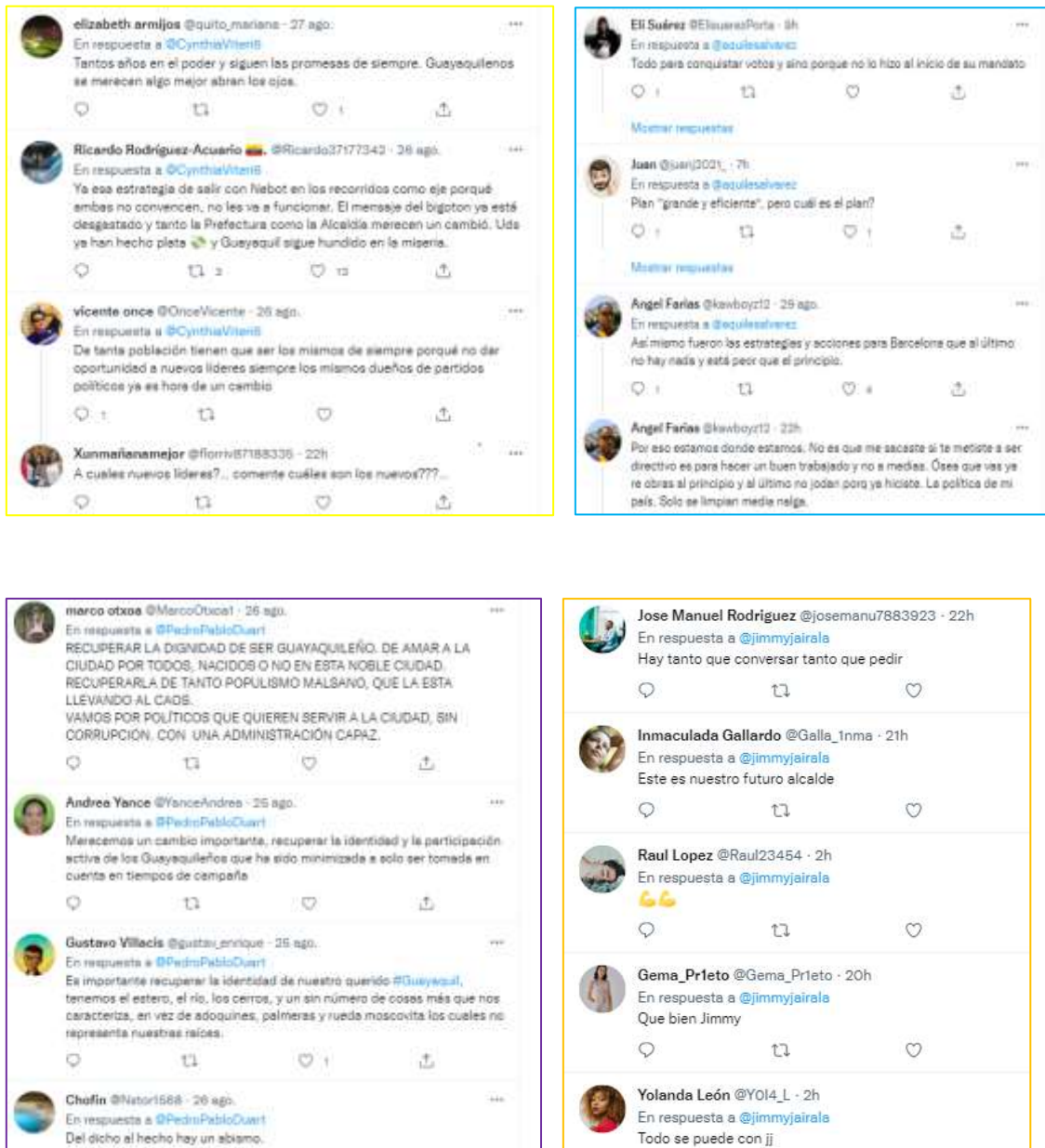


Ilustración 33. Tweets. Información tomada de la aplicación de Twitter. Elaborado por Mixi Segura

3.5.2.1. ANÁLISIS DE DATOS

Como podemos observar en el primer cuadernillo de la ilustración 34, contamos con las librerías que vamos a trabajar, las cuales tienen diferentes funciones, para poder realizar las extracciones de aquellos tweets, debemos de conectarnos con los servidores de la red social Twitter con la API, luego crear un documento web en Jupyter Notebook en el cual vamos a programar en Python.

Luego tenemos en el segundo cuadernillo las contraseñas generadas por la Api mismo, las cuales al momento de la creación de developer, nos salen ya definidas. Debemos de ponerlas en el código después de las librerías para tener el acceso a la red social y nos permita la extracción de los tweets y comentarios. Las cuales son:

- Consumer_key
- Consumer_secret
- Access_token
- Access_token_secret.

```
In [1]: # Librerias

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')

In [2]: #contraseñas

credentials = pd.read_csv('./Credentials/CredentialsStudent.csv')
consumerKey = credentials['key'][0]
consumerSecret = credentials['key'][1]
accessToken = credentials['key'][2]
accessTokenSecret = credentials['key'][3]

In [17]: # 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 34. Código de programación. Información de Python. Elaborado por Mixi Segura

3.5.2.2. EXTRACCIÓN DE DATOS

Una vez que ya estamos conectados con Twitter y tenemos el permiso a la información de dicha red social, comenzamos con la extracción de los tweets de los precandidatos mediante la función Tweepy, como se puede visualizar en la Ilustración 35.

```
In [17]: # Get Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@CynthiaViteri6").items(700):
    print(status.text)

RT @CynthiaViteri6: MÁS DE 30.000 HABITANTES DEL KM 8 X VÍA A DAULE SE BENEFICIAN CON NUESTRAS LABORES DE PA
VIMENTACIÓN

Familias de Juan M.
RT @Joffre0: @ATMGuayaquil @ATM_Transito @CynthiaViteri6 pero no se dan cuenta... así por todos lados. todo
es un caos. un desorden. que de...
RT @Morah73407026: Hombre de detalles
@JosueSanchezEc
@CynthiaViteri6
@DIANA_TMR
@lospremiosheat
@EpicoGye
MELFI X FESTIVAL HEAT GYE
RT @lahistoriaec: De realities, reinados y TV. Alcaldesa @CynthiaViteri6
presentó a los candidatos a concejales del PSC en el distrito 3 de...
@CynthiaViteri6 Señora pero y sus revuelos, los negociados de sus maridos ?
Los favores a sus familiares?
El sobreprecio de las obras ?
RT @Morah73407026: Yapingacho
-- -- --
```

```
In [6]: # Get Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@aquilesalvarez").items(700):
    print(status.text)

RT @aquilesalvarez: Guayaquileño, te invito a ejecutar el Plan de Seguridad más grande y eficiente en la his
toria de la ciudad. Estrategias...
RT @GyeconAquiles: Las mujeres guayaquileñas merecen una ciudad segura, donde exista respeto e igualdad de o
portunidades para todas y todos...
RT @aquilesalvarez: Guayaquileño, te invito a ejecutar el Plan de Seguridad más grande y eficiente en la his
toria de la ciudad. Estrategias...
RT @aquilesalvarez: Guayaquileño, te invito a ejecutar el Plan de Seguridad más grande y eficiente en la his
toria de la ciudad. Estrategias...
RT @GyeconAquiles: Las mujeres guayaquileñas merecen una ciudad segura, donde exista respeto e igualdad de o
portunidades para todas y todos...
RT @aquilesalvarez: Guayaquileño, te invito a ejecutar el Plan de Seguridad más grande y eficiente en la his
toria de la ciudad. Estrategias...
RT @Javier92_5: @aquilesalvarez Esto deberían aprender @AlfaroMoreno @LuigiMB @MunirMassuh @rafaelverduga da
rse cuenta que no van más no le...
RT @aquilesalvarez: Guayaquileño, te invito a ejecutar el Plan de Seguridad más grande y eficiente en la his
toria de la ciudad. Estrategias...
@aquilesalvarez Sin miedo lanza tu candidatura fuera de la lista Robolucion para que aterrices tu realidad
-- -- --
```

```
In [4]: # Get Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@PedroPabloDuart").items(700):
    print(status.text)

RT @PedroPabloDuart: Más que solo acuerdos. 🙏 Es un compromiso con nuestros jóvenes y nuestro futuro. Graci
as a La Casa de la Cultura que c...
@JennyHilll @__ROCCO_ @PedroPabloDuart Nunca respondió.!
Así noooo.!
RT @LuisMauretleon: Etiquetar de informales a un padre o madre de familia que sale a las calles a buscar el
pan para sus hijos es discrimin...
RT @PedroPabloDuart: Les cuento un poco de mi vida, para que me conozcan más de cerca 🙏

Desde hoy al futuro la historia la vivimos juntos...
@PedroPabloDuart Excelente ser humano 🙏
@PedroPabloDuart Me importa un pepino tu vida pendejada
@PedroPabloDuart Vamos guayaquileños honrados y trabajadores, dejemos a un lado a tanto politiquero corrupto
y pica... https://t.co/B4X4AVXn3e (https://t.co/B4X4AVXn3e)
RT @PedroPabloDuart: Más que solo acuerdos. 🙏 Es un compromiso con nuestros jóvenes y nuestro futuro. Graci
as a La Casa de la Cultura que c...
RT @PedroPabloDuart: Más que solo acuerdos. 🙏 Es un compromiso con nuestros jóvenes y nuestro futuro. Graci
as a La Casa de la Cultura que c...
RT @PedroPabloDuart: Más que solo acuerdos. 🙏 Es un compromiso con nuestros jóvenes y nuestro futuro. Graci
as a La Casa de la Cultura que c...
RT @PedroPabloDuart: Más que solo acuerdos. 🙏 Es un compromiso con nuestros jóvenes y nuestro futuro. Graci
as a La Casa de la Cultura que c...
```



```

In [4]: # Get Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@jimmyjairala").items(700):
    print(status.text)

