UNIVERSIDAD CATÓLICA DE CUENCA

UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

UNIDAD ACADÉMICA DE TECNOLOGÍAS DE LA INFORMACION Y COMUNICACIÓN

CARRERA DE SISTEMAS

ANÁLISIS DE SENTIMIENTOS EN COMUNIDADES DIGITALES UTILIZANDO TÉCNICAS DE BIG DATA PARA DETERMINAR PATRONES DE COMPORTAMIENTO ORIENTADO A FENÓMENOS SOCIALES

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMA

AUTOR: WILLAM ALFREDO LLIVICOTA BUÑAY.

DIRECTOR: ING. CRISTINA MARIUXI FLORES URGILES, MSC.

CAÑAR – ECUADOR

2021

DIOS, PATRIA, CULTURA Y DESARROLLO

UNIVERSIDAD CATÓLICA DE CUENCA

Comunidad Educativa al Servicio del Pueblo

UNIDAD ACADÉMICA DE TECNOLOGÍAS DE LA INFORMACION Y COMUNICACIÓN

CARRERA DE SISTEMAS

ANÁLISIS DE SENTIMIENTOS EN COMUNIDADES DIGITALES UTILIZANDO TÉCNICAS DE BIG DATA PARA DETERMINAR PATRONES DE COMPORTAMIENTO ORIENTADO A FENÓMENOS SOCIALES

TRABAJO DE TITULACIÓN DE PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMA

AUTOR: WILLAM ALFREDO LLIVICOTA BUÑAY.

DIRECTOR: ING. CRISTINA MARIUXI FLORES URGILES, MSC.

CAÑAR – ECUADOR

2021

DIOS, PATRIA, CULTURA Y DESARROLLO

AGRADECIMIENTO

Agradezco a Dios por haberme brindado herramientas necesarias para culminar mi tesis lo cual muestra que con trabajo y dedicación todo se puede lograr.

Un agradecimiento especial a mis padres Gabriel Llivicota Pinguil y María Mercedes Buñay Bermeo por haberme formado como persona y profesional.

A mi esposa Jhenny Fernanda Aguayza Piña y a mi hijo Mathias Gabriel Llivicota Aguayza quienes son mi principal fuente de inspiración y responsabilidad para mi formación.

A los docentes de Carrera de Ingeniería de Sistemas en especial al Ing. Jhovany Santacruz director de Carrera.

A la Ing Cristina Flores directora de trabajo de titulación por el tiempo asignado a mi persona por colaborar en el desarrollo del presente trabajo

DEDICATORIA

A Dios por haberme bendecido día a día con su sabiduría y haberme permitido culminar esta investigación.

A mis padres Gabriel Llivicota Pinguil, María Mercedes Buñay Bermeo, pilares fundamentales en mi vida a mi esposa Jhenny Fernanda Aguayza Piña por sus consejos quienes han sido mi motivación para cumplir cada una de mis metas y a mi hijo Mathias Gabriel Llivicota Aguayza.

DECLARACION

Yo, Willan Alfredo Llivicota Buñay, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Católica de Cuenca Extensión Cañar puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y la Normativa actual de la institución.

Llivicota Buñay Willan Alfredo

C.I: 0302907068

RESPONSABILIDAD

"La responsabilidad del contenido de esta tesis de grado, me corresponde exclusivamente; y el patrimonio intelectual de la misma a la Universidad Católica de Cuenca Extensión Cañar".

Llivicota Buñay Willan Alfredo

C.I: 0302907068

CERTIFICADO DE APROBACION DE TRABAJO DE TITULACION

Yo Cristhiana Mariuxi Flores Urgilés portador(a) de la cédula de ciudadanía Nº 030163837-5. En calidad de tutor certifico que el estudiante, Sr. José Enrique Patiño Guaraca, ha concluido su trabajo de titulación que lleva por nombre "ANÁLISIS DE SENTIMIENTOS EN COMUNIDADES DIGITALES UTILIZANDO TÉCNICAS DE BIG DATA PARA DETERMINAR PATRONES DE COMPORTAMIENTO ORIENTADO A FENÓMENOS SOCIALES".

El trabajo realizado a obtenido la nota de cuarenta y nueve puntos sobre cincuenta (46/50)Aprovecho la ocasión para reiterarle éxitos en el desempeño de sus funciones.

Cañar, 15 de octubre de 2021

F: ……..

Cristhina Mariuxi Flores Urgilés

C.I. 0302090535

DIRECTOR DEL TRABAJO INVESTIGATIVO

UNIVERSIDAD CATÓLICA DE CUENCA EXTENSION CAÑAR

APROBACION DE TRIBUNAL DE GRADO

El tribunal designado por el honorable consejo directivo de la Universidad Católica de Cuenca Extensión Cañar, Facultad de Ingeniería de Sistemas instalado para receptar la sustentación del trabajo final de investigación con el tema: "ANÁLISIS DE SENTIMIENTOS EN COMUNIDADES DIGITALES UTILIZANDO TÉCNICAS DE BIG DATA PARA DETERMINAR PATRONES DE COMPORTAMIENTO ORIENTADO A FENÓMENOS SOCIALES", transcurrido el tiempo reglamentario procede a consignar la calificación de (_____/100).

Cañar, _____de_____ del 2020

PRESIDENTE	**DIRECTOR**
DELEGAGO	**SECRETARIA**

CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

El presente proyecto de titulación tiene el propósito de analizar los sentimientos de los usuarios de las diferentes comunidades digitales, con el fin de extraer patrones de comportamiento sobre fenómenos sociales que afectan a la comunidad, determinando la existencia de contenido de temas relevantes en esta época postpandemica como el Covid-19 y la política ,el análisis se ha llevado a cabo mediante las fases de la metodología CRISP-DM, las cuales fueron realizadas en R, con la ejecución de las librerías más destacadas como twitteR, rtweet, tidyverse, knitr, cada una de ellas cumpliendo funciones diferentes. Para el respectivo análisis se ha procedido con la recuperación de los datos provenientes de la plataforma social Twitter, para luego realizar un análisis exploratorio del comportamiento de los datos y determinar el porcentaje de tweets con sentimientos positivos, negativos y neutros de cada tema a ser analizado. Los resultados de esta investigación demuestran que la minería de texto es una disciplina que en su estado actual puede ser útil para la toma de decisiones para compañías e individuos y que sin embargo es susceptible de ser mejorada para el aprovechamiento de la cantidad masiva de opiniones en texto emitidas por los usuarios de los diferentes medios sociales.

***Pablaras claves*:** crisp-dm, sentimientos, comunidades digitales, fenómenos sociales.

ABSTRACT

The prenent research work aims at analyzin the uners feeling of different digital communities, in order to extract behavior patterns of social phenomena that affect the community, determining the existence of revelant topics in this post-pandemic era, such as Covid-19 and politics. The analysis was carried out through the different phases of the CRISP-DM methodology, which were carried out in R, with the execution of the most outstanding libraries such as twitteR, rtweet, tidyverse, knitr each one of them fulfill different funtions. Data were collected from the online social networking Twiter. Next,an exploratory og the behavior data was conducted, in order to determine the percentage of positive, negative, and neutral tweets about each topic to be analyzed. Results show that text mining is a discipline that in its currents state can be useful for decision-making for both companies and individuals, nevertheless, it is susceptible to improvement in order to take advantage of the massive amount of opinios in text issued by users of the different social media sites.

Keywords: CRISP-DM, feeling, digital communities, social phenomena

INTRODUCCION

El mundo tecnológico de los datos va en constante crecimiento, ya que múltiples usuarios tienen acceso a diferentes redes sociales, en donde manifiestan sus sentimientos a través de sus opiniones sobre un determinado tema con comentarios positivos, negativos y neutros.

Es importante describir las opiniones de los usuarios a través de técnicas de minería de datos para analizar los sentimientos mediante patrones de comportamiento con el fin de obtener información que es útil para interactuar con los usuarios de mejor manera. Existen múltiples herramientas que permiten analizar detalladamente o de forma estadística las opiniones generadas. El presente proyecto comprende el análisis de sentimientos en Twitter de tres fenómenos sociales como son el COVID-19, la política en donde el Sr. Lenin Moreno y el Sr. Guillermo Lasso toman relevancia en este ámbito.

A continuación, se hará una breve descripción de los capítulos realizados en el proyecto.

Primer capítulo, trata sobre el marco referencial, en donde se explica el problema de la investigación, objetivo general, específicos, limitaciones y delimitaciones.

Segundo capítulo, trata sobre el marco teórico el cual recolecta toda la información necesaria para realizar un correcto análisis de sentimientos en Twitter mediante la herramienta ***R.***

En el tercer capítulo, se ha identificado la metodología de la investigación, su enfoque y las técnicas e instrumentos de recolección de datos.

El cuarto y último capítulo contiene la elaboración del análisis de sentimientos de diferentes usuarios en los temas de salud y política ecuatoriana, redactando cada proceso en las fases de la metodología CRISP-DM.

CAPITULO I

MARCO REFERENCIAL

1.1 Planteamiento del Problema

El sentimiento es innato en las personas mediante las cuales demuestran un estado anímico hacia hechos o acontecimientos suscitados en el día a día, estos sentimientos son reflejados en las actitudes y acciones de las personas al momento de desarrollar sus actividades diarias. En mucho de los casos en la actualidad debido al gran avance tecnológico los sentimientos se ven reflejados de una manera digital a través de patrones de comportamiento en las diferentes redes sociales.

Durante la última década varios actores de los ámbitos políticos, económicos y sociales utilizan las diferentes redes sociales como técnica de transmisión y análisis, en donde expresan sus ideas hacia la sociedad y estos acontecimientos dan como resultado reacciones en las personas con distintos patrones de comportamiento, estos patrones son importantes en cada uno de los ámbitos de la sociedad, ayudando a la toma de decisiones empresariales, de gobiernos seccionales, gobierno nacional, partidos políticos , etc.

El país en el presente año 2020 se encuentra atravesando una de las crisis sanitarias más preocupantes de la última década, la cual ha desencadenado en una crisis económica, política y social. Esta situación ha generado diferentes opiniones y estados anímicos en las personas, las mismas que son reflejadas a través de las redes sociales desde los diferentes puntos de nuestro país; dichos estados emocionales no puedes ser analizados de una manera sencilla por lo que se propone utilizar herramientas de BIG DATA como recurso que ayude a recopilar los diferentes patrones de comportamiento de la ciudadanía en las redes sociales frente a esta crisis.

La herramienta que sea generada en este estudio, permitirá dar las pautas y la metodología a seguir en otros casos de estudio, y la difusión de los resultados generados podrán aportar a la toma de decisiones a los diferentes actores de los ámbitos políticos, sociales y económicos dentro del país.

1.2. Formulación del Problema

¿Cuáles serán las herramientas o metodología adecuada para la realización del análisis de sentimientos y Big Data?

¿De qué manera ayudara el diseño de la herramienta en el análisis de sentimientos de los usuarios en las redes sociales de acuerdo a un fenómeno social especifico?

¿Cómo se reflejará la información en la herramienta?

¿Cuál es la influencia del comportamiento y sentimiento de los usuarios en comunidades digitales mediante estos fenómenos sociales?

1.3. Antecedentes de la Investigación

1.4. Justificación

El análisis de sentimiento es extremadamente útil para monitoreo las opiniones, actitudes, cometarios de los usuarios en las redes sociales ya que permite hacernos una idea de la opinión público general de ciertos temas.

Los sentimientos del usuario son actividades generalmente automatizadas dentro de las redes sociales, por lo que para identificar estos sentimientos se utiliza herramientas de BIG DATA como recurso que permita monitorear el comportamiento de la sociedad dentro de estas comunidades virtuales; considerando cometarios, opiniones, que manifiestan los usuarios, estos sentimientos son de mucha relevancia, por lo cual pocas personas se expresan en distintos medios del mundo digital.

Existen una gran cantidad de herramientas de BIG DATA que permiten el análisis de un gran volumen de datos, obteniendo datos significativos sobre el comportamiento de los datos, en el caso del análisis de sentimientos permite determinar el tono emocional que hay detrás de una palabras determinadas, si una frase contiene una opinión positiva o negativa sobre una fenómeno social, situación económica, institución, organización, empresa, evento o persona los sentimientos, contarios de los usuarios, siendo beneficioso para desarrollar inteligencia de negocios para toma de decisiones de organizaciones públicas, privadas, personas naturales y personas jurídicas .

Actualmente el mundo está atravesando una crisis sanitaria que está dejando consecuencias socioeconómicas incalculables, el Ecuador no es la excepción y esto ha generado un caos e incertidumbre a nuestra sociedad. Debido al gran avance tecnológico de hoy en día, y la gran cantidad de información que se genera en el Ecuador a diario; la cantidad de datos es abrumadora, por lo que la sociedad manifiesta su comportamiento mediante el mundo digital ante esta riesgosa pandemia.

1.5. Objetivos

1.5.1. **Objetivo General**

Desarrollar un método que sea capaz de analizar los sentimientos de los usuarios de las diferentes comunidades digitales, con el fin de extraer patrones de comportamiento sobre fenómenos sociales que afectan a la comunidad mediante la explotación de técnicas de BIG DATA.

1.5.2. **Objetivo Específicos**

- Realizar un estudio de estado del arte y revisión bibliográfica sobre el tema.
- Determinar la metodología que permita establecer el proceso para el análisis de datos.
- Diseñar una herramienta la cual analice el proceso de difusión de los sentimientos de los usuarios en las redes sociales de acuerdo a un fenómeno social especifico.
- Analizar los resultados obtenidos a través del BIG DATA para conocer el patrón de comportamiento de las personas al utilizar las redes sociales, frente a un fenómeno social especifico.

1.6. Limitaciones

En el avance de este proyecto se puede hallar con inconvenientes que a continuación se especifica:

Pueden estar de acuerdo a la capacidad de procesamiento de los equipos a disposición

Se limitará a los accesos permitidos por las redes sociales

1.7. Delimitaciones

Este diseño se elaborará para la observación de sentimientos y comportamientos de los usuarios en las comunidades digitales, siendo el problema que por mala información a los usuarios de estos fenómenos sociales que viene pasando en la actualidad.

El tiempo para llevar a cabo el desarrollo del proyecto está basado en la planificación realizada por la Universidad Católica de Cuenca, Extensión Cañar.

Para el contenido del proyecto de investigación, está basada en las técnicas de BIG DATA para determinar el comportamiento de los usuarios ante un fenómeno social.

CAPITULO II

2. MARCO TERICO

2.1. Inteligencia de Negocios

Para entender en si para qué sirve la inteligencia de negocios, se han tomado definiciones, en donde se detalla cuáles son sus características.

La inteligencia de negocios hace referencia a varios métodos, aplicaciones, práctica y capacidades centradas a la interpretación de información que permite tomar decisiones a los empleadores de una entidad (Jordi Conesa Caralt, 2010) n.

Según (Libros Científicos, 2015), comenta que:

"Hace referencia a un conjunto de métodos orientados en la administración y creación de conocimiento sobre el medio, mediante un análisis de los datos presentes dentro de una entidad, siendo su objetivo respaldar decisiones empresariales"

Además, se considera como un conjunto de procesos encargados de eliminar ineficiencias que agilizan la elaboración de los datos resultantes de los sistemas de gestión empresarial para el estudio de una mejora en bienestar de la empresa.

2.1.1. Características de la Inteligencia de Negocios

Se considera que está formado por cinco características esenciales, los cuales cumplen diferentes funciones, tareas y metodologías.

Siendo una herramienta esencial que sirve para la planificación empresarial, ya que cuenta con características determinadas que logran que los negocios puedan conseguir información importante con el fin de mejorar su utilidad:

Observación. Se recolectan los datos y se analiza el problema.

Comprensión. Análisis y cruce de datos para obtener información de suma importancia.

Predicción. Estimar mediante los datos un posible resultado.

Colaboración. Dada entre la difusión de los resultados y la colaboración entre los departamentos.

Decisión. Propuesta de una táctica en función del análisis y las simulaciones ejecutadas.

En definitiva, mediante la Inteligencia de Negocios se puede identificar tanto productos como usuarios del mercado, las ventas que producen alta demanda o sectores en los que no se tiene resultados positivos. Ayudando de esta forma a poder crear métodos para el beneficio de una determinada organización a futuro.

2.1.2. Enfoque de implementación de proyectos de BI

Cuando se toma las decisiones de establecer un plan de inteligencia de negocios lo cual sería una serie decisiones de las arquitecturas, diseño y el alcance para ser valoradas con éxito en los que se emplea el "divide y vencerás" sin embargo las organizaciones en el proyecto BI establece un alcance normalmente practico para poner en marcha la solución BI (Cano, 2009)

Los proyectos de inteligencia de negocios son ejecutivos para las organizaciones como proyectos que incluyen exclusivamente en los temas de la nueva era de la tecnología de la información, sin embargo, el componente tecnológico en esto proyectos lo cual es muy importante para el entendimiento fundamental para el negocio para alcanzar el éxito puesta en el funcionamiento del BI, por lo tanto es casi imposible

obtener un previo entendimiento del negocio para la identificación de la información que sirve para la toma de decisiones. (Meneses, 2013)

Los proyectos de Business Intelligence, tienen como objetivo el disponer de un sistema integrado de apoyo para la toma de decisiones que serán de gran ayuda en cuanto a la gestión de riesgo y ayudarán a dar una mayor flexibilidad aceptando cambios de mercado señalando aspectos positivos y negativos. (González Ferran, 2016, págs. 67-77)

A partir de las definiciones se tiene claro que el enfoque para implementar de proyectos BI es el responsable de las dependencias entre actividades en un mayor régimen a la retribución de recursos, técnicas y procesos de elaboración de posibles etapas y fases de organizaciones para que pueda tomas las decisiones técnicas adecuadas.