@anontalvoch en el ministerio sale asi pero en la cenescyt no consta por favor ayuda 🙏🙏 hemos invertido t
iempo dine... https://t.co/KkXXqCEILG (https://t.co/KkXXqCEILG)
@UnCafeConJJ @jimmyjairala @SolcaGuayaquil @IESSec Buena noticia.
@FernandoCedeR @Guillermolasso @BancadaUNESec @OrlandoPerezEC @teleSURtv @UnCafeConJJ @jimmyjairala @fabrici
ovelav... https://t.co/Z7XJ3wVG9p (https://t.co/Z7XJ3wVG9p)
RT @FDSPEecu: @periodicodeuna @panchojimenezs @LassoGuillermo @LorenzoCalvas @GomezRichard @paultutiven @ji
mmyjairala
@periodicodeuna @panchojimenezs @LassoGuillermo @LorenzoCalvas @GomezRichard @paultutiven @jimmyjairala
@AsambleaEcuador @FiscaliaEcuador @DianaSalazarM2 @PGEcuador @VsaquicelaE @CONAIE_Ecuador @PKnacional18... htt
ps://t.co/UdOYOp6a2o (https://t.co/UdOYOp6a2o)
RT @kevinchvarria: @FernandoCedeR @Guillermolasso @BancadaUNESec @OrlandoPerezEC @teleSURtv @UnCafeConJJ @ji
mmyjairala @fabriciovelav @Rica...
RT @JorgeAndresMC: @FernandoCedeR @Guillermolasso @BancadaUNESec @OrlandoPerezEC @teleSURtv @UnCafeConJJ @j
immyjairala @fabriciovelav @Ric...
RT @toaniel: @MonicaBritoRC @FernandoCedeR @Guillermolasso @BancadaUNESec @OrlandoPerezEC @teleSURtv @UnCafe
ConJJ @jimmyjairala @fabriciove...
RT @AromaniaPerfum: @FernandoCedeR @Guillermolasso @BancadaUNESec @OrlandoPerezEC @teleSURtv @UnCafeConJJ @j
immyjairala @fabriciovelav @Ric...
RT @toaniel: @tefyJoys @FernandoCedeR @Guillermolasso @BancadaUNESec @OrlandoPerezEC @teleSURtv @UnCafeConJJ

```

Ilustración 35. Código de programación. Información de Python. Elaborado por Mixi Segura

Luego de la extracción de los datos, creamos una carpeta para que se guarde los tweets, la cual por defecto viene con el formato CSV, en la parte de la función `items()` ingresamos la cantidad de tweets que se extrajo, en el otro cuadernillo ingresamos un código para que dicho formato se ha leído con la función `read_csv` y se nos presente en forma de tabla como podemos visualizar en la Ilustración 37, la cual viene con los siguientes atributos: fecha, hora, usuario, locación, following, entre otras.

En la función `head()` ingresamos la cantidad de columna de datos que se desee visualizar, en este caso escogimos la cantidad de 10 columnas.

```

In [19]: #Create a file with Tweets

path = Path('./DataSets/CynthiaV/Cynthia.csv.csv')
if(path.is_file()):
    os.remove('./DataSets/CynthiaV/Cynthia.csv.csv')

#Fill rows CSV
with open ('./DataSets/CynthiaV/Cynthia.csv', "w", newline= '\n', encoding="utf-8") as csvwriter:
    writer = csv.writer(csvwriter, delimiter= "|")
    for status in tweepy.Cursor(api_twitter.search_tweets, q="@Cynthiaviteris", 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", "")])

In [20]: #Read CSV

dataset = pd.read_csv('DataSets/CynthiaV/Cynthia.csv', names= ['geo_enabled/conectados', 'status.source/dispositivo',
                        'Fecha', 'user.name', 'user.location',
                        'user.following/siguiendo', 'place', 'retweet_count', 'full_text',
                        ])

dataset.head(10)

```

Ilustración 36. Código de programación. Información de Python. Elaborado por Mixi Segura

Out[20]:

_enabled/conectados	status.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count	
True	False	2022-08-30 04:53:46+00:00	Profesor del Tweet	NaN	NaN	10	False	RT
False	False	2022-08-30 04:53:38+00:00	Luz Elena Rovira M	Panamá	NaN	6	False	RT @
True	False	2022-08-30 04:53:16+00:00	ALBA MENDOZA PAREDES	Quito, Ecuador	NaN	17	False	RT
False	False	2022-08-30 04:50:45+00:00	fernando moreno	alausi	NaN	0	False	@
True	False	2022-08-30 04:49:07+00:00	LUIS SALAZAR	NaN	NaN	1	False	RT @
False	False	2022-08-30 04:46:40+00:00	Jeniffer Palma H	Guayaquil, Ecuador	NaN	1	False	ganai
False	False	2022-08-30 04:45:08+00:00	Jeniffer Palma H	Guayaquil, Ecuador	NaN	3	False	RT @
True	False	2022-08-30 04:44:54+00:00	Melany 🌟	Riobamba, Ecuador	NaN	6	False	RT @
False	False	2022-08-30 04:43:59+00:00	Azucena Jimenez	Ecuador-Guayaquil	NaN	0	False	Has
False	False	2022-08-30 04:43:27+00:00	Jeniffer Palma H	Guayaquil, Ecuador	NaN	3	False	R c

Ilustración 37. Código de programación. Información de Python. Elaborado por Mixi Segura

3.5.3. LIMPIEZA Y VALIDACIÓN DE DATOS

En esta fase realizamos la limpieza de los tweets extraídos, eliminamos datos no válidos con la función CleanText () como: emoticones, signos de puntuación, links, entre otros, los cuales pueden causar una distorsión al momento del análisis.

```
In [21]: #Encadenar el texto

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
```

Ilustración 38. Código de programación. Información de Python. Elaborado por Mixi Segura

En el siguiente cuadernillo que podemos observar en la Ilustración 39, procedemos con la librería stopwords que nos ayudará a que la lista de los textos no validados se elimine y así tener un archivo más limpio, se procede a la validación y creación del archivo con el texto limpio.

```

In [22]: #Function StopWords
import unicode

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)

In [23]: #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 Cinthya.csv
dataset['tweet'] = dataset['full_text'].apply(lambda x : removeStopWords(x,'es'))

In [24]: #Create CSV Clean
dataset.to_csv('./DataSets/CynthiaV/Cynthia-Clean.csv', sep='|')

```

Ilustración 39. Código de programación. Información de Python. Elaborado por Mixi Segura

Y finalmente podemos observar en la tabla el texto limpio, sin emoticones, sin signos de puntuación, entre otros.

```
In [27]: dataset.head(10)
```

Out[27]:

conectados	status.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count		
True	False	2022-08-30 04:53:46+00:00	Profesor del Tweet	NaN	NaN	10	False	RT @lahistoria:	ri
False	False	2022-08-30 04:53:38+00:00	Luz Elena Rovira M	Panamá	NaN	6	False	RT @Victori3539	FESTIVA
True	False	2022-08-30 04:53:16+00:00	ALBA MENDOZA PAREDES	Quito, Ecuador	NaN	17	False	RT @egiscaice	kits i
False	False	2022-08-30 04:50:45+00:00	fernando moreno	alausi	NaN	0	False	@SelimF0i	@GuilleMorei
True	False	2022-08-30 04:49:07+00:00	LUIS SALAZAR	NaN	NaN	1	False	RT @CorderoLar	el mc
False	False	2022-08-30 04:46:40+00:00	Jeniffer Palma H	Guayaquil, Ecuador	NaN	1	False	RT @Norah	ganas@JosueSai
False	False	2022-08-30 04:45:08+00:00	Jeniffer Palma H	Guayaquil, Ecuador	NaN	3	False	RT @Norah7340	bola @Jk
True	False	2022-08-30 04:44:54+00:00	Melany 🌟	Riobamba, Ecuador	NaN	6	False	RT @Victori3539	FESTIVA
False	False	2022-08-30 04:43:59+00:00	Azucena Jimenez	Ecuador- Guayaquil	NaN	0	False	Hasta cuándo?	los
False	False	2022-08-30 04:43:27+00:00	Jeniffer Palma H	Guayaquil, Ecuador	NaN	3	False	RT @Norah7:	con guandú@

Ilustración 40. Código de programación. Información de Python. Elaborado por Mixi Segura

3.5.4. MINERÍA DE DATOS

Esta etapa se trata de aplicar a los datos que ya fueron extraídos, limpiados y procesados el algoritmo a utilizar, el cual nos permitirá a llegar a nuestro objetivo, en este caso nuestro proyecto va dirigido al análisis de reputación en línea que nos ayudará a tener una mejor interpretación de los precandidatos a la Alcaldía de Guayaquil.

Para nosotros llegar a nuestro objetivo, primero debemos de analizar las frecuencias de las palabras ya sean positivas y negativas, en la siguiente ilustración 41 podremos visualizar el código, el cual con la función value nos dará a conocer la frecuencia en que las palabras son repetidas en los tweets y comentarios.

```
In [35]: #Análisis de frecuencia de palabra

lista_texto= clean_texto.split(" ")
palabras = []

#Paso eliminar palabras muy cortas y muy largas.

for palabra in lista_texto:
    if (len(palabra)>=3 and len(palabra)<18):
        palabras.append(palabra)

#Generalizador de palabras
word_count={}

for palabra in palabras:
    if palabra in word_count.keys():
        word_count[palabra][0]+=1
    else:
        word_count[palabra]=[1]

#Generar DF y Lo ordenamos
df=pd.DataFrame.from_dict(word_count).transpose()
df.columns=["freq"]
df.sort_values(["freq"],ascending=False, inplace=True)

df.head(50)
```

Out[35]:

Ilustración 41. Código de programación. Información de Python. Elaborado por Mixi Segura

Y así es como nos aparecerá en el siguiente cuadernillo las palabras con el número de frecuencia.

Out[21]:			
		freq	
	cynthiaviteri6	458	alcaldiagye 38
	guayaquil	178	viteri 36
	ver	113	cnegobec 34
	vuelcita	88	rosada 30
	video	81	campana 29
	perteneció	80	egiscaicedo 28
	pega	80	solo 27
	rafaelsi	79	correa 25
	dios"	79	soledaddiab 25
	"ciudad	79	jrodriguezecu 24
	sabr�	79	ahora 24
	alcaldesa	73	hace 23
	cynthia	47	eduandradej 23
	centr...rt	44	libertadores 23
	casa	44	atmguaquil 22
	ciudad	42	vipcreativo 22
	gente	40	sanciones 22
	a�os	38	as� 21
			millones 21
			calle 21
			pondr� 21
			adelantada 21

Ilustraci n 42. C digo de programaci n. Informaci n de Python. Elaborado por Mixi Segura

Luego debemos de traducir el documento de los tweets extra do con la funci n `translate_text()`, la cual nos ayudar  con el procedimiento.

```
In [28]: # Traducci n de 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)

In [30]: #Texto completo traducido
dataset["tweet_EN"] = dataset['tweet'].apply(translate_Tweet)
dataset.head()
```

Ilustraci n 43. C digo de programaci n. Informaci n de Python. Elaborado por Mixi Segura

Creamos el archivo CSV traducido y con la función `read_csv` lo mandamos a leer para un nuevo análisis de los datos y ver que todo este correcto, como podemos observar en la ilustración 45 nos sale ya los tweets en inglés.

```

In [31]: #Creacion del archivo csv traducido
dataset.to_csv('./Datasets/Cynthiav/Cynthia-EN.csv', sep="|")

In [38]: #Lectura del archivo traducido
sentiment = pd.read_csv('./Datasets/Cynthiav/Cynthia-EN.csv', delimiter="|")

#clean cvs
sentiment.dropna(subset=['tweet', 'tweet_EN'], inplace=True)
sentiment = sentiment.rename(columns={'Unnamed: 0': 'ID'})

sentiment['tweet_EN'] = sentiment['tweet_EN'].apply(removeStopWords)
sentiment.head()

```

Ilustración 44. Código de programación. Información de Python. Elaborado por Mixi Segura

Out[38]:

dos	status.source/dispositivo	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count	full_text	tweet	tweet_EN
True	False	2022-08-30 04:53:48+00:00	Profesor del Tweet	NaN	NaN	10	False	RT @lahistoriaec: De realidades, reinados y TV....	rt realities reinados tv alcaldesa candidatos ...	rt realities reigns tv mayor candidates council...
alse	False	2022-08-30 04:53:38+00:00	Luz Elena Rovira M	Panamá	NaN	8	False	RT @Victori35394538: MELFI X FESTIVAL HEAT GYE...	rt melfi x festival heat gye imr	rt melfi festival heat gye imr
True	False	2022-08-30 04:53:18+00:00	ALBA MENDOZA PAREDES	Quito, Ecuador	NaN	17	False	RT @egiscaicedo: Son 10 mil kits de alimentos ...	rt mil kits alimentos entregaremos dias mil ni...	rt thousand food kits deliver thousand childre...
alse	False	2022-08-30 04:50:45+00:00	fernando moreno	alausi	NaN	0	False	@SelimF007 @psyalectali @GuilleMoreira98 @Cynth...	corrupto	corrupt
True	False	2022-08-30 04:49:07+00:00	LUIS SALAZAR	NaN	NaN	1	False	RT @CorderoLasso: Toma pero el modelo exitoso ...	rt toma modelo exitoso decian pichincha	rt takes successful model pichincha

Ilustración 45. Código de programación. Información de Python. Elaborado por Mixi Segura

Luego de a ver traducido los tweets, procedemos a realizar el código para el análisis de reputación, el cual nos aparecerá las barras estadísticas de cuantas palabras positivas, neutral y negativas tiene dicho precandidato.

```
# BarGraph
result['sentiment_type']=''
result.loc[result.compound>0,'sentiment_type']='POSITIVE'
result.loc[result.compound==0,'sentiment_type']='NEUTRAL'
result.loc[result.compound<0,'sentiment_type']='NEGATIVE'

In [45]:
import math
result.sentiment_type.value_counts().plot(kind='bar',title="Análisis de reputación")

Out[45]: <AxesSubplot:title={'center':'Análisis de reputación'}>
```

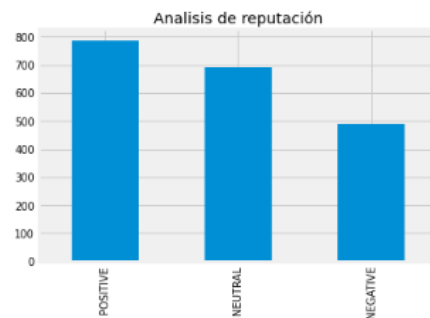


Ilustración 46. Código de programación. Información de Python. Elaborado por Mixi Segura

Para tener una idea más clara y saber de dichas palabras cuanto es el porcentaje tanto en positivo, neutral y negativo. Realizamos el código con la ayuda de la función `plot()`.

```
#Graph Pie
result.sentiment_type.value_counts().plot(kind='pie',
                                           subplots = True,
                                           title="Análisis de reputación",
                                           autopct='%1.0f%%',
                                           y=None,
                                           legend=True,
                                           labeldistance=1,
                                           figsize=(20, 10))
```