2.2. Integración de datos ETL

Consiste en el desarrollo que autoriza a empresas a distribuir datos variados desde su origen, a extraerlos, transformarlos y cargarlos en otra base.

"Es un proceso donde se obtiene datos de una fuente, los datos son cambiados y cargados en un almacenamiento," (Mostazo, 1890)

Proceso ETL

Parte de la agregación de los datos, siendo un componente principal que tiene como tarea, el impacto de toda la ejecución del enlace entre las aplicaciones y los sistemas

Es el proceso que se encarga de compilar datos, mediante varias fases:

Tabla 1: Fases del proceso ETL; Autoría:propia.

Extracción	Elaboración de los datos de diferentes orígenes.
Transformación	En esta fase dicho datos, en otras palabras, es la capacidad de reformar y limpiar estos datos cuando se vital.
Carga	En esta parte el desarrollo es más procedente que la fase de transformación, un datamart va con el objetivo de analizar y fundamentar para el proceso del negocio

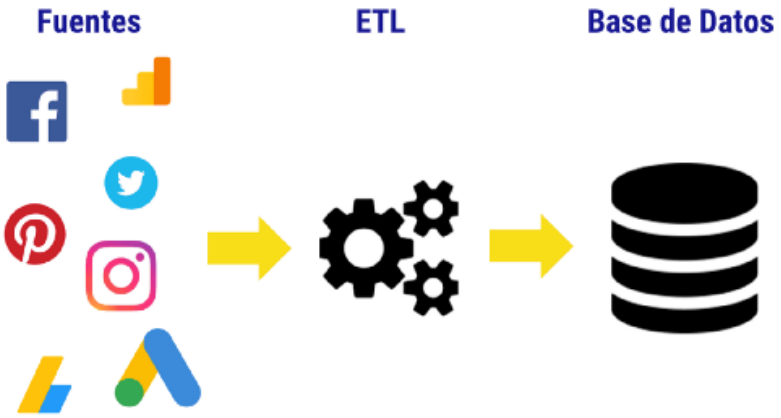

Ilustración 1: Modelo ETL; Fuente: (Gajardo, 2019).

2.3. Big Data

Según (Viktor Mayer-Schonberger, 2013), "se trata de hacer cosas a partir del análisis de inmensas cantidades de información, que simplemente no son posibles con volúmenes más pequeños" (pág. 31)

Un Big Data se puede también referirse como:

"Tratamiento y análisis de grandes repositorios de datos, que resulta imposible manejarlos con las herramientas de bases de datos y analíticas convencionales" (Argonza, 2006, pág. 3).

(Antonio Monleón Esteban Vegas, 2017) manifiesta que actualmente, las grandes compañías de la información, se valen de gran cantidad de información con el objetivo de mejorar las demandas electrónicas de los usuarios, midiendo la información digital en bytes para obtener insights que conlleven a la toma de mejores decisiones y acciones de los negocios aplicando estrategias. (págs. 12-13)

> "El término Big Data, se aplica a conjuntos de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable y por los medios habituales de procesamiento de la información" (MARQUÉS, 2015, págs. 4-8)

A partir de ciertas definiciones, en definitiva, se describe a Big Data como la cualidad de poder utilizar datos para conseguir información y conocimiento sobre el valor para emprender nuestro negocio, por ende, el Big Data tiende al estudio del procedimiento de los usuarios tomando en cuenta sumamente los datos guardados, y así poder formar pronósticos de los patrones examinados.

2.3.1. Características

Según (Enrique Martín, 2020), Big Data, está compuesta por tres principales características:

- **Volumen:** Hace referencia a la cantidad de datos que se recopilan, y que se procesan de forma constante.
- **Velocidad:** La velocidad de los datos al llegar a la web y también la rapidez de utilizarlos en una base de datos.
- **Variedad:** Big Data, es capaz de procesar diferentes datos, de distinta categoría y fuentes, para una determinada organización. (pág. 29)

2.4. Data Warehouse

Según (Mendez, 2010, págs. 19-26) afirma que el Data Warehouse "es una tecnología para el manejo de la información construido sobre la base de optimizar el uso y análisis de la misma utilizado por las organizaciones para adaptarse a los vertiginosos cambios en los mercados".

Según definiciones estudiadas de Data Warehouse es un almacén de datos para el uso fácil de analizar todos los departamentos de las empresas de grandes velocidades de respuestas para una mejor solución fiable de inteligencia de negocios, sin embargo, la información que se puede lograr de una Data Warehouse es información fiable para la toma de decisiones

2.4.1. Arquitectura

Data Warehouse por lo tanto es una colección de datos que proporciona a diferentes temas, su arquitectura está compuesta por varios componentes tales como:

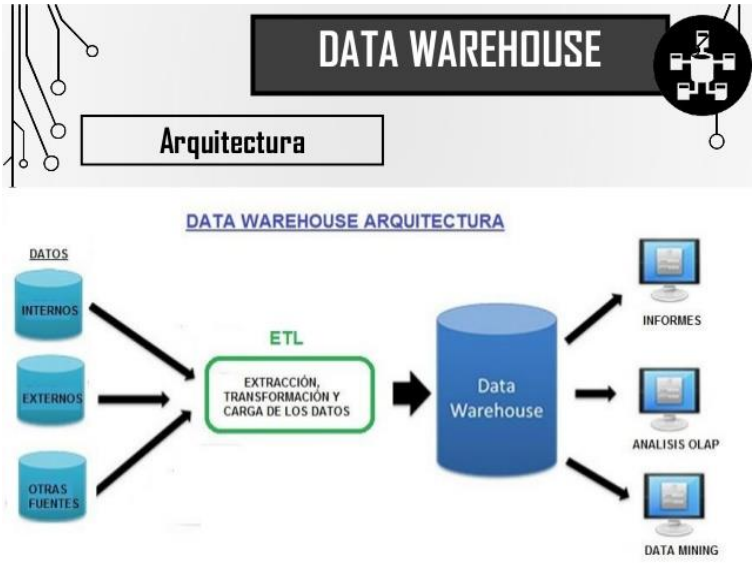

Ilustración 2: Arquitectura de Data Warehouse; Fuente: (WAREHOUSE, 2017)

2.4.1.1. OLTP (On-Line Transaction Processing)

Considerado como un SGBD, que trabaja en tiempo real y eficaz, que permite la realización de comandos como insert, update y delete. Tomando en cuenta que no son ideales para soporte de la toma de decisiones ya que no guardan un histórico de datos. Utilizado para lectura y grabación de datos, no presenta un detallado de los datos, pero tiene como ventaja una mayor velocidad en la respuesta de documentos o retorno (Gálvez, 2016, págs. 74-75).

"Un Sistema OLTP, tiene un modelo de datos de relación y un modelo orientado a la aplicación diseño de base de datos, utilizado para el procesamiento de transacciones" (G.SATYANARAYANA REDDY, 2010, pág. 2)

Conforme a la revisión de las definiciones estudiadas de OLTP, se conforma en el procedimiento lo cual facilita la gestión de aplicaciones transaccionales para el acceso de los datos, recuperación y procedimientos de los paquetes de software empleados para estas tecnologías que son basados en procedimiento cliente-servidor.

2.4.1.2. Middleware

> Según (Sevilla, 2006) considera que: "es el conjunto del software distribuido necesario para el soporte de interacciones entre clientes y servidores" (pág. 209)
>
> "Es la tecnología que realiza el paso intermedio, y que gestiona las comunicaciones con el almacén de datos, coordinando la concurrencia y gestionando los procesos" (Jaime Laviña Orueta, 2010, pág. 177)

Este sistema de software compone aplicaciones y servicios que se ejecutan en dispositivos de comunicaciones, incluye identificación, autenticación, autorización, conmutación por software, certificación y seguridad, en cuanto a comunicaciones este

ensambla e interpreta de forma inmediata aplicaciones de comunicaciones flexibles. (Mahmoud, 2004)

Software encargado de facilitar el desarrollo de programas, está incluido dentro del SSI, por lo que está basado en la arquitectura cliente-servidor que permite independizar la aplicación del entorno hardware de explotación (Senén Barro Ameneiro, 2002, pág. 61)

Según (Commerce, 2009) afirma que Middleware "es el software que conecta o integra componentes de software entre aplicaciones y sistemas distribuidos, que permite la transferencia eficaz de datos entre aplicaciones" (pág. 109)

De acuerdo con las definiciones estudiadas Middleware, se dispone entre un S.O y las aplicaciones que realizan en él, lo cual el software presenta servicios y funciones generales para las aplicaciones. Por lo tanto, Middleware ayuda a los desarrolladores a proyectar aplicaciones con mayor eficacia a través de un hilo de datos y los usuarios.

2.4.1.3. OLAP (On-Line Analytical Process)

Integra el pensamiento analítico en línea junto a la minería de datos con el fin de que se puedan recolectar en distintas bases de datos o niveles de abstracción.

Un artículo realizado por (Larraín, 2021), comenta que:

Esta técnica de procesamiento de información está diseñada en patrones que se encargan de detallar información de grandes volúmenes para poder desarrollar un informe detallado de un determinado tema, basándose en estructuras multidimensionales, realizando su almacenamiento en un vector multidimensional.

(Angelino Feliciano Morales, 2016), manifiesta que:

> OLAP es una tecnología la cual permite el acceso rápido a datos mediante la herramienta Data Warehouse, proporciona una búsqueda inmediata de los datos y de acuerdo a la cantidad de estos, un tiempo determinado, varias compañías utilizan OLAP ya que es considerado como sistema confiable que se encarga del procesamiento de datos que tiene como fin mediante un análisis, la mejora de la entidad que la utiliza (pág. 2)

> Las bases de datos analíticas OLAP son usadas para consulta de grandes cantidades de datos, en marketing, minería de datos, tiene como ventaja la rapidez de respuesta, pero para consultas multitabla es lenta. Las sentencias SQL que se ejecutan en esta base de datos, son de tipo SELECT. (León, 2018, pág. 229)

Este sistema analítico, permite a las empresas que las usan, realizar análisis multidimensional de grandes cantidades de información a través de cubos, cumpliendo con varias reglas establecidas como, el tener una visión multidimensional de los datos, pensar en dimensiones y métricas del negocio (Kamagate, 2013, págs. 1-7).

> A partir de las definiciones tomadas se tiene un claro concepto que OLAP permite extraer fácilmente y de manera selectiva datos y examinarlos desde diferentes perspectivas para agilizar grandes cantidades de datos multidimensionales o sistemas transaccionales.

2.5. Aplicaciones

2.5.1. EIS (Executive Information System)

Son sistemas de Inteligencia empresarial que proporcionan información táctica, que consiste en observar informes de la organización tanto interna como externa de la misma.

2.5.2. DSS (Decission Support System)

DSS, es una herramienta de Inteligencia de Negocios, considerado como un sistema que requiere de una base de datos como fuente para ayudar a la toma de decisiones empresariales, se encarga de tomar información que facilita a los empleadores un fácil entendimiento de las situación interna y externa de la compañía.

2.6. Modelamiento

Es un conjunto de base de datos la forma lógica de la base que determina como se guarda los datos y como se acceded a ellos. Existen tres tipos de diseños de esquemas multidimensionales:

- **Esquema en Estrella:** Este esquema representa una vista de la organización, incluyendo ventas y mercadeo, se encarga de consolidar hechos en relación a dimensiones o filtros, ya que contiene un conjunto de tablas de datos que se relacionan entre sí siendo el centro del esquema la tabla de hechos que relaciona muchas tablas de dimensiones y teniendo una clave principal. (Aguilar)

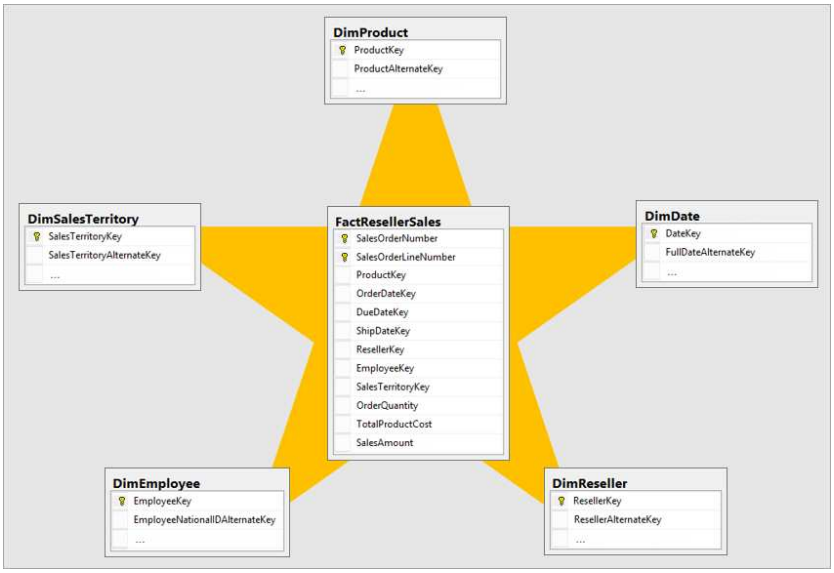

Ilustración 3. Esquema Estrella. Fuente: (Microsoft , 2019)

- **Esquema Copo de Nieve**

Este esquema, contiene puntas del esquema estrella las cuales se pueden dividir en más puntas, las tablas de dimensión tienen relación con otras, tiene como ventaja un menor espacio de almacenamiento, aunque incrementa el número de tablas y su sistema de consulta es complejo (Trujillo, 2006, pág. 4)

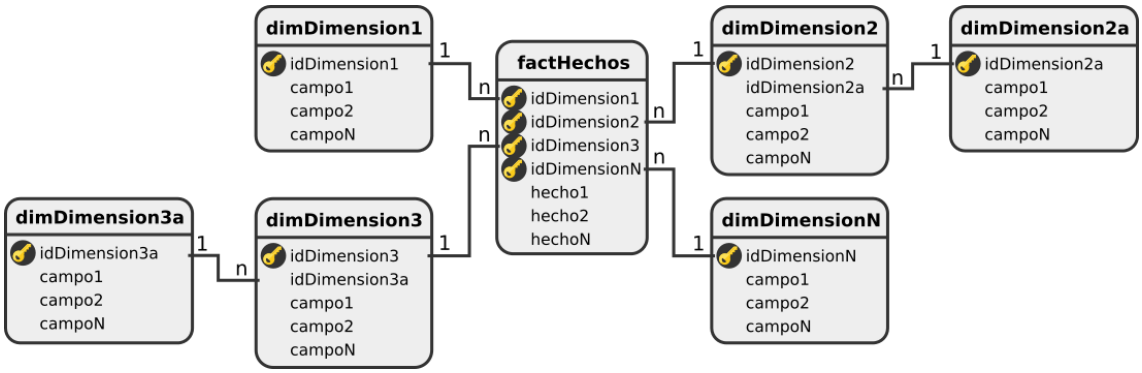

Ilustración 4. Esquema Copo de Nieve. Fuente: (Bernabeu R. Dario, 2021)

Esquema Constelación

"Es un esquema complejo ya que contiene múltiples tablas de hechos, se caracteriza por ser flexible y su utilidad es que al tener dimensiones que pueden ser compartidas por cubos se tiene un mejor uso de espacio de almacenamiento"

(Martínez., 2011, pág. 37)

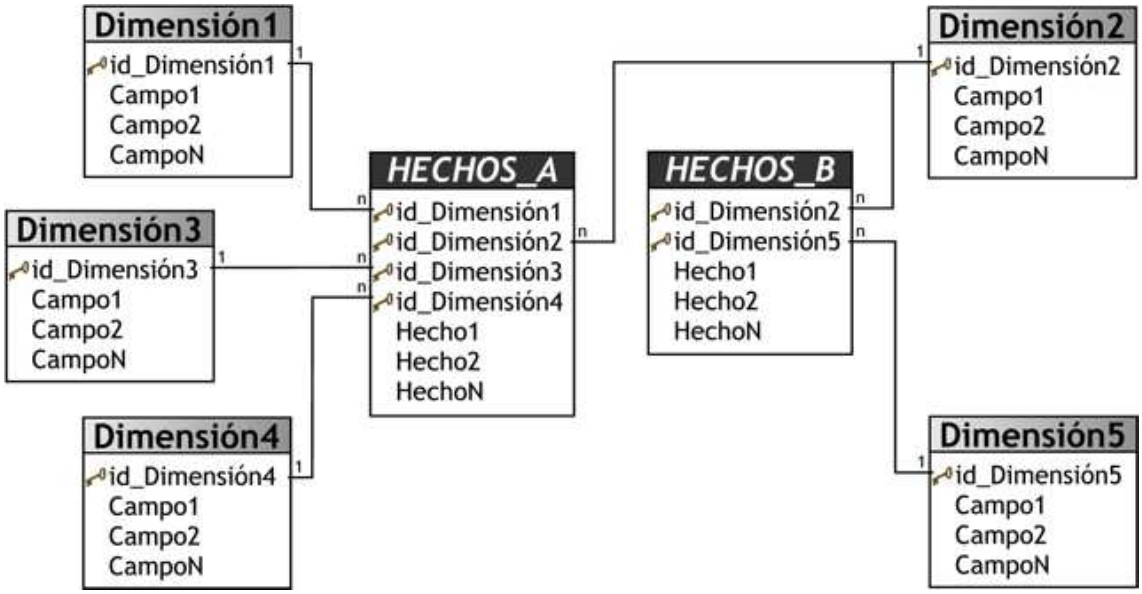

Ilustración 5. Esquema Constelación. Fuente: (Darío, 2006)

2.7. Minería de datos y de texto

2.7.1. Minería de datos

Se define como un proceso para el descubrimiento automático de nuevas técnicas para examinar en grandes bases de datos información que se encuentra almacenada de modo ordenado usando "tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos" (PEREZ LOPEZ, 2007)

Según (José C. Riquelme, 2006), indica que:

"Conforme a la creciente demanda de la tecnología de Internet, se ha visto la necesidad de desarrollar más herramientas de minería de datos para poder interpretar de mejor manera información almacenada en bases de datos" (págs. 13-14)

Una tesina realizada por (Martínez, 2001), comenta que la minería de datos influye en las empresas de una forma contable mejorando sus esquemas de administración, ayudando a la toma de decisiones y a mejorar una organización. Se encarga de descubrir comportamientos de los usuarios mediante patrones con el objetivo de evaluar información implícita (pág. 18).