```
Out[46]: array([<AxesSubplot:ylabel='sentiment_type'>], dtype=object)
```

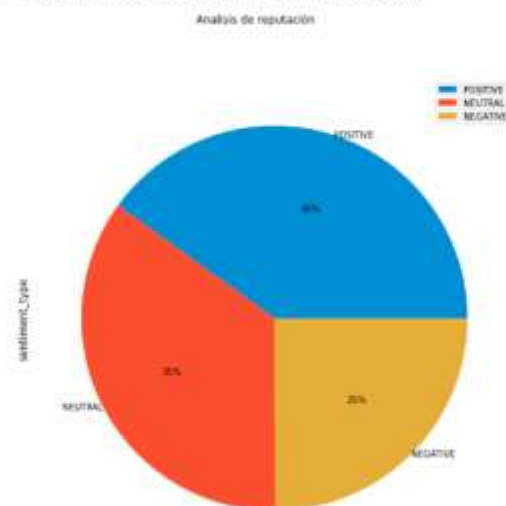


Ilustración 47. Código de programación. Información de Python. Elaborado por Mixi Segura

Para nosotros poder visualizar y analizar lo que los electores digitales piensan de los precandidatos, tuvimos que realizar el siguiente código que pueden observarlo en la Ilustración 48, con la ayuda de la función WordCloud que nos permitió crear una nube de las palabras positivas como negativas.

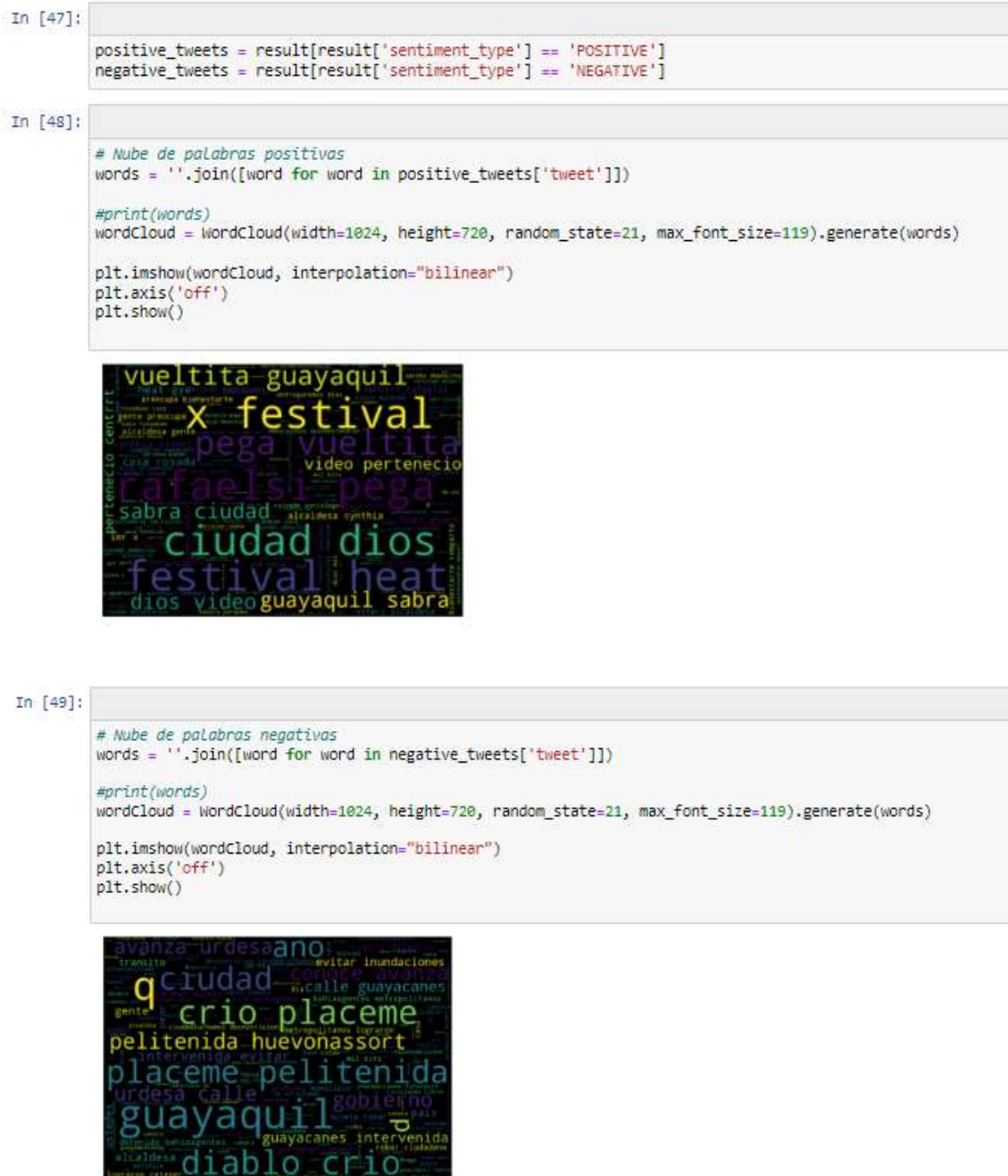


Ilustración 48. Código de programación. Información de Python. Elaborado por Mixi Segura

Y finalmente podemos observar en la Ilustración 49 la clasificación de los tweets, los cuales están manifestados en positivos y negativos. Lo cual nos ayuda a analizar la reputación que tienen cada uno de los precandidatos en la red social Twitter.

result.head(10)

out[50]:

	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count		full_text	tweet	tweet_EN	sentiment_type
#	2022-08-30 04:53:46+00:00	Profesor del Tweet	NaN		NaN	10	False	RT @lahistoriaec: De realities, reinados y TV...	rt realities reinados tv alcaldesa candidatos ...	rt realities reigns tv mayor candidates counci...	NEUTRAL
#	2022-08-30 04:53:38+00:00	Luz Elena Rovira M	Panamá		NaN	6	False	RT @Victori35394538: MELFI X FESTIVAL HEAT GYE...	rt melfi x festival heat gye imr	rt melfi festival heat gye imr	POSITIVE
#	2022-08-30 04:53:16+00:00	ALBA MENDOZA PAREDES	Quito, Ecuador		NaN	17	False	RT @egiscaicedo: Son 10 mil kits de alimentos ...	rt mil kits alimentos entregaremos dias mil ni...	rt thousand food kits deliver thousand chldre...	POSITIVE
#	2022-08-30 04:50:45+00:00	fernando moreno	alausi		NaN	0	False	@SelimF007 @osyalecali @GuilleMoreira98 @Cynth...	corrupto	corrupt	NEUTRAL
#	2022-08-30 04:49:07+00:00	LUIS SALAZAR	NaN		NaN	1	False	RT @CorderoLasso: Toma pero el modelo exitoso ...	rt toma modelo exitoso decian pichincha	rt takes successful model pichincha	POSITIVE
#	2022-08-30 04:46:40+00:00	Jeniffer Palma H	Guayaquil, Ecuador		NaN	1	False	RT @Norah73407026: Las ganas@JosueSanchezEo@Cy...	rt ganasimr x festival heat gye	rt ganasimr festival heat gye	POSITIVE
#	2022-08-30 04:46:08+00:00	Jeniffer Palma H	Guayaquil, Ecuador		NaN	3	False	RT @Norah73407026: Caldo de bola @JosueSanchez...	rt caldo bola imr x festival heat gye	rt broth ball imr festival heat gye	POSITIVE
#	2022-08-30 04:44:54+00:00	Melany 🌟	Riobamba, Ecuador		NaN	6	False	RT @Victori35394538: MELFI X FESTIVAL HEAT GYE...	rt melfi x festival heat gye imr	rt melfi festival heat gye imr	POSITIVE
#	2022-08-30 04:43:56+00:00	Azuena Jimenez	Ecuador- Guayaquil		NaN	0	False	Hasta cuándo? Ya no sólo es a los locales come...	locales comerciales docentes dia dirigimos...	business premises teachers day day run institu...	NEUTRAL

Ilustración 49. Código de programación. Información de Python. Elaborado por Mixi Segura

3.5.5. VISUALIZACIÓN DE DATOS

En esta fase de visualización de datos, utilizamos la herramienta de un Dashboard - Power BI, la cual nos da una presentación mediante varios diagramas dinámicos e interactivos y así tener una mejor interpretación al análisis de la reputación en línea de cada uno de los precandidatos mediante colores representativos.

Al momento de desarrollar la herramienta tomamos en cuenta la cantidad de tweets, mediante un recuadro dinámico el cual nos facilita la visualización del total.

Posteriormente, se añadió categorías en las cuales fueron segmentadas las interacciones de los usuarios, procediendo a presentar un diagrama circular, con los porcentajes tanto positivos, negativos y neutrales.

Finalmente, nos valimos de un segundo diagrama de barra, en el cual denotamos las palabras más frecuentes extraída de los tweets de los electores digitales.



Ilustración 50. Dashboard. Información de Power BI. Elaborado por Mixi Segura



Ilustración 51. Dashboard. Información de Power BI. Elaborado por Mixi Segura

3.6. PRESENTACIÓN DE RESULTADOS

Se presenta como resultado final el análisis de la reputación en línea que tienen los precandidatos: @Candidato amarillo, @Candidato naranja, @Candidato celeste, @Candidato morado, en la red social Twitter. Lo cual fue desarrollado en Jupyter Notebook, con el lenguaje de Python 3 y con la metodología que consta de 5 fases: Selección, Análisis y extracción de datos, Limpieza y validación de datos, Minería de datos e Interpretación.

En este proyecto se realizó un enfoque mixto, en los referentes a lo cuantitativo nos basamos en una cierta cantidad de tweets (2000) y en relación con lo cualitativo, procedimos a validar las diferentes opiniones de los electores digitales sobre la reputación en línea de los precandidatos a la alcaldía de Guayaquil.

Se realizó el análisis de los tweets extraídos de cada uno de los precandidatos, con el análisis de frecuencia pudimos observar las palabras que son más nombradas por los electores digitales, con relación a la imagen que proyecta cada uno de los precandidatos en sus cuentas de Twitter. Luego para llegar a nuestro objetivo, utilizamos el análisis de reputación, que nos permitió clasificar los tweets, en positivos, negativos y neutrales. Aquellos resultados son mostrados en diagrama de barra con la cantidad de tweets tanto positivos, negativos y neutrales y en diagrama circular con los porcentajes de dichas cantidades. Para tener una idea más detallada de los datos positivos realizamos una nube de palabras y así mismo se ejecutó el proceso para los datos negativos.