“La minera de datos puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos” (CESAR PEREZ LOPEZ, 2008, pág. 5)

En base a los conceptos, se tiene claro que la Minería de datos es un conjunto de métodos y tecnologías que facilita explotar inmensos volúmenes de datos, de manera semiautomática, y poder encontrar las tendencias y reglas que exprese el comportamiento de los datos.

2.7.1.1. Proceso de Minería de Datos

La minería de Datos, se encarga de procesar grandes cantidades de datos utilizando patrones y tendencias, consta de un proceso que implica tres fases tales como:

- Selección y adquisición de la información, recopilando los datos y definiendo el método y herramienta, que se aplicará para el procedimiento.
- Preparación y proceso de los datos.
- Interpretación e integración de los resultados.

2.7.2. Knowledge Discovery in Databases (Descubrimiento de conocimiento de base de datos)

“Consiste en el análisis exploratorio y modelado de grandes repositorios de datos e involucra áreas de conocimiento como inteligencia artificial, aprendizaje automático, estadística, sistemas de gestión de base de datos y medios que apoyan a la toma de decisiones” (Lautaro Ramos, 2019, pág. 1)

“Corresponde a la aplicación de algoritmos para hallar patrones ocultos en los datos, utiliza varias técnicas de aprendizaje automático desarrollados en Machine

Learning, basados en herramientas de análisis de datos" (Gabriela Mancilla-Vela, 2020, págs. 2-3)

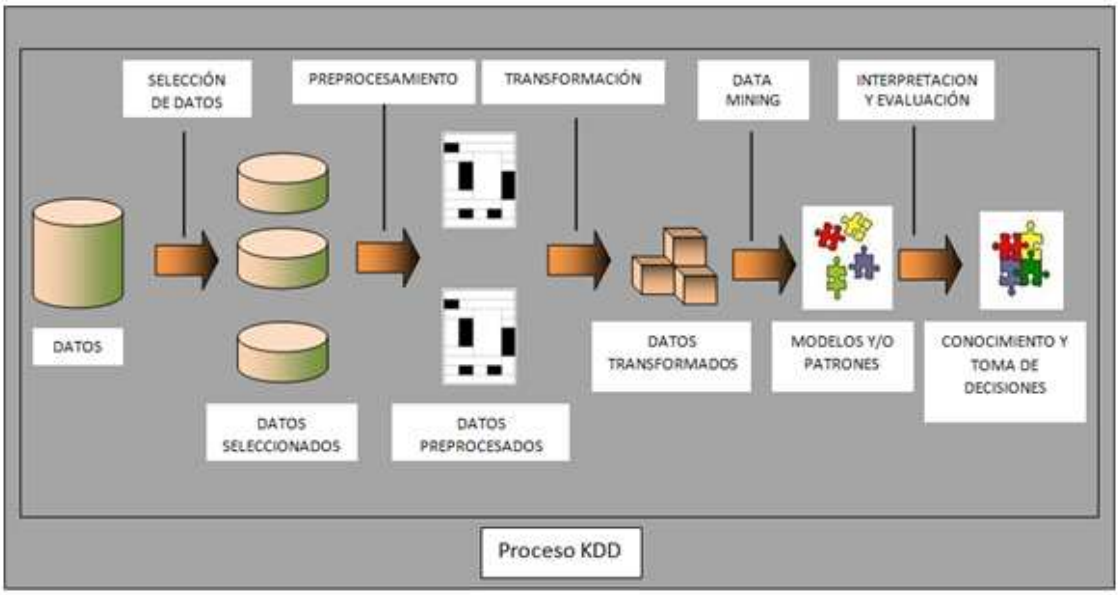

Ilustración 6. Proceso KDD. Fuente: (Cond, 2017)

2.8. Modelos de minería de datos

Las tareas predictivas de la minería de datos, contienen dos modelos que tienen como fin predecir el comportamiento del usuario.

- **Clasificación:** (María Uvidia Fassler, 2018) expone que:

  Este método consiste en identificar particularidades de un objeto determinado, para poder asignarle a una categoría definida de acuerdo a sus características, dentro de este método se encuentran varias técnicas como las redes neuronales, las redes bayesianas, árboles de decisión, máquina de soporte Vectorial y basado en Instancia. (págs. 4-5)

- **Regresión: "**Permite predecir respuestas a partir de un muestreo de datos de forma aleatoria, es usada para pronosticar valores ausentes guiándose en una relación con variables semejantes" (María Uvidia Fassler, 2018, pág. 5)

2.8.1. Técnicas de minerías de datos

Según Paola García García (2008), manifiesta que:

Las técnicas de la minería de datos proceden de la inteligencia artificial y también de la estadística, no son nada más que algoritmos diseñados para lograr resultados de un conjunto de datos, la minería de datos aplicada en redes sociales, trabaja conjuntamente con técnicas que sobresalen en este ámbito como son:

- **Redes neuronales**: este algoritmo interconecta neuronas en una red que produce un estímulo de salida, tienen procesamiento automático.
- **Árboles de decisión**: Se trata de un modelo de predicción usado en machine learning, que, mediante una base de datos, estos algoritmos representan y categorizan condiciones que suceden de forma sucesiva para poder resolver un determinado problema. (pág. 2)
- **Clustering:** "Procedimiento de agrupación de una serie de vectores según criterios a distancia, seleccionando los vectores de entrada, más cercanos y que contienen las mismas características" (pág. 3)

2.8.2. Algoritmos para minería de datos

- **Algoritmo de árboles de decisión:** Es empleado dentro de un conjunto de datos que está diseñado para la predecir resultados, utilizando valores relevantes para luego poder relacionar mediante métodos.

- **Algoritmo Naive Bayes:** Es un algoritmo de clasificación, es utilizado en minería de datos ya que detecta las relaciones entre columnas de entrada y también las de predicción, facilita una buena precisión de la clasificación en tiempo real dentro de un conjunto de datos. (Microsoft, 2018)

- **Algoritmo Multinomial Naive Bayes:** Consiste en una versión especializada del algoritmo anterior, que está definido para contener información de frecuencia de palabras dentro de un determinado tema, dentro de una secuencia ordenada de palabras.

2.8.3. Aplicaciones de minerías de datos

Se considera que la minería de datos, está formado por aplicaciones, los cuales cumplen diferentes funciones y tareas, además (Maria Consuelo Justicia de la Torre, 2017) los describe en metas y aplicaciones.

- **Mejorar las capacidades de marketing** esto se enfoca a la orientación de campañas hacia determinados clientes.
- **Detectar patrones anormales** se encarga de la prevención de fraudes.
- **En la industria** se refiere al control de producción y logística.
- **En el ámbito bancario** se enfoca al estudio de análisis de riesgo, créditos, prevención de morosos.
- **En gestión de BDD** dirigidos a la ingeniería inversa, mejorar la calidad de los datos, mejorar las consultas de descubrimiento dependencias funcionales. (págs. 13-14)
- **Descubrimiento de conocimiento en texto** vale decir que es uno de los nuevos campos de aplicación de las herramientas de minería de datos.

2.8.4. Minería de texto

De acuerdo con (Manuel Montes, 2001):

> La minería de texto, considerada como el descubrimiento de patrones para la colección de textos, contiene tres etapas de los sistemas de minería de texto, la etapa de preprocesamiento, el tipo de representación y el tipo de

descubrimiento, la primera encargada de extraer información, la segunda realiza una secuencia de palabras y una tabla de los datos seleccionados y la última se encarga de relacionar entidades. (pág. 4)

Considerada también como un campo de investigación, que conjuntamente con sus técnicas, es eficaz y confiable la extracción correcta en la retribución de categorías de los textos, de acuerdo a la calidad del diccionario para detectar información dentro de un concluyente tema.

Otros autores consideran a la minería de texto como una "herramienta que tiene la capacidad de explorar grandes cantidades de textos no organizados con patrones, para ayudar a la toma de decisiones dentro de un ámbito específico" (Eleazar Botta Ferrer, 2007, pág. 1).

2.8.4.1. Herramientas de minería de texto

2.8.4.1.1. *R*

R, es considerada como una herramienta para el análisis estadístico para minería de datos siendo un lenguaje de programación interactivo que está diseñado para realizar un análisis de los datos recolectados, realizando computación estadística y facilitando funciones gráficas del mismo, utilizando árboles de decisión para resolver problemas de clasificación (Bellosta, 2018, pág. 8)

Ilustración 7. Software R. Fuente: (Bellosta, 2018, pág. 8)

2.8.4.1.2. *Python*

Python es un lenguaje de programación multiplataforma y multiparadigma el cual contiene un código comprensible, se centra en la exploración de datos ya que contiene variedad de bibliotecas que se encargan de la exploración de datos, modelado y también la visualización de estos, proporcionando las herramientas necesarias para llevar a cabo la minería de datos, siendo fácil de usar.

Este lenguaje ofrece también la posibilidad de ejecutar tareas de mejor manera, teniendo permanencia, modularidad y legibilidad del código, trabajando con inteligencia artificial, machine learning entre otros para poder explorar gran cantidad de datos y transformarlos en información útil.

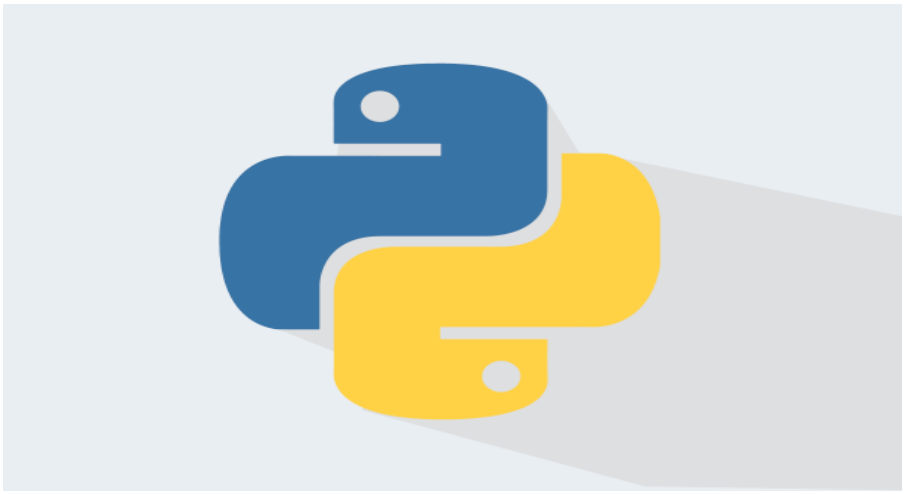

Ilustración 8. Lenguaje de programación Python. Fuente: (Pulido, 2021)

2.8.4.1.3. Tableau

Esta herramienta, permite la visualización de grandes volúmenes de datos, diseñada para inteligencia de negocios, que tiene como fin ayudar a comprender datos de un determinado tema, siendo sus resultados de fácil comprensión. Tableau es flexible y eficaz, entre sus funcionalidades están las visualizaciones de información multidimensional de forma inmediata y sencilla.

Ilustración 9. Software Tableau. Fuente: (Bellosta, 2018, pág. 8)

2.9. Cuadro Comparativo entre los lenguajes de programación

Las herramientas informáticas que ayudan a realizar minería de datos son múltiples, sin embargo, entre los lenguajes de programación más utilizados para realizar cálculos científicos, numéricos y estadísticos, así como para crear gráficas que ayudan a la toma de decisiones están el lenguaje R y Python.

Tabla 2 Cuadro comparativo de los lenguajes de programación para minería de datos. Autoría Propia..

	Python	**R**
General	Python es un lenguaje de programación de uso general para el análisis de datos y la computación científica.	R es un entorno y lenguaje de programación funcional para gráficos y computación estática
Objetivo	Ciencia de datos, Web, Desarrollo web.	Ciencia de datos y modelado estadístico
Entorno	**iPython, Pycharm, Jupyter Notebook, Spyder**	Rstudio, R, GUI, R KWARD
Recopilación de datos	Admite archivos CSV, SQL, JSON y webscraping con BeautifulSoup	También puede importar archivos csv con una biblioteca de lectura incorporada. La biblioteca de R, SCurl, proporciona una forma sencilla de

		realizar solicitudes de API, similar al paquete de solicitudes de Python.
Análisis de datos	Elbora marcos de datos con filtrado y ordenación de Pandas.	Realiza un análisis exploratorio de datos detallado.
Paquetes y bibliotecas esenciales	Numpy, Pandas, matplotlib, scipy, scikit-learn, TensorFlow	intercalador, cuerda, ggplot2, knitr, tldyverse, rebaja, brillante, pronosticar, refugio
capacidad de manejo de la base de datos	Puede manejar fácilmente datos grandes porque hay menos restricciones para el uso de la memoria	R maneja grandes cantidades de datos para elaborar diferentes cálculos: científicos, numéricos y estadísticos creando figuras de gran calidad en la memoria.
Visualización de los datos	A pesar de las capacidades de las herramientas de visualización de datos como Matplotlb y Seaborn, Python no está a la altura de las funciones	Desarrollado por y para datos estadísticos, R tiene funciones completas de visualización de datos

	de visualización de datos de R	
Sintaxis	El 'zen de Python es que hay una forma adecuada de escribir código	R no tiene este conjunto de reglas. Además, la indización comienza en 1, lo que puede considerarse poco convencional para los programadores generales
Curva de aprendizaje	La estructura de código simple y legible facilita el aprendizaje de los principiantes. También permite la programación orientada a objetos. También ofrece una amplia gama de estructuras de datos que no esperaría de un lenguaje de propósito general.	La sintaxis funcional de R no es fácil para los principiantes, pero no es demasiado desafiante para aquellos que están bien versados en programación. También ofrece algunas estructuras de datos, maneja grandes cantidades de datos.

De acuerdo a la tabla comparativa de los lenguajes de programación, se identifica a **R** como el lenguaje adecuado para realizar minería de datos ya que ofrece una amplia gama de librerías y maneja grandes cantidades de datos, siendo una herramienta útil para analizar detalladamente sentimientos positivos, negativos y neutros de varios temas.

2.10. Metodologías para proyectos de minería de datos

Actualmente, para realizar un análisis de sentimientos en redes sociales existen varias metodologías que permiten obtener resultados apropiados y también realizar una correcta toma de decisiones mediante una serie de pasos ordenados.

- **CRISP-DM:** Metodología más utilizada en proyectos de Data Mining, que permite la extracción y manipulación de la información para generar reportes de forma estadística conformada por cuatro niveles.

  El primer nivel compuesto por seis fases:

  1. **Comprensión del negocio:** Comprende cómo funciona la expresión de actitudes y sentimientos por parte del usuario en redes sociales.
  2. **Comprensión de los datos:** Recolecta los datos iniciales de la determinada red social con el fin de realizar una hipótesis de acuerdo a un tema en específico.
  3. **Preparación de los datos:** Se selecciona los datos más importantes para categorizarlos.
  4. **Modelado:** Se elije los métodos para procesar a los datos seleccionados mediante herramientas para poder construir el modelo
  5. **Evaluación:** Se ejecuta el modelo construido realizando pruebas de verificación.

6. **Implementación:** Se lleva a cabo informes finales representando así resultados que son necesarios para analizar el tema propuesto.

El segundo nivel de esta metodología se enfoca al cumplimiento de objetivos del proyecto mediante algunas actividades. El tercer nivel hace referencia a un mapeo de las tareas las cuales definen tanto el modelo como las actividades a desarrollar, y el último nivel sujeta al proceso de la toma de decisiones mediante los resultados obtenidos.

- **Metodología KDD:** KDD, es un proceso iterativo, el cual contiene cinco fases en donde la fase de data mining destaca ya que esta parte realiza varios procesos que ayudan a tener buenos resultados en el proyecto. Esta metodología se basa en herramientas que realizan análisis de datos tradicionales y también en tecnologías avanzadas de machine learning para dar resultados con pronósticos válidos. (Gervilla García, Jiménez López, Montaño Moreno, Sesé Abad, & Cajal, 2009, págs. 3-4)

  En cuanto a las etapas de esta metodología se detalla las siguientes:

  1. **Etapa de selección**: Se crea un conjunto de datos objetivo, seleccionando una muestra de los datos para realizar el proceso de descubrimiento.
  2. **Etapa de limpieza:** Se analizan la cantidad de los datos, aplicando estrategias como el missing y emty para eliminar datos duplicados y datos nulos.
  3. **Etapa de transformación:** Aplica una búsqueda de características necesarias para representar los datos, mediante la reducción de dimensiones o de

transformación para reducir el número de variables no necesarias.

4. **Etapa de minería de datos:** Esta etapa se encarga de la búsqueda y descubrimiento de patrones, a través de tareas como la clasificación, clustering, patrones secuenciales y asociaciones.
5. **Etapa de interpretación:** Los patrones son interpretados consolidando el conocimiento para documentar los resultados y verificar y resolver conflictos potenciales previamente descubiertos.

- **Metodología SEMMA:** Metodología creada por SAS Institute, es un conjunto de herramientas para la selección, exploración y modelado de grandes volúmenes de datos con el fin de encontrar patrones que permitan analizar de forma estadística un posible resultado, está compuesta por cinco fases:

1. **Muestreo**

   En esta etapa se extraen los datos con el objetivo de reunir datos con las mismas características.

2. **Exploración**

   Esta fase se realiza mediante técnicas estadísticas que están diseñadas para identificar relaciones y tendencias entre los datos formulando hipótesis.

3. **Modificación**

   Se selecciona los datos de acuerdo a las variables presentadas para el modelado de datos.

4. **Modelado**

   Mediante las técnicas de minerías de datos que permiten asociar y combinar datos, se realiza el modelado de los datos.

5. **Evaluación**

   De acuerdo al modelado, los resultados son evaluados midiendo su exactitud y tratando de que estos sean precisos y confiables (Gervilla García, Jiménez López, Montaño Moreno, Sesé Abad, & Cajal, 2009, págs. 3-4)

Tabla 3. Cuadro comparativo de las metodologías utilizadas para proyectos de minería de datos. Autoría Propia.

Metodologías	CRISP-DM	KDD	SEMMA
Estructura Fases	**Niveles** 1.- Comprensión del negocio 2.- Comprensión de los datos 3.- Preparación de datos 4.- Modelado 5.- Evaluación 6.- Implementación	**Fases** 1.- Comprensión del dominio de aplicación 2.- Creación del conjunto de datos 3.- Limpieza y preprocesamiento 4.- Reducción y proyección de los datos 5.- Determinación de la minería de datos 6.-	**Fases** 1.- Muestreo 2.- Exploración 3.- Modificación 4.- Modelado 5.- Evaluación

		Determinar el algoritmo de la minería 7.- Minería de datos 8.- Interpretación	
Presentación final de documentos	Presenta una guía de usuario con un modelo de referencia	No especifica actividades puntuales	No presenta una guía de actividades determinadas

2.10.1. Tabla de metodologías para proyectos de minería de datos

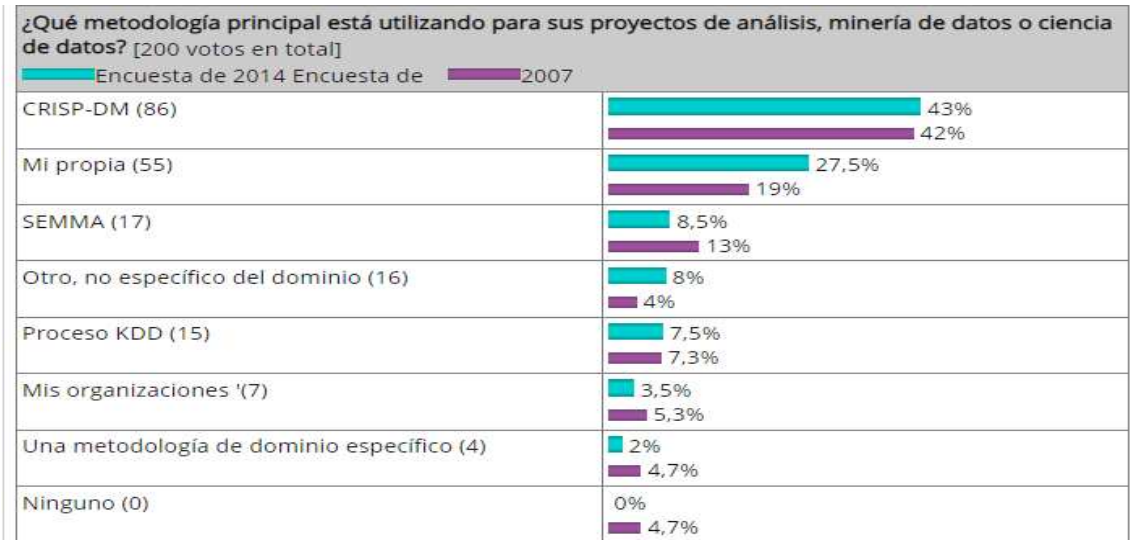

Ilustración 10. Metodologías utilizadas en Data Mining. Fuente: (Kdnuggest, 2014)

En base a una comparación desarrollada de metodologías para minería de datos, realizada por (Wendy Jahayra Pinargote Mendoza, 2018), (Kdnuggest, 2014), en su proyecto "Aplicación de técnicas de minería de datos para el análisis de sentimientos en la red social Facebook sobre el servicio de telefonía móvil en Ecuador", determina que la mejor metodología para minería de datos, es la CRISP-DM, ya que cuenta con seis fases que son bidireccionales con el objetivo de corregir errores.

Es por esta razón se ha optado por la metodología CRISP-DM, para el desarrollo del presente proyecto, en el cual se dará cumplimiento en cada una de las fases establecidas para la ejecución de los tweets en minería de datos.

2.11. Social Media "Inteligencia de Negocios"

Actualmente las redes sociales han crecido, ya que cada vez son más los usuarios que se integran a estas, siendo un medio de transmisión de todo tipo de información en las cuales las personas expresan sus sentimientos. Es por eso que las organizaciones pueden realizar análisis de las necesidades de los usuarios mediante herramientas diseñadas que van en beneficio de un determinado tema de publicidad o marketing. (DOMÍNGUEZ, 2010, págs. 3-15)

Un artículo elaborado por (Donald Córdova, 2020)

> El impacto de la Inteligencia de Negocios en redes sociales, es fundamental para la detección del comportamiento por parte de los usuarios, siendo la inteligencia artificial de gran ayuda para recolectar información y posteriormente lograr un objetivo procesando los datos en una BDD que ayuda a la toma de decisiones. (pág. 6)

2.11.1. Las redes sociales

Según (MSc. Kelly Deysi Hernández Mite, 2017) afirma que las diferentes redes sociales que existen, son plataformas que tienen como objetivo presentar múltiples espacios de información que pueden satisfacer las necesidades y preferencias de los usuarios, mediante imágenes, textos, videos entre otros, facilitando la interacción de entidades y clientes en productos o servicios.

Para (Crovi, 2009, pág. 15) “Las redes son una estructura sistémica y dinámica que involucra a un conjunto de personas u objetos, organizados para un determinado objetivo, que se enlazan mediante una serie de reglas y procedimientos.”

2.11.2. Análisis de Sentimiento con BI

Conocido también como minería de opinión, encargada de detectar de forma automática los sentimientos y actitudes de los diferentes usuarios que se dan mediante texto sobre un evento u objeto, para luego ser clasificados tanto de forma positiva como negativa o neutral. (Juan Antonio Vicente Virseda, 2019, pág. 87)

Según (Javi Fernández, 2011):

La expansión de la Web 2.0 ha provocado en varios autores, analizar sentimientos y actitudes de los usuarios para un determinado tema que ayuda a categorizar efectos positivos, negativos y neutrales sobre un objeto u servicio, beneficiando en temas de competitividad gracias a herramientas que permiten el correcto procesamiento de los datos, convirtiéndolos en información fundamental. (págs. 2-3)

Business Intelligence, ayuda a la toma de decisiones a los usuarios de una organización, y en el ámbito del análisis de sentimientos, mediante sus tecnologías como OLAP, Data warehouse, entre otras se puede realizar el análisis de sentimientos en redes sociales, apoyándose en una base de datos, para de esta manera aportar información clasificada que ayuda a ser más eficaz y confiable para una entidad (Díaz, 2012, págs. 18-19-20).

Es importante saber cómo se puede convertir los datos en datos procesables y usarlos para mejorar el análisis. Estamos acostumbrados al término "minería de datos" pero aquí, realmente estamos analizando la "minería de opinión" mediante el análisis de sentimientos. Cuando se usa y analiza correctamente, el análisis de sentimientos es una

herramienta muy poderosa. Permite a las marcas comprender el comportamiento de los usuarios y reaccionar en función de esos hallazgos a que realizan análisis de opinión. (Guenther, 2019)

Hoy en día estamos viviendo la era digital en lo cual los usuarios requieren más de las redes para manifestar sus opiniones, sin embargo, reflejan su ser o comportamiento ante un fenómeno social que determinan un nivel de minería de texto positiva o negativa.

El análisis de sentimientos obtiene información sobre todo tipo de fenómenos sociales en línea por lo tanto facilita de una forma más masiva y relativamente para la identificación, recopilación y análisis de las experiencias u opiniones de los usuarios.

- **Análisis de sentimientos en redes sociales**

  Realizar un análisis de sentimientos en las diferentes redes sociales, se debe efectuar mediante herramientas las cuales se encargan del análisis de texto con el fin de realizar un procesamiento natural del lenguaje para poder clasificar sentimientos de los usuarios sobre un determinado tema, dándose esto mediante patrones (Leonardo M. Moreno, 2019)

  Debido a que las redes sociales son plataformas en las cuales los usuarios manifiestan sus opiniones de manera pública, las aplicaciones de la inteligencia artificial son diseñadas para detectar y analizar opiniones con varias técnicas que permiten la generación de un informe detallado y estadístico en tiempo real de los sentimientos de los usuarios en la web. (Javi Fernández Y. G., 2015)

Las redes sociales creemos que pueden considerarse útil instrumento de control político, obtención de información por los ciudadanos y participación en el debate político informando, siendo un instrumento de difusión. desde otra perspectiva de análisis dinamizan el sentimiento social por la preocupación que se expresa en la selección de los asuntos públicos que son del interés sociales. (Garza, 2015, pág. 179)

Las redes sociales creemos que pueden considerarse un útil instrumento de control político, obtención de información por los usuarios y participación en el debate político, siendo un instrumento de difusión de mensajes de camdinamizan el sentimiento social por la preocupación que se expresa en la selección ante un fenómeno social. (Garza L. M., 2006, pág. 176)

- **Fenómenos sociales**

El internet y la facilidad de acceder a las redes sociales, han determinado ciertos cambios en nuestras vidas, ya que estas plataformas abordan varios ámbitos a nivel mundial. Combinando aspectos sociales, económicos y políticos, que permiten a las personas dar su opinión y expresar sus sentimientos frente a diferentes situaciones de cada uno de los fenómenos antes mencionados, siendo importante realizar un análisis de comportamiento basándose en textos, imágenes entre otros, con el fin de establecer una conclusión para una correcta toma de decisiones.

Debido a que individuos a nivel mundial interactúan entre sí, la política ha sido un campo que ha destacado en las principales redes sociales, especialmente Twitter, existiendo de esta forma “partición democrática haciendo de la red un medio más social” (Dijck, 2019, pág. 4) siendo así esta

plataforma un recurso valioso en el cual se puede analizar los sentimientos que van de acuerdo a la política.

El análisis de redes sociales, en el campo económico ha sido bastante significativo ya que permite analizar a la sociedad, el uso de estos medios para incrementar su actividad financiera tanto en organizaciones como personas en general. Y en cuanto al ámbito social las herramientas de Big Data permiten realizar un análisis detallado sobre las actitudes y emociones de las personas al realizar diferentes actividades sociales y culturales.

CAPITULO III

3. ENFOQUE DE LA INVESTIGACIÓN

3.1. ENFOQUE DE LA INVESTIGACIÓN

En el presente trabajo se considera la necesidad de analizar sentimientos expresados por los usuarios de Twitter en los diferentes mensajes o "tweets", relacionados a varios temas de interés como es la política ecuatoriana y la salud, utilizando un enfoque mixto ya que manejan variables tanto cualitativas como cuantitativas.

3.2. NIVEL DE INVESTIGACIÓN

La presente investigación es de tipo descriptiva, ya que se realiza una recolección de datos para luego ser procesados con diferentes técnicas de minería de datos y de texto para luego obtener un informe detallado de resultados.

3.3. POBLACIÓN Y MUESTRA

3.3.1. Población

La población será 3.200 registros correspondientes al número de tweets que la herramienta selecciona del total de los tweets relacionados a cada uno de los temas producto de la investigación.

3.3.2. Muestra

Al ser la población solamente de 3.200 registros, no se realizará muestreo y se trabajará sobre el universo íntegro de los datos.

Realizando el muestreo en el mes de junio del 20 al 24 del presente año.

3.4. MÉTODOS DE INVESTIGACIÓN

El proyecto presenta un método deductivo, ya que el proceso de análisis de los datos va de lo general a lo particular. Además, se utilizará el método CRISP-DM, ya que es considerado como la mejor metodología en proyectos de minería de datos, como se puede observar en la ilustración 6 y a la tabla 1 comparativa realizada en el capítulo II del proyecto.

3.5. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN

Para realizar la recolección de los datos, se realiza mediante técnicas de minerías de datos. Se ejecutarán diferentes librerías que permiten el acceso a Twitter para la recolección de datos de los diferentes temas, las cuales se encuentran detalladas a continuación:

3.5.1. LIBRERIAS PARA EXTRACCIÓN DE INFORMACIÓN

Las librerías en **R** son un conjunto de instrucciones funcionales codificadas en un lenguaje de programación que permiten el tratamiento de múltiples datos. Se encuentran dentro de paquetes que deben ser preinstalados.

(rtweet): Ofrece a los usuarios una variedad de funciones diseñadas para la extracción de datos de las API de transmisión y REST de Twitter.

(tidyverse): Es una librería orientada a la manipulación, exploración y visualización de datos. Facilitando el trabajo estadístico.

(knitr): Facilita la mezcla de cualquier tipo de texto con cualquier código

(wordcloud): Permite la elaboración de gráficas estadísticas.

(RColorBrewer): Gestiona colores en la herramienta para los gráficos estadísticos.

(tidytext): Contiene funciones y conjuntos de datos de apoyo para permitir la conversión de texto en formatos ordenados.

(igraph): Librería para análisis de redes, proporcionan un conjunto de tipos de datos y funciones para la implementación de algoritmos de grafos.

(ggraph): Tiene la función de mostrar una descripción general de los datos de la red, permite colorear, etiquetar y dimensionar nodos, así como características de esta.

3.5.2. Recuperación de tweets

Los diferentes comandos utilizados para la ejecución de la recuperación de Tweets en cuando a los diferentes temas son:

- **searchTwitter:** Se encarga de la extracción de datos que contienen una cadena en particular
- **as_tibble:** Sirve para crear un nuevo tibble a partir de vectores individuales,
- **write.csv:** Este comando se encarga de exportar datos generados por as.data.frame.
- **exclude:retwets:** Mediante un filtro se excluye los tweets.

3.6. TRATAMIENTO DE LA INFORMACIÓN

Existe un gran número de datos que se pueden encontrar en Twitter, mismos que son consideradas como fenómenos sociales; entre las más relevantes en esta época tenemos efectos de la postpandemia, el Covid-19 y la política, son temas que se han vuelto virales en redes sociales, cuyas publicaciones causan tendencia tanto a favor como en contra de las mismas.

Es por eso que se ha seleccionado estos temas como prueba para el caso de estudio, basándose en un artículo realizado por (Sofía Cabrera, 2020), el

cual hace referencia a la tendencia que causa el COVID-19 en científicos ecuatorianos, quienes manifiestan sus diferentes puntos de vista y emociones ante la enfermedad, contando con varios tweets de usuarios, analizando su postura de opinión.