3.6.1. @ Candidato Amarillo

Para nosotros poder analizar la reputación en línea que tiene el precandidato del color amarillo, fueron extraídos 2001 tweets, los cuales fueron distribuidos en tres categorías como podemos observar en la ilustración 50, estas interacciones fueron validadas por la técnica de minería de datos, la cual nos arroja un 39,98% como reputación positiva, mientras que un 35.03% son neutrales, esto quiere decir que no están ni a favor ni en contra de la imagen que está proyectando el precandidato, finalizando la explicación del diagrama podemos darnos cuenta que un 24,99% de los electores catalogan con una reputación negativa.

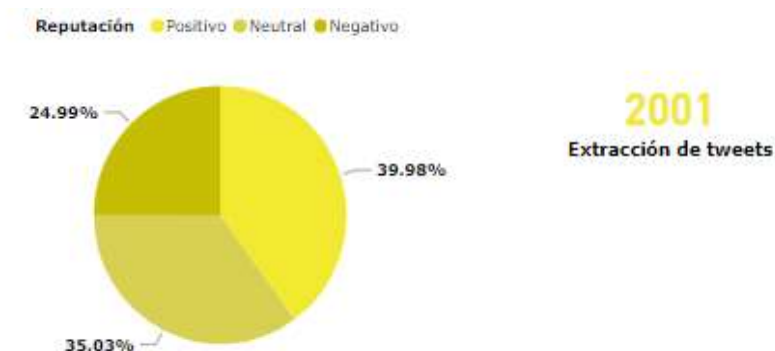


Ilustración 52. Dashboard. Información de Power BI. Elaborado por Mixi Segura

En la siguiente ilustración 51, visualizamos las palabras más frecuentes que fueron mencionadas en los tweets, comentarios. Las cuales los electores digitales postearon referente al precandidato y que fueron seleccionadas al momento de realizar el análisis de frecuencia.



Ilustración 53. Dashboard. Información de Power BI. Elaborado por Mixi Segura

3.6.2. @Candidato Naranja

Analizando la reputación en línea que deriva del precandidato del color naranja, fueron extraídos 1121 tweets, visualizando la ilustración 52, se proyecta un resultado del 45,94% de los tweets referentes a los electores digitales que muestran una respuesta positiva acerca a la reputación, el 44,07% mantiene una opinión neutral, en cambio se refleja una reputación negativa de los usuarios que bordea el 9,99%.

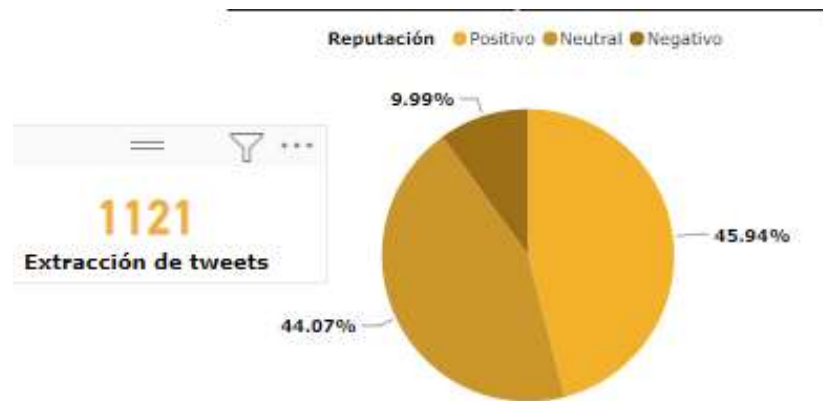


Ilustración 54. Dashboard. Información de Power BI. Elaborado por Mixi Segura

Representando el diagrama de barra extraídos del análisis de frecuencia podemos darnos cuenta de que existe palabras con cantidades similares que utilizan muchos de los internautas referente al precandidato clasificadas como positivas y negativas.



Ilustración 55. Dashboard. Información de Power BI. Elaborado por Mixi Segura

3.6.3. @Candidato Celeste

Como se interpreta en el diagrama de pastel del precandidato del color celeste, en el cual se extrajeron 2000 tweets de la red social Twitter, está clasificado mediante una proporción del 33% la cual corresponde a una reputación positiva e identificamos que el 45% de los tweets no se encuentran de acuerdo con la proyección del precandidato, por lo tanto, tiene una reputación negativa, se concluye con el 22% de usuarios que se mantienen indecisos por lo tanto están neutrales en su opinión.

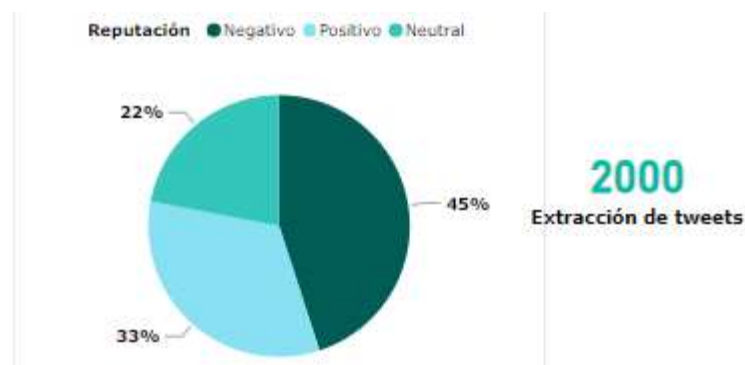


Ilustración 56. Dashboard. Información de Power BI. Elaborado por Mixi Segura

En la ilustración 55, apreciamos con mayor claridad la reiteración de palabras obtenidas mediante el análisis, las cuales se tomaron para visualizar las opiniones de los electores digitales hacia el precandidato.

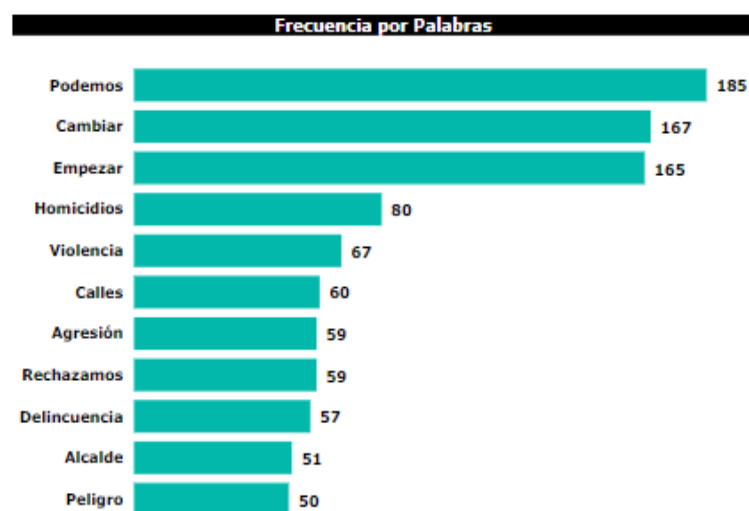


Ilustración 57. Dashboard. Información de Power BI. Elaborado por Mixi Segura

3.6.4. @Candidato Morado

Los datos obtenidos en la extracción de 1428 tweets del precandidato del color morado, nos será útil para presentar el porcentaje positivo del 38.94%, mientras que el porcentaje considerado como negativo en relación con la reputación es de 21.01% y adicionalmente se considera un 40.06% que le es indiferente por lo tanto mantienen una opinión neutral.

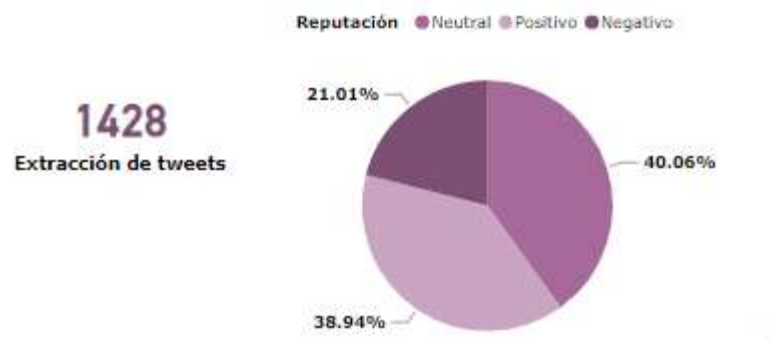


Ilustración 58. Dashboard. Información de Power BI. Elaborado por Mixi Segura

Es evidente mediante el análisis de la gráfica que continuamente se utilizan palabras repetitivas en los tweets que nos sirven para medir un nivel de aceptación o rechazo hacia la reputación del precandidato.

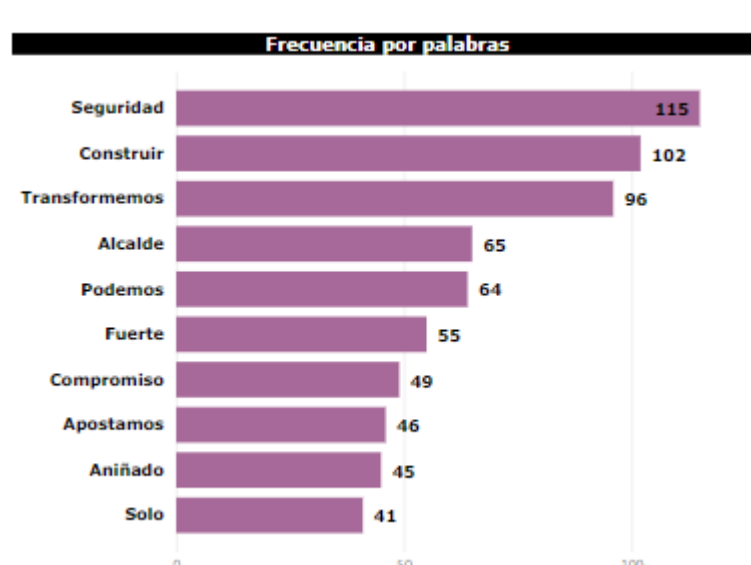


Ilustración 59. Dashboard. Información de Power BI. Elaborado por Mixi Segura.

CONCLUSIONES

- Finalizando con el proyecto de investigación podemos concluir que la red social Twitter es una aplicación masiva de información, la cual al conectarse con el servidor de la API podemos analizar los tweets, comentarios mediante la extracción de datos y darnos cuenta de la reputación que están proyectando los precandidatos en sus cuentas.
- En base a la metodología utilizada podemos concluir que es la más acertada en relación con extracción de datos, la cual nos ayuda a la selección de la herramienta, a la integración, limpieza, procesamiento y a la minería de datos, que se basa en el algoritmo a utilizar, el cual nos permitirá procesar un conjunto de datos para obtener una nueva información sobre ese mismo dataset y por último nos brinda una interpretación sobre los patrones y modelos utilizados.