En cuanto a la política (Manuel Antonio Conde, 2021), analiza a 19 presidentes en la red social Twitter y la tendencia que causan mediante mensajes de cada propietario y las reacciones de los usuarios ante sus publicaciones. Por otra parte, una tesis elaborada por (Pérez, 2020), analiza las elecciones del año 2019, en donde interviene el actual presidente Guillermo Lasso, observando la carga emocional de los tweets dirigidos a la política ecuatoriana y los partidos políticos más votados con sus respectivos representantes.

CAPÍTULO IV

PROPUESTA

4. TÍTULO DE LA PROPUESTA

ANÁLISIS DE SENTIMIENTOS EN COMUNIDADES DIGITALES UTILIZANDO TÉCNICAS DE BIG DATA PARA DETERMINAR PATRONES DE COMPORTAMIENTO ORIENTADO A FENÓMENOS SOCIALES

4.1 Presentación

El presente proyecto tiene como objetivo desarrollar un método que sea capaz de analizar los sentimientos de los usuarios de las diferentes comunidades digitales, con el fin de extraer patrones de comportamiento sobre fenómenos sociales que afectan a la comunidad mediante la explotación de técnicas de BIG DATA.

4.2. Justificación

Actualmente, el mundo gira en torno a las redes sociales, ya que a través de estas las personas expresan sus sentimientos ante diferentes temas, que con el paso del tiempo se vuelve un fenómeno social.

Considerando que las TIC, nos ofrecen herramientas de minería de datos, las cuales se encargan de realizar un análisis detallado sobre los temas y las palabras que causan tendencia en la sociedad, se ha optado por la utilización de la herramienta R para un análisis de sentimientos en la red social Twitter, en la cual se analiza los temas como: **COVID-19, POLÍTICA,** con el fin de extraer patrones de comportamiento sobre estos fenómenos sociales que aportan relevancia entre los usuarios.

Para poder llegar al objetivo determinado se basa en las fases de la metodología CRISP-DM, la cual permite la extracción y manipulación de la información para generar reportes de forma estadística.

4.3. Desarrollo del proyecto

Para llevar a cabo el desarrollo del proyecto de análisis de sentimientos se seguirá cada una de las fases de la metodología antes mencionada, comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación, implementación en la herramienta ***R***.

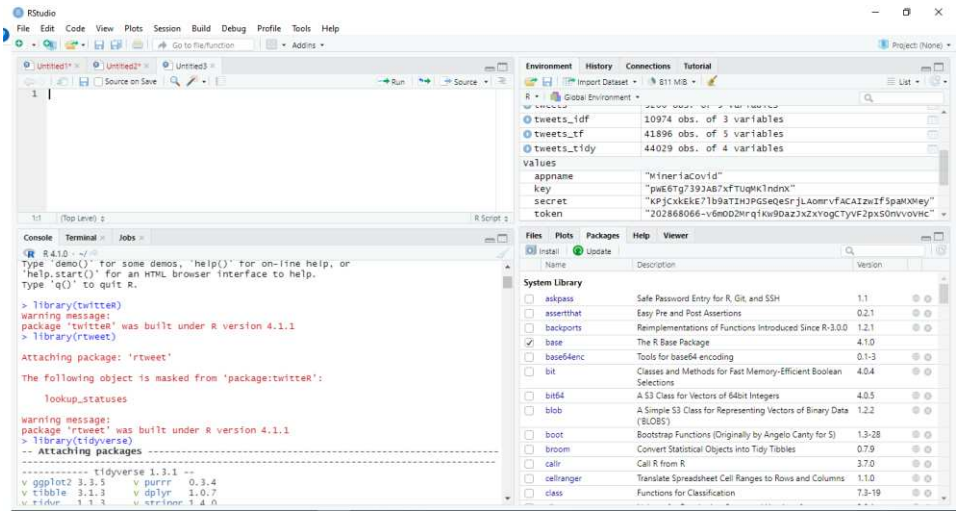

Ilustración 11. Interfaz de R. Autoría Propia.

4.3.1. Fases de la metodología CRISP-DM

Debido a los tres temas analizados en la herramienta y a la metodología propuesta, los resultados son descritos en cada una de las etapas descritas a continuación:

1. **Propósito de la minería de texto**

En la presente fase se determinó los objetivos de búsqueda, los cuales están relacionados con temas de la actualidad, que son relevantes y los más mencionados en las redes sociales en el Ecuador.

En lo que respecta a la política ecuatoriana se analizó temas de actualidad, el primer caso de análisis es el caso del actual presidente de la República de Ecuador, **Guillermo Lasso,** quien ha causado tendencia en varias redes sociales, resaltando también al ex presidente **Lenin Moreno** quien toma relevancia en publicaciones y comentarios de los diferentes usuarios. Por otra parte, en el tema de la salud, en los últimos años el **COVID-19** ha sido uno de los temas con mayor mención a nivel mundial.

Por otro lado, se ha seleccionado a la red social **Twitter** como objeto de nuestro estudio, en vista de que es en esta red social que dichos personajes son muy activos manteniendo sus cuentas (@Lenin, @lasso), con publicaciones diarias**.** Además, que todos los temas relacionados con COVID-19 son tendencia en dicha red social.

Luego del análisis realizado se determinó como objetivos de búsqueda las interacciones realizadas en la red social Twitter sobre los temas Lenin Moreno, Guillermo Lasso y COVID-19, información que mediante la aplicación de técnicas de minería de texto y análisis de sentimientos permitirá obtener un mejor entendimiento de la interacción de los usuarios con respecto a estos temas.

2. Recuperación de la información

Esta fase corresponde a la recolección de los datos necesarios para el desarrollo de la presente investigación, los cuales serán obtenidos de la red social Twitter, mediante la herramienta ***R,*** y su ***RESTAPI,*** que realiza búsqueda sobre tweets, usuarios, timelines y otros objetos. Entregando a un usuario un conjunto de tweets u otros objetos que cumplen las condiciones de la consulta a realizar, que para el motivo de la investigación son Guillermo Lasso, Lenin Moreno, COVID-19, esta herramienta proporcionó acceso a los tweets de los últimos 6 a 9 días, sobre dichos temas, teniendo como entradas a las publicaciones y comentarios de perfiles.

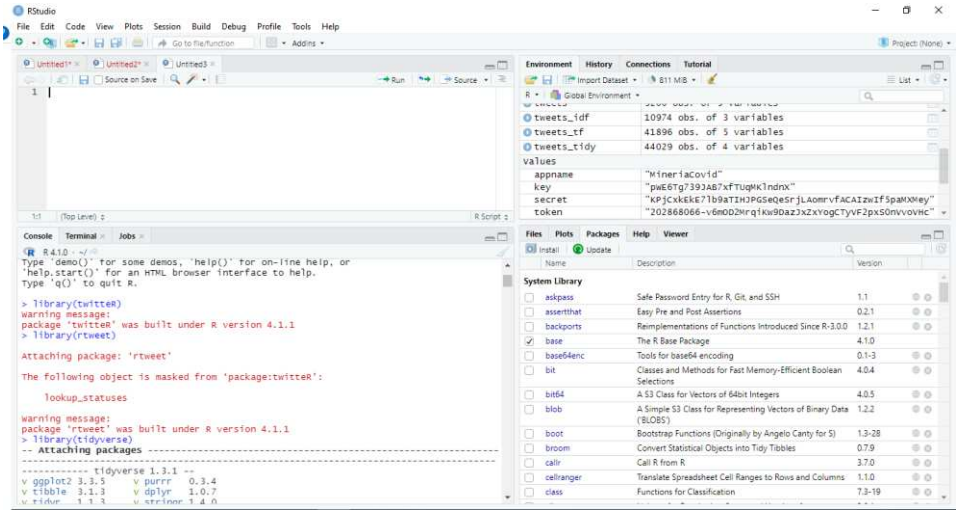

Ilustración 12. Instalación de librerías. Autoría Propia.

Para ello se determinó un total de tweets a ser analizados de cada uno de los temas antes mencionados, a través de la ejecución de varios comandos que se encuentran en el anexo 2 que nos permite tomar una ventana de tiempo correspondientes a los temas de Twitter para un análisis posterior.

Se explora las diferentes opiniones de los usuarios de Twitter que han comentado sobre la política del Sr. Guillermo Lasso, a través de perfiles y hashtags obteniendo un total de 3200 tweets.

```
>                   total_documentos = tweets_tidy $tweet_id %>% unique() %>% length()
>                   total_documentos
[1] 3200
```

Ilustración 13. Total de tweets de Guillermo Lasso. Autoría Propia.

Con lo que respecta al tema de Lenin Moreno, se realizó un análisis similar al primer tema en el cual se obtuvo un total de 2624 tweets.

```
>                   total_documentos = tweets_tidy $tweet_id %>% unique() %>% length()
>                   total_documentos
[1] 2624
```

Ilustración 14. Total de tweets de Lenin Moreno. Autoría Propia

Debido a que la herramienta limita el número de tweets a ser analizados se obtuvo un total de 3200 tweets, para el tema del COVID-19, los cuales serán procesados para obtener el análisis de sentimientos por parte de los usuarios.

```
>                   total_documentos = tweets_tidy $tweet_id %>% unique() %>% length()
>                   total_documentos
[1] 3200
>
```

Ilustración 15. Total de tweets del COVID-19. Autoría Propia

3. Extracción de la información

Esta fase se encarga de realizar la tokenización y limpieza de los datos, eliminando del texto todo aquello que no aporta ninguna información como: caracteres innecesarios, removiendo signos de puntuación y palabras que no aportan información necesaria, sobre los temas tanto de política como de salud.

La presente imagen interpreta la limpieza de las palabras (ted, de, que, la, el) en el tema de la política de Lasso, palabras que no contribuyen información en el contenido.

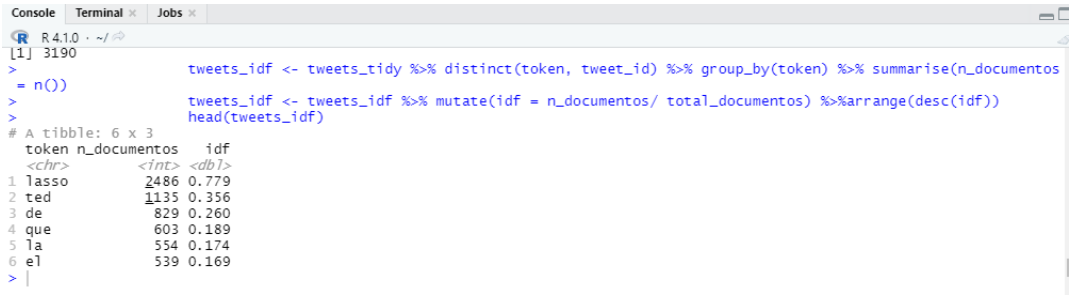

Ilustración 16. Limpieza y tokenización (Guillermo Lasso). Autoría Propia

De la misma manera en el tema de Lenin Moreno, se realizó la tokenización como se muestra en la figura 17, y la limpieza de las palabras (de, que, la, el), que se presenta en la figura 18 realizando este proceso a través de las librerías **tokenizers** o **quanteda.**

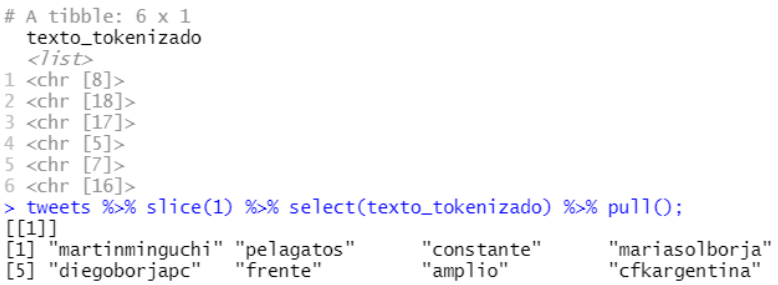

Ilustración 17. Tokenización (Lenin Moreno). Autoría Propia

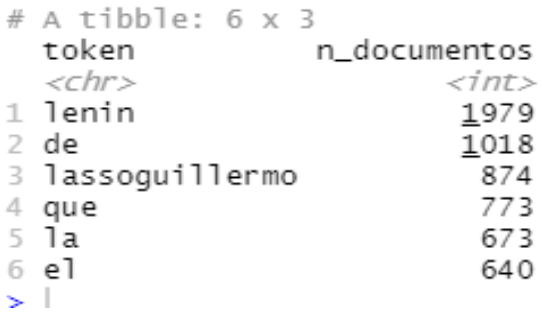

Ilustración 18. Limpieza de datos (Lenin Moreno). Autoría Propia

Al igual que los temas de la política ecuatoriana, en el tema del COVID-19 se realizó también este procedimiento para poder llevar a cabo la siguiente etapa, seleccionando las palabras innecesarias (the, to, in, de, of) como se muestra en la siguiente figura 19.

```
>                   tweets_idf <- tweets_tidy %>% distinct(token, tweet_id) %>% group_by(token) %>% summarise(n_docum
>                   tweets_idf <- tweets_idf %>% mutate(idf = n_documentos/ total_documentos) %>%arrange(desc(idf))
>                   head(tweets_idf)
# A tibble: 6 x 3
  token n_documentos   idf
  <chr>        <int> <dbl>
1 covid         2137 0.668
2 the            778 0.243
3 to             603 0.188
4 in             532 0.166
5 de             530 0.166
6 of             512 0.16
>
```

Ilustración 1. Limpieza y tokenización (COVID-19). Autoría Propia.

4. Procesamiento de texto

Esta fase evalúa técnicas de procesamiento de texto, generando reportes estadísticos, determinando la frecuencia de cada uno de los términos.

Del análisis del tema Guillermo Lasso, se obtiene el término con mayor mención (lassoguillermo), el más utilizado en los tweets.

Ilustración 20. Términos más utilizados en los tweets (Guillermo Lasso). Fuente: Autoría Propia

De los tweets que mencionan al ex presidente Lenin Moreno, se obtiene un reporte estadístico desde el 22 hasta el 24 de junio del presente año, siendo los términos **Lenin** y **dlasamericas,** las palabras más utilizadas por los usuarios, por otra parte, los términos **país, peor, traidor, mejor, Venezuela, ahora, ifmolinao,** ocupan menos frecuencia.

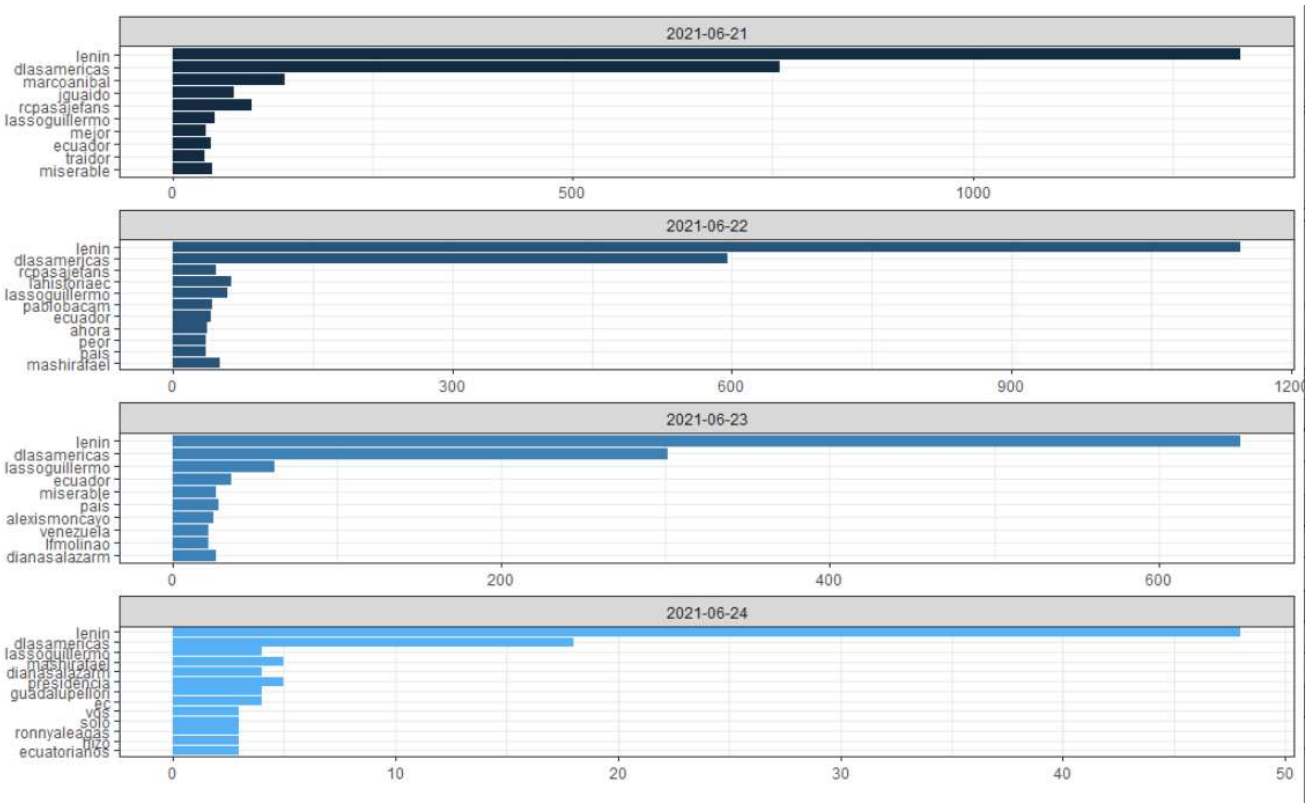

Ilustración 2. Palabras con mayor mención de forma estadística. Autoría Propia.