- Observando los resultados del precandidato del color amarillo, concluiremos que la información extraída de la red social Twitter paso por un proceso de selección, análisis, extracción, limpieza y validación, la cual nos ayuda analizar que el 40% de los electores digitales se manifiestan de forma positiva, mientras que el 35% son neutrales, esto quiere decir que no están ni a favor ni en contra de la imagen que el precandidato está proyectando en su cuenta de Twitter, mientras que el 25% restante se expresan con tweets y comentarios negativos.
- Observando los resultados del precandidato del color naranja, concluiremos que la información extraída de la red social Twitter paso por un proceso de selección, análisis, extracción, limpieza y validación, la cual nos ayuda analizar que el 46% de los electores digitales se manifiestan de forma positiva, mientras que el 44% son neutrales, esto quiere decir que no están ni a favor ni en contra de la imagen que el precandidato está proyectando en su cuenta de Twitter, mientras que el 10% restante se expresan con tweets y comentarios negativos.

- Observando los resultados del precandidato del color celeste, concluiremos que la información extraída de la red social Twitter paso por un proceso de selección, análisis, extracción, limpieza y validación, la cual nos ayuda analizar que el 33% de los electores digitales se manifiestan de forma positiva, mientras que el 22% son neutrales, esto quiere decir que no están ni a favor ni en contra de la imagen que el precandidato está proyectando en su cuenta de Twitter, mientras que el 46% restante se expresan con tweets y comentarios negativos.

- Observando los resultados del precandidato del color morado, concluiremos que la información extraída de la red social Twitter paso por un proceso de selección, análisis, extracción, limpieza y validación, la cual nos ayuda analizar que el 39% de los electores digitales se manifiestan de forma positiva, mientras que el 40% son neutrales, esto quiere decir que no están ni a favor ni en contra de la imagen que el precandidato está proyectando en su cuenta de Twitter, mientras que el 21% restante se expresan con tweets y comentarios negativos.

RECOMENDACIONES

- Se recomienda que antes de elegir la cuenta a la cual le van a realizar la extracción de datos, deben de llevar a cabo un estudio previo, monitoreando inicialmente la que tenga mayor interacción con los electores digitales, para abarcar una mayor cantidad de tweets y que los resultados sean más confiables.
- Recomendamos que al momento de la creación del developer guardemos las contraseñas que nos brinda la API, las cuales son Consumer_Key, Consumer_Secrets, Access_Token, Acces_Token_Secret, que al momento de programar en Jupyter Notebook con Python debemos de ingresarlas para poder tener conexión con la red social y poder extraer la información.
- Antes de trabajar en Python, se recomienda conocer con que versión vas a realizar el proyecto y cuáles son las librerías compatibles, ya que cada versión se trabaja con algunas librerías diferentes.
- Recomiendo trabajar con la herramienta Power Bi para una mejor visualización al momento de proyectar los resultados finales, ya que da una interfaz con paneles interactivos, lo cual hace más entendible los datos ingresados.

ANEXOS

ANEXOS 1

CREACIÓN DE CUENTA EN DEVELOPER – DESARROLLADO DE TWITTER

1. En el navegador se ingresa al siguiente link: <https://developer.twitter.com/en>



2. Hacer clic en iniciar sesión, previo a esto nosotros debemos de tener ya una cuenta en Twitter, llenamos la información requerida.

Developer Portal

Docs Community Updates Support

m*****@chirp****.com

Continue

What country are you based in?

Ecuador

What's your use case?

We need this information for data protection. [Learn more](#)

Student

Will you make Twitter content or derived information available to a government entity or a government affiliated entity?

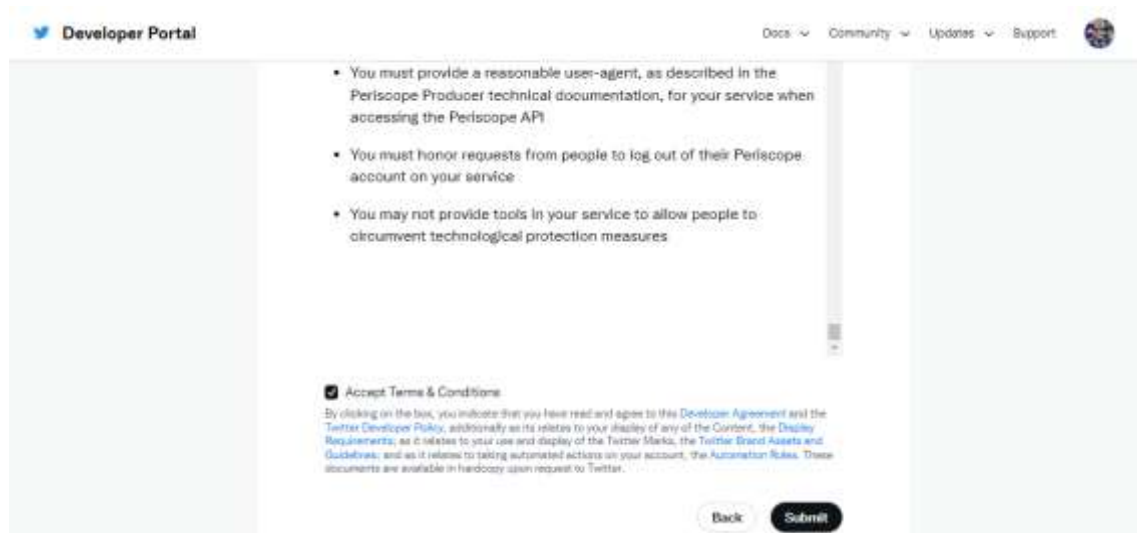
No

Get involved and updated

☐ Sign me up for the latest Twitter Dev news, tips and updates

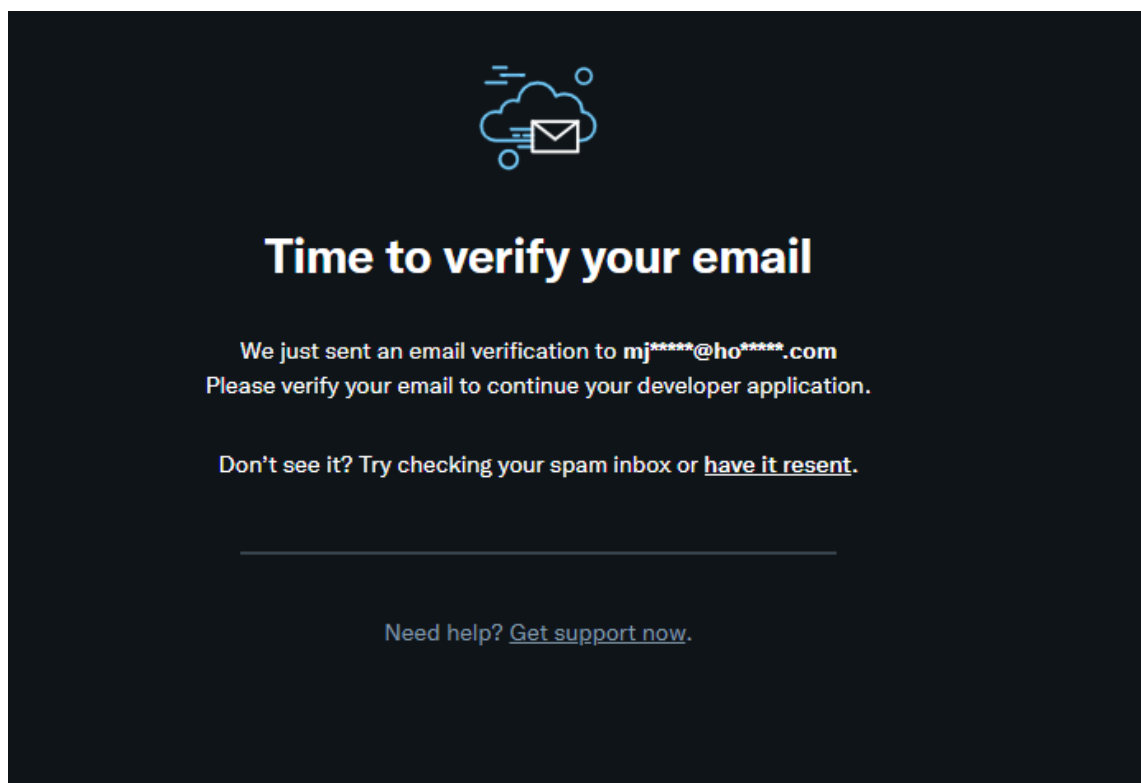
Let's do this

3. Luego aceptamos los términos, condiciones y hacemos clic en Submit.



The screenshot shows the Twitter Developer Portal interface. At the top, there's a navigation bar with 'Developer Portal' on the left and 'Docs', 'Community', 'Updates', and 'Support' on the right. The main content area displays three bullet points regarding API usage: providing a reasonable user-agent, honoring log out requests, and not providing tools to circumvent protection measures. Below these is a section titled 'Accept Terms & Conditions' with a checkbox that is checked. A small paragraph of legal text follows. At the bottom right, there are 'Back' and 'Submit' buttons.

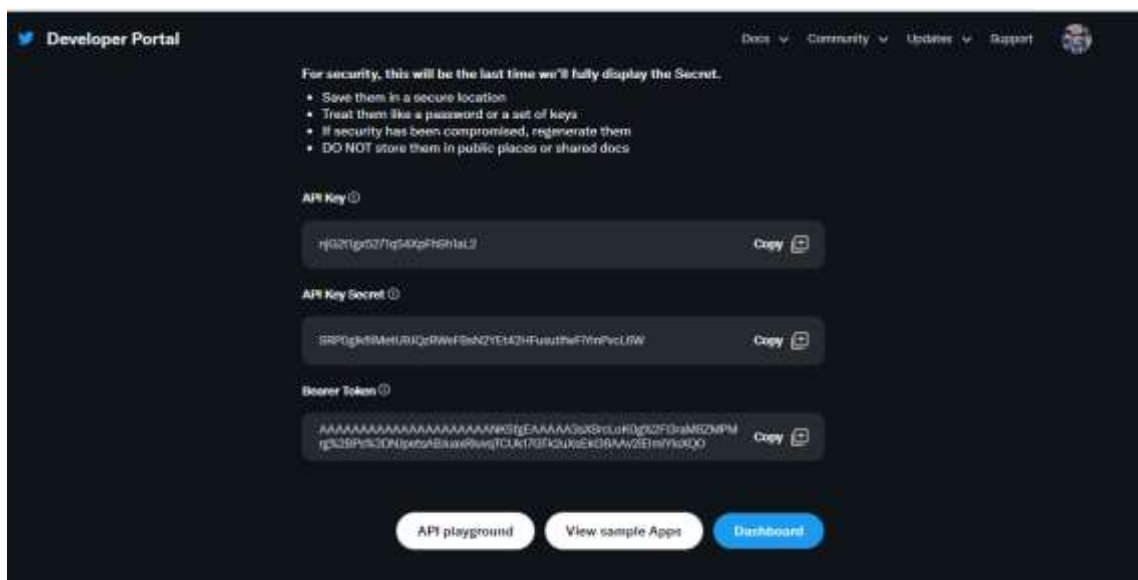
4. Nos mandaran un email al correo que pusimos, debemos de abrirlo y dar confirmación para que nos aprueben la cuenta de desarrollador.



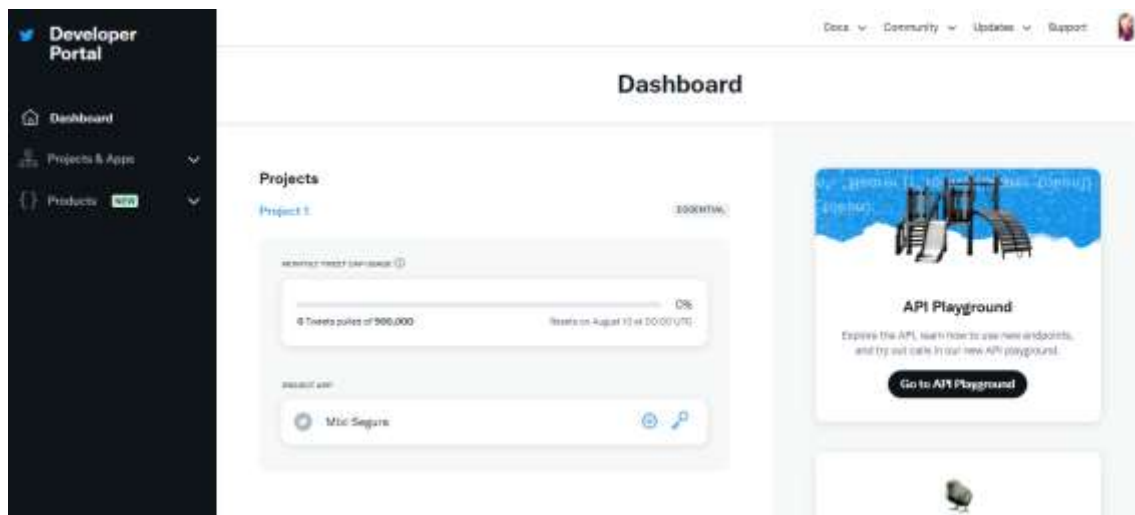
5. Al momento que confirmas tu correo, te dan la bienvenida a la plataforma de desarrolladores en Twitter y debemos de escribir el nombre de usuario y hacemos clic en Get Keys.



6. Al momento que hicimos clic en Get Keys, se nos generó las contraseñas, las cuales serán importantes al momento de la extracción de tweets.



7. Ya tendremos la cuenta creada en developer.



ANEXOS 2

JUPYTER NOTEBOOK – CÓDIGO DE PROGRAMACIÓN

- Declaración de las librerías