En el siguiente gráfico se puede visualizar la palabra (**lenin**), como el término con mayor mención, seguida de **dlasamericas**.

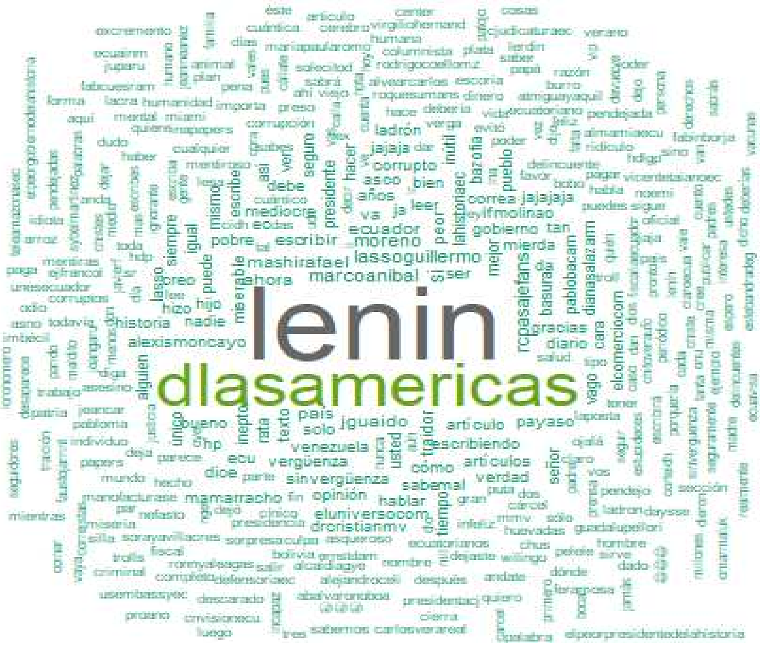

Ilustración 6. Términos más utilizados en los tweets (Lenin Moreno). Autoría

Dentro del tema del COVID-19, se determina la frecuencia desde 0 a 150, resaltando los términos más y menos utilizados por los usuarios, siendo el término ***vaccine,*** con más repetición en los tweets.

Por otra parte, las palabras que ocupan menos frecuencia son ***medidas, help, deaths.***

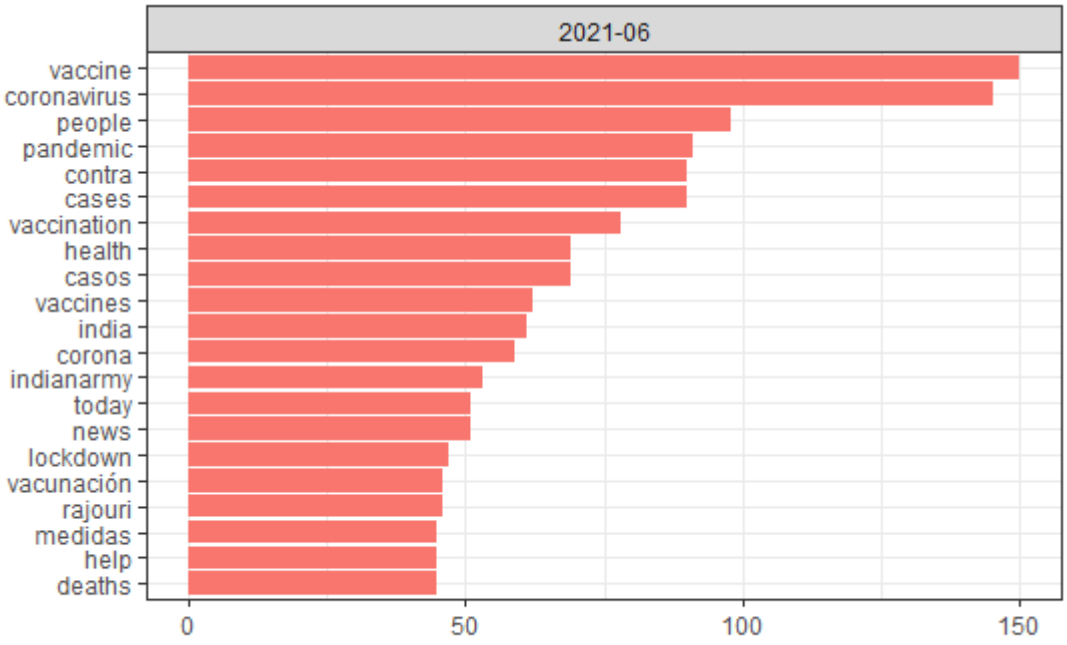

Ilustración 23. Reporte estadístico de las palabras con mayor mención (COVID-19). Autoría Propia.

De la misma forma la siguiente ilustración permite visualizar el término más utilizado (vaccine) con el número de letra mayor a las demás analizadas, obteniéndose mediante la librería wordcloud.

Ilustración 24. Palabras con mayor mención en Twitter del tema COVID-19. Autoría Propia.

5. Aplicación del método de minería de texto

En base a los tweets que toman mayor relevancia en la herramienta, las librerías instaladas se encargan de relacionar a las palabras en bigramas, a través del código disponible en el anexo de la página 111.

En lo que respecta al tema de política, especificando al actual presidente Guillermo Lasso se realiza un análisis de la relación de palabras en bigramas con el fin de obtener una forma más visual e informativa de análisis. Estableciendo las relaciones:

GuillermoLasso-aborrero

GuillermoLasso-mashirafael

GuillermoLasso-andreajserrano

GuillermoLasso-lahoraecuador

GuillermoLasso-salud

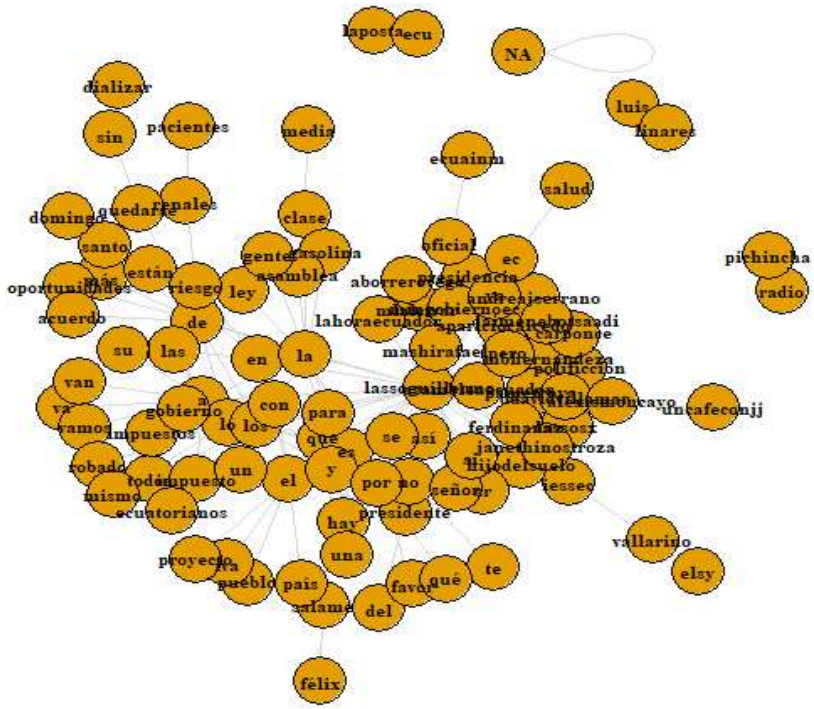

Ilustración 25. Relación de los términos (Guillermo Lasso). Autoría Propia.

De la misma manera en el tema de Lenin Moreno se obtiene un bigrama con las relaciones entre palabras mediante el uso de networks, el cual permite una visualización más explicativa, descartando los términos (salud, ec, Venezuela, evito, la posta, ecu) que se encuentran fuera del nodo principal, el cual une a cada una de las palabras.

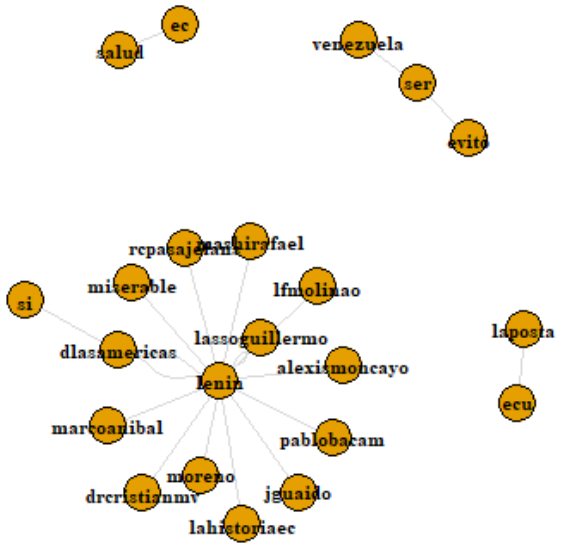

Ilustración 26. Relación de los términos (Lenin Moreno). Autoría Propia.

Lenin-lassoguillermo

Lenin-moreno

Lenin-alexismoncayo

Lenin-miserable

Lenin-mashirafael

Lenin-pablobacam

Aquí se nota las relaciones existentes con otros políticos de la actualidad como el presidente actual Guillermo Lasso, el ex presidente Rafael Correa, etc., además improperios contra su persona como la palabra miserable.

En la siguiente ilustración, se observa la agrupación de palabras por su frecuencia de uso común en lo que respecta al COVID-19, conociendo así los distintos temas que se tratan en los tweets y relacionando cada uno de ellos.

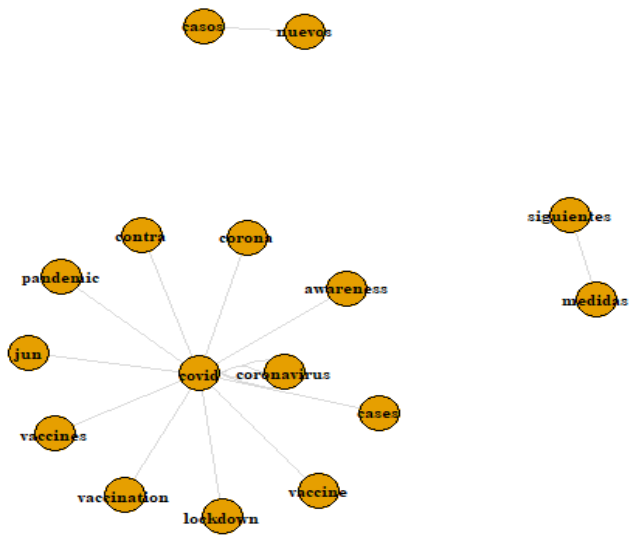

Ilustración 27. Relación de los términos (COVID-19). Autoría Propia.

Siendo la relación común las siguientes:

Covid-vaccines

Covid-lockdown

Covid-vaccination

Covid-pandemic

Covid-awareness

Lo que da a entender que las relaciones más comunes son las de Covid con el proceso de vacunación.

6. Análisis de resultados e interpretación

Esta última etapa comprende a la obtención del total de sentimientos en base a los tweets analizados, clasificándolos en positivos, negativos y neutros, obteniendo como resultado los siguientes análisis de los temas propuestos.

El análisis de sentimientos se obtiene representado en números tales como:

-1, 1, 0, siendo el -1 un sentimiento negativo, el 1 sentimiento positivo y el 0 un sentimiento neutro.

```
> tweets_sent %>% group_by(tweet_id,mes_anyo) %>%summarise(sentimiento_promedio = sum(valor)) %>%head()
`summarise()` has grouped output by 'tweet_id'. You can override using the `.groups` argument.
# A tibble: 6 x 3
# Groups:   tweet_id [6]
  tweet_id            mes_anyo sentimiento_promedio
  <chr>               <chr>                   <dbl>
1 1405531186592579587 2021-06                    -1
2 1405531259573444610 2021-06                     1
3 1405531433095995395 2021-06                    -1
4 1405531487890403331 2021-06                    -1
5 1405531552780414990 2021-06                    -1
6 1405531653926096896 2021-06                    -2
```

Ilustración 28. Resultados representados de forma numérica (Guillermo Lasso) Autoría Propia.

El actual presidente Guillermo Lasso, adquiere como resultados: 50.9 %positivos, 46.8%negativos y 2.34%neutros, determinando que la mayoría de usuarios realizan publicaciones positivas.

```
# A tibble: 1 x 3
  positivos neutros negativos
      <dbl>   <dbl>      <dbl>
1      50.9    2.34       46.8
```

Ilustración 29. Resultados del análisis de sentimientos (Guillermo Lasso). Autoría Propia.

De la misma manera, se puede visualizar en la siguiente ilustración, los resultados de forma estadística del tema Guillermo Lasso, representado por tres colores, verde (neutros), azul (positivos) y rojo (negativos).

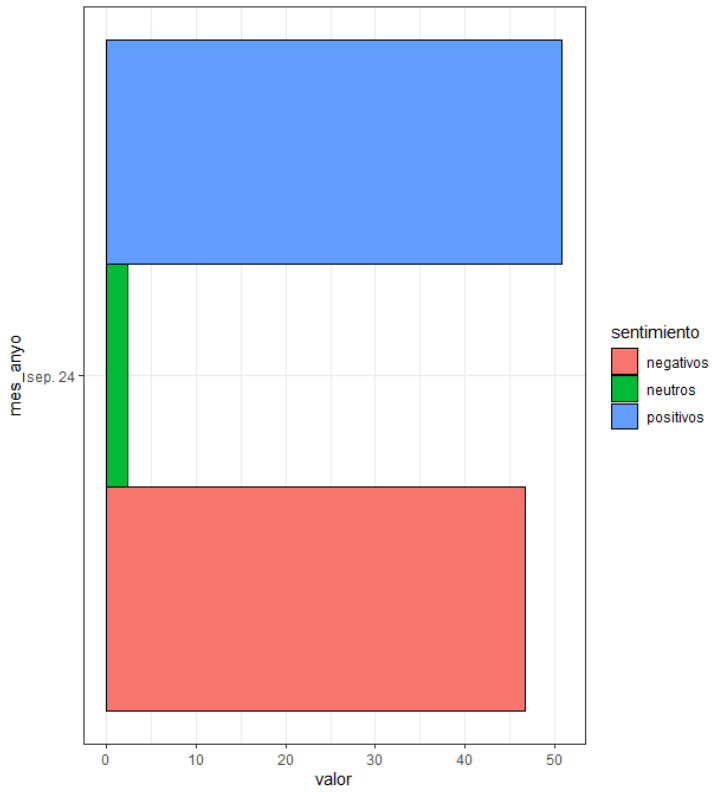

Ilustración 30. Resultados representados de forma estadística (Guillermo Lasso). Autoría Propia.

El ex presidente Lenin Moreno obtuvo un total de 2624 tweets de los cuales en el análisis de sentimientos se determina que un 80.7 %, son sentimientos negativos, 17.7 % corresponde a sentimientos positivos y 1.61 % son sentimientos neutros, es decir que dicho personaje llega a los usuarios de manera negativa.

```
# A tibble: 1 x 3
  positivos neutros negativos
      <dbl>   <dbl>      <dbl>
1      17.7    1.61       80.7
```

Ilustración 31. Resultados del análisis de sentimientos (Lenin Moreno). Autoría Propia

Los resultados obtenidos en el gráfico anterior, se muestran a continuación de forma estadística, representado por colores:

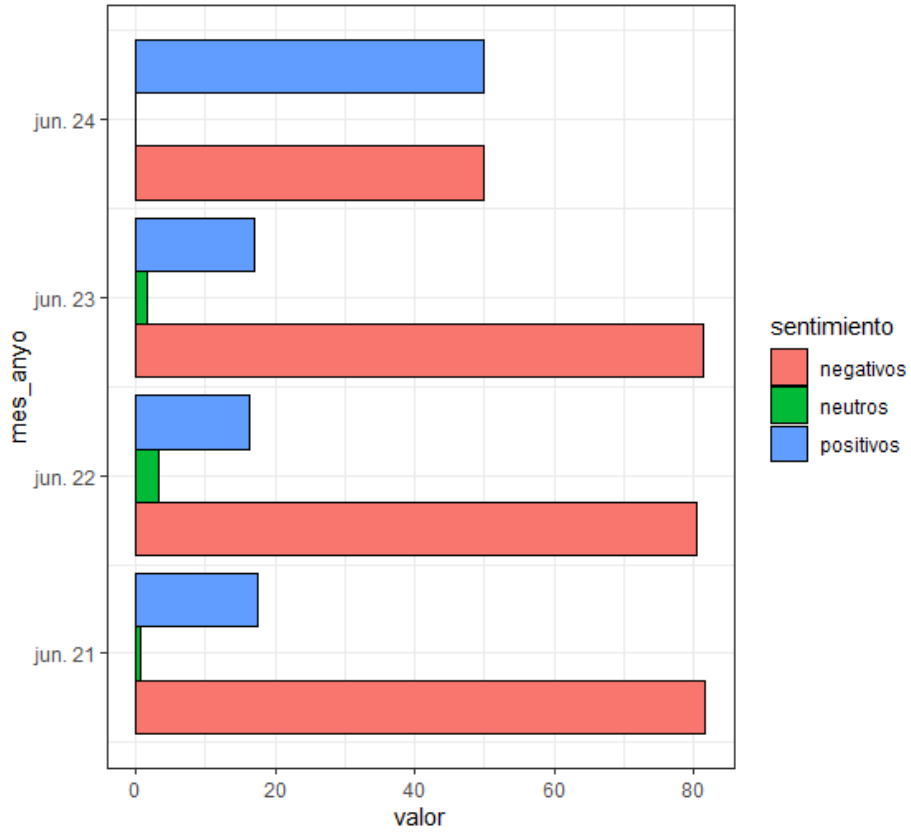

Ilustración 32. Resultados representados de forma estadística (Lenin Moreno). Autoría Propia.

Para obtener los resultados de sentimientos se utilizó la librería **Bing,** la cual permite determinar si son positivos, negativos o neutros.

```
> sentimientos <- sentimientos %>% mutate(valor = if_else(sentiment == "negative", -1, 1))
> tweets_sent <- inner_join(x = tweets_tidy, y = sentimientos, by = c("token" = "word"))
> tweets_sent %>% group_by(tweet_id) %>% summarise(sentimiento_promedio = sum(valor)) %>%head()
# A tibble: 6 x 2
  tweet_id            sentimiento_promedio
  <chr>                             <dbl>
1 1405531186592579587                  -1
2 1405531259573444610                   1
3 1405531433095995395                  -1
4 1405531487890403331                   0
5 1405531552780414990                  -1
6 1405531653926096896                  -2
```

Ilustración 7. Resultados representados de forma numérica (COVID-19) Autoría Propia.

En cuanto al COVID-19, se obtuvo un 45.7 % de resultados positivos, 46.2 % negativos y 8.12% resultados neutros, concluyendo que esta enfermedad genera en la mayoría de los usuarios sentimientos negativos, tal y como se muestra en la ilustración 34.

	mes_anyo	positivos	neutros	negativos
	<chr>	<dbl>	<dbl>	<dbl>
1	2021-06	45.7	8.12	46.2

Ilustración 8. Resultados del análisis de sentimientos (COVID-19). Autoría Propia

CONCLUSIONES

En el presente trabajo se ha estudiado las diferentes definiciones propuestas por diferentes autores, en los que respecta el "Análisis de sentimientos" y los temas relacionados en minería de textos, analizando cada uno de los puntos abordados para el desarrollo de la investigación.

Se concluye que la metodología CRISP-DM, ayuda a comprender de una manera fácil el comportamiento de la información, a través de cada una de las fases.

La herramienta para análisis de sentimientos fue seleccionada en base a una matriz comparativa realizada en el capítulo 2, siendo R apto para este trabajo, esta herramienta permitió analizar los sentimientos en la red social Twitter, obteniendo datos reales de dicha red social relacionados a todo tipo de usuarios o temas de discusión, que en este caso se analizaron temas correspondientes a Guillermo Lasso, Lenin Moreno y el COVID-19, logrando reconocer las opiniones o sentimientos de los usuarios en todas la publicaciones referentes a los temas mencionados, permitiendo también entender la situación momentánea y relevante de esos temas.

RECOMENDACIONES

Se recomienda a la Carrera de Ingeniería de Sistemas de la Universidad Católica de Cuenca:

- Emprender proyectos de investigación relacionados a minería de texto, análisis de sentimientos en vista de que son temas que se encuentran en auge y que servirán a los estudiantes para ampliar sus posibilidades laborales.
- Incorporar conocimientos sobre minería de texto que permiten mejorar el perfil profesional de los estudiantes graduados de la carrera.

REFERENCIAS

Aguilar, J. (s.f.). *UTPL.* Obtenido de http://www.ing.ula.ve/~aguilar/actividad-docente/TIN/transparencias/clase31.pdf

Angelino Feliciano Morales, R. E. (2016). Procesamiento Analítico con Minería de Datos. *Revista Iberoamericana de las Ciencias Computacionales e Informática,* 1-22.

Antonio Monleón Esteban Vegas, F. R. (2017). *Big Data. Hacia la cuarta revolución industrial.* Edicions Universitat Barcelona.

Argonza, J. S. (2006). BIG DATA EN LA EDUCACIÓN. *revista digital unioversitaria*, 1-16.

Bellosta, C. J. (2018). *R para profesionales de los datos: una introducción.* Obtenido de https://datanalytics.com/libro_r/_main.pdf

Bernabeu R. Dario, G. M. (04 de 06 de 2021). *troyanx*. Obtenido de https://troyanx.com/Hefesto/copo-de-nieve.html

Cano, J. (3 de 12 de 2009). *Estrategias BI*. Obtenido de https://enfoquepractico.com/2009/12/03/estrategias-bi/

CESAR PEREZ LOPEZ, D. S. (2008). *Mineria de Datos. Tecnicas y Heramientas.* Madrid, España: Paraninfo.

Commerce, O. o. (2009). *Operaciân del servicio.* Reino Unido: TSO.

Cond, D. (22 de 05 de 2017). *wordpress*. Obtenido de https://datosmineriainformacion.wordpress.com/2017/05/22/que-es-el-proceso-de-kdd-mineria-de-datos-cuales-son-las-etapas-en-que-se-divide-el-proceso/

Crovi, L. y. (2009). *Redes Sociales Análisis y aplicaciones.* Mexico: Plaza y Valdés.: EDIMPRO.

Dario, B. (06 de 05 de 2006). *dataprix*. Obtenido de https://www.dataprix.com/es/data-warehousing-y-metodologia-hefesto/34-datawarehouse-manager

Díaz, J. C. (2012). *Introducción al Business Intelligence.* Barcelona: UOC.

Dijck, J. V. (2019). *La cultura de la conectividad: Una historia crítica de las redes sociales.* Siglo XXI Editores.

DOMÍNGUEZ, D. C. (2010). *Las Redes Sociales. Tipología, uso y consumo de las redes 2.0 en la sociedad digital actual.* Obtenido de https://revistas.ucm.es/index.php/DCIN/article/view/DCIN1010110045A/18656

Donald Córdova, J. D. (2020). El Impacto de Inteligencia de negocios en las Redes Sociales. *Risti*, 14.

Eleazar Botta Ferrer, J. E. (2007). Minería de textos: una herramienta útil para mejorar la gestión del bibliotecario en el entorno digital. *Acimed*, 12.

Enrique Martín, R. C. (2020). *Las bases de Big Data.* Madrid: Los Libros De La Catarata, 2020.

G.SATYANARAYANA REDDY, M. P. (2010). DATA WAREHOUSING, DATA MINING, OLAP. *International Journal on Computer Science and Engineering*, 9.

Gabriela Mancilla-Vela, P. L.-G.-O.-S. (2020). Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en lína: un análisis en minería de datos. *Formación Universitaria*, 14.

Gajardo, D. (24 de 02 de 2019). *https://medium.com/*. Obtenido de https://medium.com/@parasitodigital/lo-b%C3%A1sico-que-necesitas-para-armar-un-dashboard-de-redes-sociales-82e9d67ec389

Garza, D. L. (2015). *REDES SOCIALES INSTRUMENTO DE PARTICIPACION DEMOCRATICA, Analisis de las tecnologias implicadas y nuevas tendencias.* Madrid: DYKINSON, S.L.

Garza, L. M. (2006). *Redes sociales, instrumentos de participación democrática.* Madrid: Dykinson S.L.

Gervilla García, E., Jiménez López, R., Montaño Moreno, J. J., Sesé Abad, A., & Cajal. (2009). La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes. *adicciones, 21*(1), 17. Obtenido de https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

González Ferran, X. R. (2016). *¿Cómo planificar un proyecto de inteligencia de negocio?* Barcelona: Editorial UOC.

Guenther, E. (16 de 04 de 2019). *beneficios del análisis de opinión con Power BI*. Obtenido de https://aptude.com/es/power-bi/entrada/Cinco-beneficios-del-an%C3%A1lisis-de-sentimientos-utilizando-Power-Bi/

Jaime Laviña Orueta, L. M. (2010). *Libro Blanco de la Universidad Digital 2010.* Madrid: Ariel S.A.

Javi Fernández, J. M. (2011). Análisis de Sentimientos y Minería de Opiniones: el corpus EmotiBlog . *Procesamiento del Languaje Natural*, 1-10.

Javi Fernández, Y. G. (2015). Social Rankings: an´alisis visual de sentimientos en redes sociales. *Procesamiento del Lenguaje Natural*, 5.

Jordi Conesa Caralt, J. C. (2010). *inteligencia de negocios.* Carcelona: Editorial UOC.

José C. Riquelme, R. R. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial*, 1- 8.

Juan Antonio Vicente Virseda, J. G. (2019). *METODOS DE DATA SCIENCIE APLICADOS A LA ECONOMIA Y A LA DIRECCION Y ADMINISTRACION DE EMPRESAS.* Madrid: UNED.

Kamagate, A. (2013). OLAP – Online Analytical Processing. *Investigación e Innovación en Ingenierías*, 1-7.

Kdnuggest. (01 de 10 de 2014). *kdnuggets.com*. Obtenido de https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html

Larraín, C. H. (10 de 5 de 2021). Obtenido de https://www.dcc.uchile.cl/sites/default/files/info-cursos/Cc72J.pdf

Lautaro Ramos, E. S. (26 de 04 de 2019). *RIA*. Obtenido de https://ria.utn.edu.ar/bitstream/handle/20.500.12272/4662/WICC2019_Descubrimiento%20de%20conocimiento%20en%20bases%20de%20datos.pdf?sequence=1&isAllowed=y

León, S. T. (2018). *UF2213 - Modelos de datos y visión conceptual de una base de datos.* Elearning, S.L. Obtenido de https://books.google.com.ec/books?id=LV9WDwAAQBAJ&pg=PA229&dq=O

LAP+QUE+ES&hl=es-419&sa=X&ved=2ahUKEwi06JqsgZrxAhVETDABHfg1BwIQ6AEwAXoECAUQAg#v=onepage&q=OLAP%20QUE%20ES&f=false

Leonardo M. Moreno, J. Á. (2019). Análisis de Sentimientos en Redes Sociales (Twitter) Mediante Python para la detección de Oportunidades de Negocio. *Tecnolocultura*, 9.

Libros Cientificos. (2015). *Modelado predictivo para la inteligencia de negocios / Predictive Modeling for Business Intelligence.* EISENBRAUNS.

Mahmoud, Q. H. (2004). *Middleware para comunicaciones.* Copyright.

Manuel Montes, G. (2001). Obtenido de https://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf

Maria Consuelo Justicia de la Torre, D. D. (2017). *Nuevas tecnicas de mineria de texto.Aplicaciones* . Obtenido de file:///C:/Users/Usuario/Desktop/tesis/capitulo%20II%20tesis/233.pdf, https://digibug.ugr.es/handle/10481/46975

María Uvidia Fassler, A. C. (2018). Minería de datos para la toma de decisiones. *Dialnet*, 1-14.

MARQUÉS, M. P. (2015). *BIG DATA - Técnicas, herramientas y aplicaciones.* Mexico: Alfaomega.

Martínez, B. B. (2001). *Minería de datos.* Obtenido de https://www.cs.buap.mx/~bbeltran/NotasMD.pdf

Martínez., M. A. (06 de 2011). *https://dspace.uclv.edu.cu/.* Obtenido de https://dspace.uclv.edu.cu/bitstream/handle/123456789/8879/Tesis%20Final%20Merly%20Arrizabalaga.pdf?sequence=1&isAllowed=y

Mendez, A. M.-M. (2010). *Fundamentos de Data Warehouse.* Obtenido de http://artemisa.unicauca.edu.co/~ecaldon/docs/bd/fundamentosdedatawarehouse.pdf

Meneses, H. (2013). *El Enfoque de Procesos como Requisito para el Éxito de un Proyecto de BI*. Obtenido de https://ciclusgroup.wordpress.com

Microsoft. (09 de 09 de 2019). *Microsoft*. Obtenido de https://docs.microsoft.com/es-es/power-bi/guidance/star-schema

Microsoft. (08 de 05 de 2018). *microsoft*. Obtenido de https://docs.microsoft.com/es-es/analysis-services/data-mining/microsoft-naive-bayes-algorithm?view=asallproducts-allversions

Mostazo, J. M. (1890). *Desarrollo de componentes software y consultas del sistema de almacen de datos.* España: ELEARNING S.L.

MSc. Kelly Deysi Hernández Mite, M. J. (2017). LAS REDES SOCIALES Y ADOLESCENCIAS. REPERCUSIÓN EN LA ACTIVIDAD FÍSICA. *Revista Universidad y Sociedad*, 10-18.

Paola García García, C. A. (08 de 09 de 2008). Obtenido de http://www.it.uc3m.es/jvillena/irc/practicas/08-09/08.pdf

PEREZ LOPEZ, C. S. (2007). *Minería de datos. Técnicas y herramientas.* Madrid: Paraninfo.

Pulido, R. U. (03 de 01 de 2021). *maquinasvirtuales.eu.* Obtenido de https://www.maquinasvirtuales.eu/python-aprendiendo-desde-cero-viii-scripting/

Senén Barro Ameneiro, A. J. (2002). *Fronteras de la computación.* Madrid España: Diaz de Santos.

Sevilla. (2006). *TECNICOS DE SOPORTE INFORMATICO DE LA COMUNIDAD DE CASTILLA Y LEON.* España: Editorial MAD.

Trujillo, A. C. (2006). Modelo Multidimensional. *Ingeniería Industrial*, 5.

Viktor Mayer-Schonberger, K. C. (2013). *Big Data. La Revolución de los Datos Masivos.* Madrid: Houghton Mifflin Harcourt.

WAREHOUSE, D. (10 de 10 de 2017). Obtenido de https://es.slideshare.net/Mardemago/data-warehouse-81078486

Wendy Jahayra Pinargote Mendoza, B. F. (01 de 01 de 2018). *repositorio.uleam.edu.ec.* Obtenido de https://repositorio.uleam.edu.ec/bitstream/123456789/2655/1/ULEAM-INFOR-0071.pdf

ANEXO 1

Titulación

Instructivo del Protocolo de Tesis

Unidad Académica de Tecnologías de la Información y la Comunicación (TIC)

Marzo 2020

Anexo: Formato del Anteproyecto.

A. TÍTULO
Análisis de sentimientos en comunidades digitales utilizando técnicas de BIG DATA para determinar patrones de comportamiento orientado a fenómenos sociales

B. DOMINIO, LÍNEA Y ÁMBITOS DE INVESTIGACIÓN			
Tecnologías de Información y Comunicación	**Ciencias exactas, naturales y tecnológicas**	Inteligencia de Negocios	X
		Sistemas de Información	
		Gobierno y administración de tecnologías de información	
		Auditoría Informática	
		Seguridad Informática	
		Redes y comunicación	
		Arquitectura de Hardware	
		Arquitectura de desarrollo de software	
		Ingeniería de Software	
		Gestión y gobierno de proyectos de tecnología informática	
		Ingeniería de requerimientos	
		Algoritmos y programación	
		Ciencias exactas y naturales (Matemáticas, Física, Química, Biología, etc.)	
		Modelaje y simulación	

C. PLANTEAMIENTO DEL PROBLEMA
El sentimiento es innato en las personas mediante las cuales demuestran un estado anímico hacia hechos o acontecimientos suscitados en el día a día, estos sentimientos son reflejados en las actitudes y acciones de las personas al momento de desarrollar sus actividades diarias. En mucho de los casos en la actualidad debido al gran avance tecnológico los sentimientos se ven reflejados de una manera digital a través de patrones de comportamiento en las diferentes redes sociales. Durante la última década varios actores de los ámbitos políticos, económicos y social utilizan las redes sociales como medio de difusión y análisis, en donde expresan sus ideas hacia la sociedad y estos acontecimientos dan como resultado reacciones en las personas con distintos patrones de comportamiento, estos patrones juegan un papel importante en cada uno de los ámbitos de la sociedad, ya sea como herramienta para la

toma de decisiones empresariales, de los gobiernos seccionales, gobierno nacional, partidos políticos , etc.