```
In [2]: # Librerías

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')
```

- Conexión del servidor de la Api con la red social Twitter

```
In [3]: #contraseñas

credentials = pd.read_csv('./Credentials/Credentialsstudent.csv')
consumerKey = credentials['key'][0]
consumerSecret = credentials['key'][1]
accessToken = credentials['key'][2]
accessTokenSecret = credentials['key'][3]
```

```
In [4]: # 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)
```

- Análisis de los datos

```
In [4]: # Get Tweets - Test
for status in tweepy.Cursor(api_twitter.search_tweets, q="@jimmyjairala").items(700):
    print(status.text)
```

@amontalvoh en el ministerio sale asi pero en la cenescyt no consta por favor ayuda 🙏🙏 hemos invertido tiempo dine... <https://t.co/KkXKqCEILG>

@UnCafeConJJ @jimmyjairala @SolcaGuayaquil @IESSec Buena noticia.

@FernandoCeder @Guillermolasso @BancadaUNESec @OrlandoPerezEC @telesurtv @UnCafeConJJ @jimmyjairala @fabriciovelav... <https://t.co/27XJ3WV69p>

RT @FOSPEecu: @periodicodeuna @panchojimenezs @LassoGuillermo @LorenzoCalvas @GomezRichard @paultutiven @jimmyjairala @periodicodeuna @panchojimenezs @LassoGuillermo @LorenzoCalvas @GomezRichard @paultutiven @jimmyjairala @AsambleaEcuador @FiscaliaEcuador @DianaSalazarM2 @PGEEcuador @VsaquicelaE @CONAIE_Ecuador @PKNacional18... <https://t.co/UdovOp6a20>

RT @Kevinchvarria: @FernandoCeder @Guillermolasso @BancadaUNESec @OrlandoPerezEC @telesurtv @UnCafeConJJ @jimmyjairala @fabriciovelav @Rica...

RT @JorgeAndresMC: @FernandoCeder @Guillermolasso @BancadaUNESec @OrlandoPerezEC @telesurtv @UnCafeConJJ @jimmyjairala @fabriciovelav @Ric...

RT @toaniel: @MonicaBritorC @FernandoCeder @Guillermolasso @BancadaUNESec @OrlandoPerezEC @telesurtv @UnCafeConJJ @jimmyjairala @fabriciovelav @Ric...

RT @AromamiaPerfum: @FernandoCeder @Guillermolasso @BancadaUNESec @OrlandoPerezEC @telesurtv @UnCafeConJJ @jimmyjairala @fabriciovelav @Ric...

RT @toaniel: @tefyJoys @FernandoCeder @Guillermolasso @BancadaUNESec @OrlandoPerezEC @telesurtv @UnCafeConJJ @jimmyjairala @fabriciovelav @Ric...

- Creación de la ruta que se va a guardar el archivo CSV y extracción de los tweets.

```
In [6]: #Create a file with Tweets

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

#Fill rows csv
with open ('./DataSets/Jimmy/Jimmy.csv','w',newline= '\n',encoding="utf-8") as csvwriter:
    writer = csv.writer(csvwriter,delimiter= ",")
    for status in tweepy.Cursor(api_twitter.search_tweets, q="@jimmyjairala", 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.full_text.replace("\n","")])
```

- Limpieza y filtración de los datos

```
In [8]: #Encadenar el texto

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
```



```

In [9]: #Function Stopwords
import unicode

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)

In [10]: #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 Jimmy.csv
dataset['tweet'] = dataset['full_text'].apply(lambda x : removeStopwords(x,'es'))

```

- Análisis de frecuencia de palabras

```

In [35]: #Análisis de frecuencia de palabra

lista_texto= clean_texto.split(" ")
palabras = []

#Paso eliminar palabras muy cortas y muy largas.

for palabra in lista_texto:
    if (len(palabra)>=3 and len(palabra)<18):
        palabras.append(palabra)

#Generalizador de palabras
word_count={}

for palabra in palabras:
    if palabra in word_count.keys():
        word_count[palabra][0]+=1
    else:
        word_count[palabra]=[1]

#Generar DF y lo ordenamos
df=pd.DataFrame.from_dict(word_count).transpose()
df.columns=["freq"]
df.sort_values(["freq"],ascending=False, inplace=True)

df.head(50)

```

Out[35]:

- Traducción de Texto

```
In [13]: # Traducción de 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)
```

- Análisis de reputación

```
In [22]: # BarGraph
result['sentiment_type']=''
result.loc[result.compound>0,'sentiment_type']='POSITIVE'
result.loc[result.compound==0,'sentiment_type']='NEUTRAL'
result.loc[result.compound<0,'sentiment_type']='NEGATIVE'
```

```
In [23]: import math
result.sentiment_type.value_counts().plot(kind='bar',title="Análisis de reputación")
```

- Visualización de los datos en diagrama de pastel

```
In [24]: #Graph Pie
result.sentiment_type.value_counts().plot(kind='pie',
                                             subplots = True,
                                             title="Análisis de reputación",
                                             autopct='%1.0f%%',
                                             y=None,
                                             legend=True,
                                             labeldistance=1,
                                             figsize=(20, 10))
```

- Nubes de palabras positivas y negativas

```
In [25]: positive_tweets = result[result['sentiment_type'] == 'POSITIVE']
negative_tweets = result[result['sentiment_type'] == 'NEGATIVE']
```

```
In [26]: # Nube de palabras positivas
words = ''.join([word for word in positive_tweets['tweet']])

#print(words)
wordCloud = WordCloud(width=1024, height=720, random_state=21, max_font_size=119).generate(words)

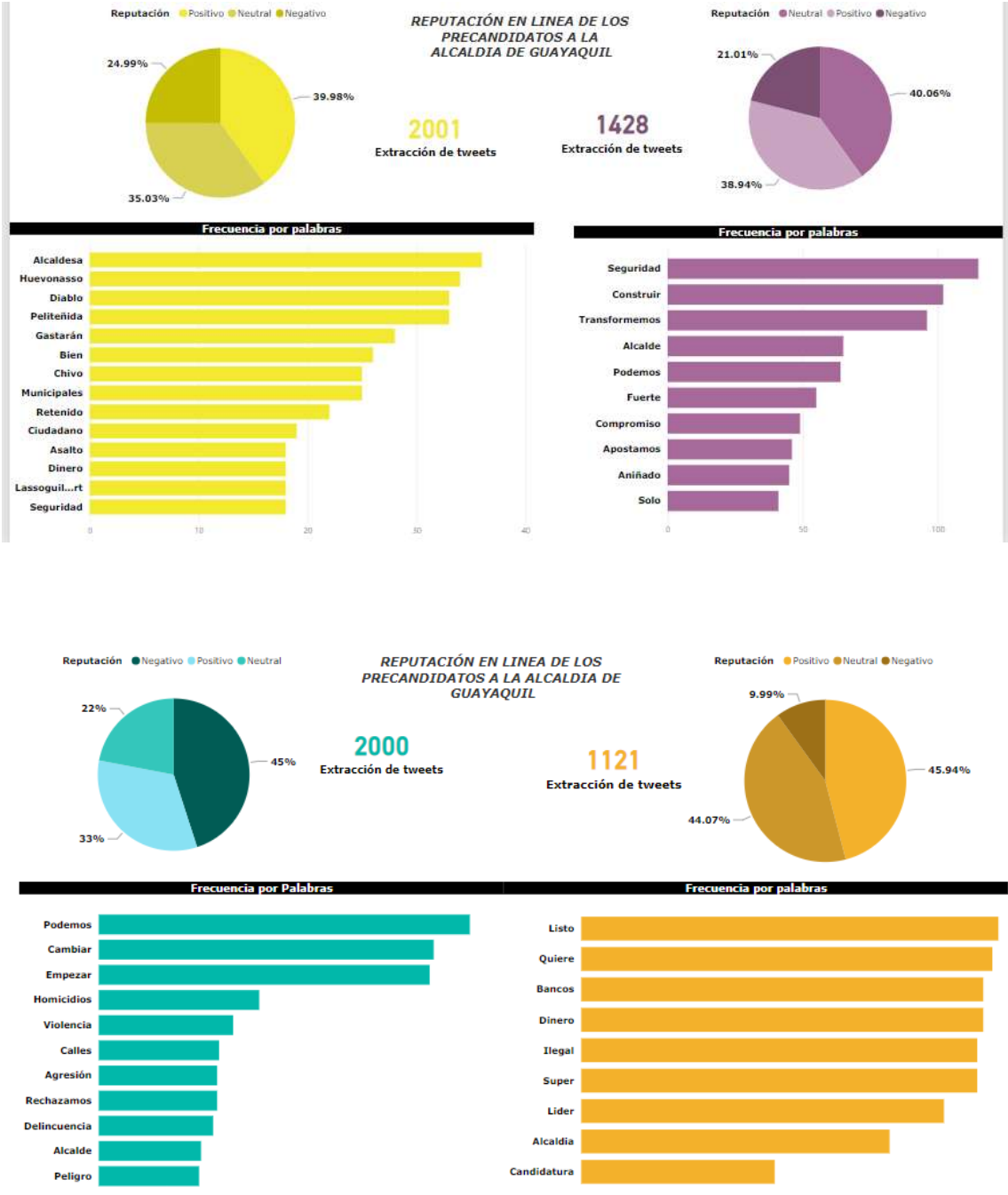
plt.imshow(wordCloud, interpolation="bilinear")
plt.axis('off')
plt.show()
```

- Visualización de los tweets positivos, negativos y neutrales.

In [28]:	result.head(10)									
Out[28]:	Fecha	user.name	user.location	user.following/siguiendo	place	retweet_count	full_text	tweet	tweet_EN	sentiment_type
	2022-08-29 19:35:53+00:00	Brismar Borja	Chimborazo, Ecuador	Place(_api= <twee.py.apl API object at 0x000001B...	0	False	@amontalvoch en el ministerio sale asi pero en...	ministerio sale cenescyt consta favor ayuda in...	ministry cenescyt states invested money fair e...	POSITIVE
	2022-08-29 18:50:33+00:00	Pedro Santos Molina	Guayaquil - Ecuador	NaN	0	False	@UnCafeConJJ @jimmyjairala @SolcaGuayaquil @IE...	noticia	news	NEUTRAL
	2022-08-29 18:20:00+00:00	Benjamin Fernández	Guayaquil, Ecuador	NaN	0	False	@FernandoCedeR @Guillermolasso @BancadaUNESec ...	vagos	bums	NEUTRAL
	2022-08-29 17:43:53+00:00	FRENTE DE ORGANIZACIONES SOCIALES DEL ECUADOR	Ecuador	NaN	1	False	RT @FOSPEecu: @periodicodeuna @panchojimenez ...	rt	rt	NEUTRAL
	2022-08-29 16:23:49+00:00	Orlando V. Fuentes	Ecuador	NaN	0	False	@AsambleaEcuador @FiscaliaEcuador @DianaSalaza...	ecuador	ecuador	NEUTRAL
	2022-08-29 14:42:10+00:00	LUNA  MENGUANTE	Guayaquil, Ecuador	NaN	1	False	RT @kevinchvarria: @FernandoCedeR @Guillermola...	rt	rt	NEUTRAL
	2022-08-29 14:41:54+00:00	LUNA  MENGUANTE	Guayaquil, Ecuador	NaN	3	False	RT @JorgeAndresMC: @FernandoCedeR @Guillermola...	rt jorgeandresmc	rt jorgeandresmc	NEUTRAL
	2022-08-29 14:41:44+00:00	LUNA  MENGUANTE	Guayaquil, Ecuador	NaN	1	False	RT @toaniel: @MonicaSintoRC @FernandoCedeR @Gu...	rt	rt	NEUTRAL
	2022-08-29 14:41:30+00:00	LUNA  MENGUANTE	Guayaquil, Ecuador	NaN	1	False	RT @AromamiaPerfum: @FernandoCedeR @Guillermola...	rt	rt	NEUTRAL
	2022-08-29 14:41:06+00:00	LUNA  MENGUANTE	Guayaquil, Ecuador	NaN	0	False	RT @toaniel: @tefyJoys @FernandoCedeR @Guiller...	rt	rt	NEUTRAL

ANEXO 3

VISUALIZACIÓN DE LOS RESULTADOS EN POWER BI



ANEXO 4

LEY ORGÁNICA DEL USO RESPONSABLE DE REDES SOCIALES

15



ASAMBLEA NACIONAL
REPÚBLICA DEL ECUADOR
2017-2021



ASAMBLEA PROVINCIAL DE MANABÍ
REPÚBLICA DEL ECUADOR
2017-2021

Oficio No. AN-DIMA-022
Quito, 25 de febrero de 2019

Señora economista
Elizabeth Cabezas Guerrero
PRESIDENTA
ASAMBLEA NACIONAL DEL ECUADOR
En su despacho. -

De mi consideración:

Por medio del presente y fundamentado en lo dispuesto en los artículos 134 numeral 1 y 136 de la Constitución y los artículos 54, 55 y 56 del Código Orgánico de la Función Legislativa, me permito remitir el Proyecto de Ley: **LEY ORGÁNICA DEL USO RESPONSABLE DE REDES SOCIALES**.

En tal virtud y por su intermedio se correrá traslado con el mismo al Consejo de Administración Legislativa, CAL., a efectos de que se continúe con el procedimiento que corresponde.

Para fines de índole legal acompañan al presente el formulario de respaldo de las firmas necesarias de conformidad con lo dispuesto en la ley.

Anticipo mi agradecimiento por su favorable atención que se dé a mi solicitud.

Con sentimientos de respeto y estima.



Daniel Isaac Mendoza Arévalo
ASAMBLEISTA POR MANABÍ
ASAMBLEA NACIONAL DEL ECUADOR

Oficio a Fygr
Anexo 12 Fygr

BIBLIOGRAFÍA

- Alcázar, J. (2017). *RANKING REDES SOCIALES, SITIOS WEB Y APLICACIONES MÓVILES ECUADOR*. Obtenido de <https://blog.formaciongerencial.com/ranking-redes-sociales-sitios-web-aplicaciones-moviles-ecuador-2017/>
- Ana, Z. (2021). *Técnicas de investigación*. Obtenido de <https://www.diferenciador.com/tecnicas-de-investigacion/>
- arimetrics. (2022). *Comportamiento Online*. Obtenido de <https://www.arimetrics.com/glosario-digital/comportamiento-online>
- Atribus. (2022). *Minería de datos de redes sociales*. Obtenido de [https://atribus.com/social-media-data-mining/#:~:text=La%20miner%C3%ADa%20de%20datos%20de%20redes%20sociales%20recopila%20varios%20tipos,%2C%20ETC.%20.%20\)](https://atribus.com/social-media-data-mining/#:~:text=La%20miner%C3%ADa%20de%20datos%20de%20redes%20sociales%20recopila%20varios%20tipos,%2C%20ETC.%20.%20))
- Blas, C. (2018). *¿Qué es Instagram?* Obtenido de <https://webescuela.com/que-es-instagram-para-que-sirve/>
- Cañarte, J. (2021). *Construcción de la reputación online de Sebastian Piñera durante el periodo 2019-2021 en la plataforma de microblogging twitter*. Obtenido de <http://dspace.casagrande.edu.ec:8080/bitstream/ucasagrande/3210/1/Tesis3281CA%c3%91c.pdf>
- Centribal. (2022). *Instagram API: la red social se adapta a tu negocio*. Obtenido de <https://centribal.com/es/instagram-api/>
- Coelho, F. (2019). *Qué es Metodología*. Obtenido de <https://www.significados.com/metodologia/>
- Delgado, P. (2020). *Ciudadanos digitales*. Obtenido de <https://observatorio.tec.mx/edu-news/ciudadania-digital-pandemia>
- Deloitte. (2022). *¿Qué es Power BI?* Obtenido de <https://www2.deloitte.com/es/es/pages/technology/articles/que-es-power-bi.html>
- Developer. (2022). *Twitter Desarrolladores*. Obtenido de <https://developer.twitter.com/es/developer-terms/policy>
- Enrique, R. (2020). *Tipos de investigación*. Obtenido de <https://economipedia.com/definiciones/tipos-de-investigacion.html>
- Evotic. (2022). *Ventajas Power BI*. Obtenido de <https://evotic.es/business-intelligence-bi/que-es-power-bi-ventajas-y-tipos-de-licencias/>
- GCF. (2022). *¿Qué es Twitter y para qué sirve?* Obtenido de <https://edu.gcfglobal.org/es/como-usar-twitter/que-es-twitter-y-para-que-sirve/1/>
- Gonzalez, E. (2021). *Facebook: todo sobre la red social más usada en el mundo*. Obtenido de <https://rockcontent.com/es/blog/facebook/>
- Herran, S. (2021). *Análisis de los recursos de odio hacia la población migrante en las redes sociales*. Obtenido de <https://uvadoc.uva.es/bitstream/handle/10324/49440/TFG-G5104.pdf?sequence=1&isAllowed=y>

- IBM. (2021). *Metodología CRISP-DM*. Obtenido de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- IEBS. (2022). *¿Qué es el minado de Datos o Data Mining?* Obtenido de <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/>
- Kyocera. (2021). *Dashboard y su significado estratégico*. Obtenido de <https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/business-challenges/procesos/dashboard-y-su-significado-estrategico.html>
- Leon, E. (2021). *Módulo Minería de Datos*. Obtenido de https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf
- Lobillo, G., & Gallart, R. (2020). *Análisis de la crisis reputacional universitaria a través de las redes sociales*. Obtenido de https://repositorioinstitucional.ceu.es/bitstream/10637/10838/1/es_m10_stamped.pdf
- Mitchell, J. (2022). *¿Qué es API de Facebook?* Obtenido de https://techlandia.com/api-facebook-info_503379/
- Navarrete, J. (2017). *Desarrollo de una aplicacion de mineria de datos que permita detectar preferencias y patrones de texto en la red social twitter*. Obtenido de <https://www.dspace.espol.edu.ec/retrieve/128822/D-106431.pdf>
- Olarte, E., Panizzi, M., & Bertone, R. (2018). *Segmentacion de mercado usando tecnica de mineria de datos en redes sociales*. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/73226/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Ordoñez, K. (2018). *Aplicación de técnicas de minería de datos*. Obtenido de <https://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez%20Brice%C3%B1o%20Karla-%20Informatica.pdf>
- Ortega, O. (2018). *Reputacion en linea*. Obtenido de <https://revista.seguridad.unam.mx/numero-16/reputaci%c3%b3n-en-l%c3%adnea>
- Roales, N. (2019). *Detección de tendencias en twitter utilizando minería de datos adaptiva*. Obtenido de https://repositorio.uam.es/bitstream/handle/10486/662510/roales_gonzalez_natalia_tfg.pdf?sequence=1
- Rodriguez, Y., & Diaz, A. (2019). *Herramientas de Minería de Datos*. Obtenido de <https://www.redalyc.org/pdf/3783/378343637009.pdf>
- SAS. (2022). *Minería de datos*. Obtenido de https://www.sas.com/es_mx/insights/analytics/data-mining.html
- Tristán, E. (2021). *Dashboard*. Obtenido de <https://tristanlosegui.com/2014/10/27/que-es-y-para-que-sirve-un-dashboard/>
- TuDashboard. (2022). *Dashboard*. Obtenido de <https://tudashboard.com/que-es-un-dashboard/>
- Twitter. (2022). *Información sobre las API de Twitter*. Obtenido de <https://help.twitter.com/es/rules-and-policies/twitter-api>
- Valdez, A. (2018). *PYTHON CON ANACONDA*. Obtenido de <https://www.researchgate.net/profile/Aldo-Valdez->

Alvarado/publication/338519767_PYTHON_CON_ANACONDA/links/5e191be0299bf10bc3a346c1/PYTHON-CON-ANACONDA.pdf

Veigler. (2022). *La importancia del Data Mining*. Obtenido de <https://veigler.com/data-mining/>

Villacis, S. (2021). *Reputacion online*. Obtenido de <http://dspace.casagrande.edu.ec:8080/bitstream/ucasagrande/3213/1/Tesis3284VILr.pdf>

Viteri, R. (2016). *Analisis de sentimiento en twitter de los socios de un club de futbol*. Obtenido de <http://repositorio.ug.edu.ec/bitstream/redug/11493/1/PTG-B-CISC%20889%20%20VITERI%20ALVARADO%20RICARDO%20JAVIER.pdf>

Yorly, O. (2020). *ANÁLISIS DEL COMPORTAMIENTO DEL CONSUMIDOR DIGITAL EN REDES SOCIALES MEDIANTE TÉCNICA DE MINERÍA DE TEXTO*. Obtenido de <http://repositorio.ug.edu.ec/bitstream/redug/51604/1/ORT%c3%8dZ%20MOR%c3%81N%20YORLY%20MELISSA.pdf>