El país en el presente año 2020 se encuentra atravesando una de las crisis sanitarias más preocupantes de la última década, la cual ha desencadenado en una crisis económica, política y social. Esta situación ha generado diferentes opiniones y estados anímicos en las personas, las mismas que son reflejadas a través de las redes sociales desde los diferentes puntos de nuestro país; dichos estados emocionales no puedes ser analizados de una manera sencilla por lo que se propone utilizar herramientas de BIG DATA como recurso que ayude a recopilar los diferentes patrones de comportamiento de la ciudadanía en las redes sociales frente a esta crisis.

La herramienta que sea generada en este estudio, permitirá dar las pautas y la metodología a seguir en otros casos de estudio, y la difusión de los resultados generados podrán aportar a la toma de decisiones a los diferentes actores de los ámbitos políticos, sociales y económicos dentro del país.

D. OBJETIVO GENERAL

Desarrollar un método que sea capaz de analizar los sentimientos de los usuario de las diferentes comunidades digitales, con el fin de extraer patrones de comportamiento sobre fenómenos sociales que afectan a la comunidad mediante la explotación de técnicas de big data.

E. OBJETIVOS ESPECÍFICOS

1. Realizar un estudio de estado del arte y revisión bibliográfica sobre el tema
2. Determinar la metodología que permita establecer el proceso para el análisis de datos
3. Diseñar una herramienta la cual analice el proceso de difusión de los sentimientos de los usuarios en las redes sociales de acuerdo a un fenómeno social especifico
4. Analizar los resultados obtenidos a través del big data para conocer el patrón de comportamiento de las personas al utilizar las redes sociales, frente a un fenómeno social especifico.

F. JUSTIFICACIÓN

El análisis de sentimiento es extremadamente útil para monitoreo las opiniones, actitudes, cometarios de los usuarios en las redes sociales ya que permite hacernos una idea de la opinión público general de ciertos temas.

Los sentimientos del usuario son actividades generalmente automatizadas dentro de las redes sociales, por lo que para identificar estos sentimientos se utiliza herramientas de BIG DATA como recurso que permita monitorear el comportamiento de la sociedad dentro de estas comunidades virtuales; considerando cometarios, opiniones, que manifiestan los usuarios, estos sentimientos son de mucha relevancia, por lo cual pocas personas se expresan en distintos medios del mundo digital.

Existen una gran cantidad de herramientas de BIG DATA que permiten el análisis de un gran volumen de datos, obteniendo datos significativos sobre el comportamiento de los datos, en el caso del análisis de sentimientos permite determinar el tono emocional que hay detrás de una palabras determinadas, si una frase contiene una opinión positiva o negativa sobre una fenómeno social, situación económica, institución, organización, empresa, evento o persona los sentimientos, contarios de los usuarios, siendo beneficioso para desarrollar inteligencia de negocios para toma de decisiones de organizaciones públicas, privadas, personas naturales y personas jurídicas .

Actualmente el mundo está atravesando una crisis sanitaria que está dejando consecuencias socioeconómicas incalculables, el Ecuador no es la excepción y esto ha generado un caos e incertidumbre a nuestra sociedad. Debido al gran avance tecnológico de hoy en día, y la gran cantidad de información que se genera en el ecuador a diario; la cantidad de datos es abrumadora, por lo que la sociedad se manifiestan su comportamiento mediante el mundo digital ante esta riesgosa pandemia.

G. ALCANCE

Con el desarrollo del presente proyecto se pretende dar un alcance a todo el territorio ecuatoriano, con la utilización de herramientas tecnológicas de BIG DATA que permita recopilar información acerca de los patrones de comportamiento de las personas frente a la emergencia sanitaria 2020, con la utilización de las redes sociales como fuente de información.

H. CONCEPTOS RELACIONADOS

1. Big Data

Según IBM (2014) cada día se generan más de 1 QB, que surgen de fuentes tan diferentes como los datos de clientes, proveedores, operaciones financieras en línea u obtenidos de dispositivos móviles, análisis de redes sociales, ubicación geográfica mediante GPS. En muchos países se gestionan gigantescas bases de datos que contienen datos de impuestos, censo de población, registros médicos, etc. **[1]**

"Big Data puede ser considerada como una tendencia en el avance de la tecnología que ha abierto la puerta a un nuevo enfoque para la comprensión y la toma de decisiones, que se utiliza para describir las enormes cantidades de datos (estructurados, no estructurados y semi-estructurados) que sería demasiado largo y costoso para cargar una base de datos relacional para su análisis. Así, el concepto de Big Data se aplica a toda información que no puede ser procesada o analizada utilizando herramientas o procesos tradicionales." **[2]**

1.2. Las características fundamentales de BIG DATA

La última característica en añadirse, pero no menos importante es el valor de los datos. Consideramos importante profundizar dentro de estas cinco dimensiones del Big Data para comprender mejor el concepto:

1.2.1. Volumen

Nos referimos a cantidades enormes de datos generadas a cada segundo. No estamos hablando de Terabytes, sino más bien de Zettabits. Hoy en día generamos cada minuto la misma cantidad de datos que los generados en el mundo desde el principio de los tiempos hasta el año 2008 1 o Brontobytes2 [3]. Esto hace que la mayoría de los datos sean muy grandes para ser almacenados y complicados de analizar usando la tecnología actual de bases de datos. Las nuevas herramientas de Big Data y analizar datos a través de bases de datos que están repartidas por todo el mundo.

1.2.2. Variedad

Como comentábamos en un inicio los datos no estructurados se presentan en muy diversos formatos, ya sea vídeo, imágenes, emails, sensores de geolocalización, redes sociales y un amplio etcétera. Antiguamente estábamos centrados únicamente en los datos estructurados que cabían perfectamente en tablas o bases de datos relacionales.

En realidad, el 80% de los datos del mundo se presentan en formatos no estructurados [3]. Es por ello por lo que resulta esencial conocer la información que ese porcentaje de datos nos puede ofrecer. Gracias a diferentes herramientas que se han ido desarrollando para gestionar Big Data podemos analizar y reunir información sobre conversaciones, fotos, vídeos o grabaciones de voz. Además del volumen, esta característica es la que hace que analizar estos tipos de datos sea una ardua tarea.

1.2.3. Velocidad

Al hablar de este término en relación con Big Data nos referimos por una parte a la velocidad con la que se crean datos actualmente y por otro la velocidad de procesamiento y análisis de estos. Con el "internet de las cosas" se puede extraer más información del usuario, ya que éste da información, por ejemplo, acerca de sus gustos cinematográficos y televisivos si hablamos de un televisor con internet; o un "Smartwatch" sabrá todo sobre nuestras rutinas, datos personales, etc. IDC afirma que actualmente hay 13 billones "cosas" conectadas, y estima que en el año 2020 habrá aproximadamente 212 millones en todo el mundo (IDC, 2015). Por otro lado, en las redes sociales podemos percibir a la velocidad que viaja la información, por ejemplo, cuando mensajes o videos se hacen virales en pocos segundos. Teniendo esto en cuenta, la velocidad a la que se producirán los datos en un futuro será titánica y por ello debemos adelantarnos, gestionándolos, transformándolos en información y aportando respuestas rápidas en el momento preciso. La cantidad de segundos que se tarde en procesar los datos, se considera un factor fundamental para marcar diferencias entre empresas.

1.2.4. Veracidad

Se refiere a la yerra o integridad de los datos. Hay estructuras en Big Data, en los que no es posible controlar con exactitud y fiabilidad los datos. Por ejemplo, los Datos de visitas a páginas webs Publicaciones en Twitter Publicaciones en Facebook Contenido web Web y redes sociales Uso de contadores de lectura inteligentes Lectores de RFID (radio frequency identification) Sensores de lectura de plantas petrolíferas Señales GPS Machine-tomachine Reconocimiento facial Genética Biometría Grabaciones de centrales de llamadas Email Expedientes médicos electrónicos Generado por el hombre Reclamaciones de salud Registros de llamadas de telecomunicaciones Registros de facturación de servicios públicos Transacciones de big data Figura 1. Tipos de datos de

Big Data. Fuente: Elaboración propia a partir de Barranco Fragoso (2012). ~ 9 ~ tweets con "hashtags", las abreviaturas, errores tipográficos o el habla coloquial. En la actualidad gracias a la tecnología, aunque no parezca posible, podemos realizar análisis con este tipo de datos, aunque sea complejo. Grandes volúmenes de datos pueden subsanar la falta de calidad o precisión [3].

1.2.5. Valor

Este componente es quizás el más importante. Resulta complicado que las empresas se informaticen al nivel que se necesita el Big Data, y a su vez la rentabilidad de esa inversión (ROI3) deberá ser alta. El valor que se extraiga de los datos depende de la cantidad almacenada de los mismos y su tratamiento, y viceversa. Si conseguimos muchos datos, pero no extraemos valor de ellos no tendremos nada.

2. Patrones de comportamiento

Cuando ciertas reacciones de la persona, se hacen muy frecuentes en determinados ambientes o situaciones, constituyen lo que llamamos un patrón de comportamiento. Un patrón de comportamiento es una forma constante que tiene una persona, de pensar, sentir, reaccionar físicamente y actuar en determinada situación. [4]

Nuestros patrones de comportamiento tienen el siguiente origen:

1. Los copiamos o aprendemos de las personas que han compartido la vida con nosotros: padres, abuelos, tíos, maestros y de cualquier personaje importante con el cual hayamos tenido un contacto significativo a través de la TV, cine, videos, iglesia, paseos, retiros espirituales, etc., sean estos personajes seres humanos, animales o dibujos animados. [4]

2. Provienen de nuestras propias reacciones: esto quiere decir, que guardamos dentro de nosotros las reacciones que tenemos frente a otros o aún ante animales, dibujos animados, películas y frente a la naturaleza (un río, una montaña, la lluvia, truenos, relámpagos, tormentas, huracanes, etc.). También grabamos y guardamos nuestras reacciones cuando satisfacemos o no las necesidades y deseos. Nuestras diversas reacciones frente al hambre, sed, contacto, compañía, afecto, seguridad, protección, etc. [4]

- **PENSAMIENTOS, CREENCIAS E IDEAS.**
- **EMOCIONES Y SENTIMIENTOS E IMÁGENES.**

- **CONDUCTAS.**
- **REACCIONES DEL CUERPO**

Las diferentes reacciones de las personas que son repetitivas o frecuentes en determinados ambientes o situaciones son los llamados patrones de comportamiento. Por su parte Gonzales S. Marcos en su artículo de revista define "un patrón de comportamiento es una forma constante que tiene una persona, de pensar, sentir, reaccionar físicamente y actuar en determinada situación" **[4]**

3. Análisis de sentimientos

Somos seres sensibles. El conflicto, no exento de connotaciones negativas, suele activar en nosotros emociones no siempre deseadas, agitar sentimientos, influir en nuestros estados de ánimo. Estas emociones, sentimientos y estados anímicos de naturaleza sombría suelen tener una mayor intensidad cuando el conflicto surge entre personas que han mantenido relaciones de afectividad sostenidas en el tiempo, como sucede en los conflictos de familia. Si a todo ello se añade la experiencia de la confrontación judicial, un stress añadido resulta inevitable**. [5]**

4.1. **El análisis de opiniones en las redes sociales**

Las redes sociales (comunidades virtuales de comunicación en internet) son una herramienta que permite a las personas comunicarse e interactuar, en ellas se puede escribir o leer opiniones de otros usuarios sobre algún tema de interés. Si bien, la opinión en redes sociales se colocó rápidamente en un lugar privilegiado en la sociedad en general, también despertó el interés por parte de la comunidad académica y empresarial, al llegar con ellas la posibilidad de medir y analizar las opiniones de libre acceso.

El análisis de opinión en las redes sociales, o minería de opinión y análisis de sentimiento aplicado a redes sociales, es un campo joven y que a la fecha se encuentra desarrollo, por lo que, día con día se proponen métodos que 23 permiten un mejor análisis de texto y mejores resultados. No obstante, sigue siendo una tarea difícil debido a la subjetividad que esta tarea implica, e incluso, los seres humanos muchas veces no coincidimos cuando intentamos clasificar una opinión o comentario **[6]**.

4. Detección de sentimientos

Nadie puede querer aquello que no conoce. Por lo tanto, si intentas primero comprender algunos hechos sobre las emociones, entonces te será más sencillo hablar de ellos. Por lo tanto, recuerda que los sentimientos:

- implica una reacción en todo el cuerpo
- están influidos por los pensamientos y las percepciones
- son simples y complejos
- son contagiosos

5. Redes sociales

Las Redes Sociales constituyen una canal para la apropiación social de la Ciencia y la Tecnología que permite entrar en el ADN de la sociedad, logrando de esta forma hacerse parte de ella.

Existen múltiples definiciones y teorías sobre qué son y qué no son las redes sociales, pero existe poco consenso todavía sobre las mismas. La gran mayoría de autores coinciden en que una red social es: "un sitio en la red cuya finalidad es permitir a los usuarios relacionarse, comunicarse, compartir contenido y crear comunidades", o como una herramienta de "democratización de la información que transforma a las personas en receptores y en productores de contenidos". [7]
En el año 2007, fue publicado un artículo en el Journal of Computer Mediated Communication1 que arrojaba interesante información sobre el fenómeno de las redes sociales en Internet. En dicho trabajo se definieron las redes sociales como: "servicios dentro de las webs que permiten al usuario

1) construir un perfil público o semi público dentro de un sistema limitado.

2) articular una lista de otros usuarios con los que comparte una conexión.

3) visualizar y rastrear su lista de contactos y las elaboradas por otros usuarios dentro del sistema. La naturaleza y nomenclatura de estas conexiones suele variar de una red social a otra". [7]

5.1. Facebook

En la actualidad, Facebook es una de las redes sociales más populares en todo el mundo, con la excepción de China. Seguramente ya tienes un perfil en Facebook, porque según la data, para la fecha esta plataforma cuenta con 1860 millones de usuarios.

Por otro lado, Facebook sigue haciendo de los dispositivos móviles su baluarte de crecimiento. Esta empresa ya posee 1230 millones de usuarios activos diarios y, de estos, 1150 millones se conectan al sitio desde un celular.

Esta herramienta social ha funcionado para conectar personas, descubrir y crear nuevas amistades, subir imágenes y compartir contenidos. [8]

¿por qué ha sido tan abrumador el éxito de Facebook?

La sencillez para lograr compartir contenidos a través de links: desde fotos hasta videos.

- La falta de límite para colgar fotografías.
- Su interface sencilla.
- Lo fácil de crearse una cuenta y convertirse en un miembro.
- La posibilidad de comunicarse con los amigos en tiempo real porque se tiene un chat.
- La integración de mensajes y mails.
- Buenas recomendaciones de amigos, las que casi siempre son acertadas.
- Los fans pagos exitosos acerca de negocios, artistas o marcas.
- La posibilidad para los desarrolladores de integrar aplicaciones y ganar dinero con ello

5.2.Twitter

Twitter es una red social de microbloggin, en el que se pueden compartir contenidos en forma de texto con un límite de 140 caracteres. Es una de las plataformas sociales más importantes del mundo en donde la dinámica consiste en seguir y ser seguidos. Uno de sus puntos más fuertes es la posibilidad de compartir noticias en tiempo real mediante cualquier dispositivo, pero es más común que las personas lo hagan a través de sus teléfonos móviles. Muchos usuarios no conciben Internet sin esta herramienta social, he ahí la magnitud de su popularidad. [8]

Muchos de los sitas en Internet incluyen un botón que permite compartir con los seguidores en esta plataforma cualquier página que se considere que pueda resultar interesante y útil, de esa forma adicionalmente se mantiene tu perfil constantemente actualizado. [8]

5.3. Instagram

Es una de las aplicaciones para teléfonos móviles más popular ya que permite editar, retocar, y agregarle filtros a las fotografías que son tomadas desde el dispositivo. Permite compartir estas mismas fotos en otras redes sociales como Facebook y Twitter y ya que logra explorar fotos y compartir contenido con otros usuarios, Instagram es considerada como una red social. Disponible en español y otros muchos idiomas para los dispositivos que usan iOS (iPhone, iPad) y todos los que usan el sistema Android. Hace poco se ha integrado la posibilidad de acceder a una cuenta de Instagram mediante la PC. Así se puede navegar a través del Timeline, dejar comentarios y dar «Likes» a las imágenes, sin tener que usar un dispositivo móvil o una Tablet. [8]

I. TRABAJOS RELACIONADOS

En un estudio realizado por la empresa tecnológica Cisco, entre el 2011 y el 2016 los datos móviles crecerán anualmente un 78% y el número de dispositivos móviles que están conectados a Internet superará la población de la Tierra. Así según un cálculo realizado en 2016 habrá unos 19 mil millones de dispositivos conectados a la red, más de 2 por habitante del planeta; entonces el tráfico global de datos móviles alcanzará 130 EB anuales. Este volumen de tráfico previsto para 2016 equivale a 33 mil millones de DVDs anuales, simplemente inacabables. [9]

En el año 2016, este artículo conceptualiza el término big data y describe su relevancia en la investigación social y las prácticas periodísticas. Explicamos las técnicas de análisis de texto a gran escala, como el análisis de contenido automatizado, la minería de datos, el aprendizaje automático, el modelado de temas y el análisis de sentimientos, que pueden ayudar al descubrimiento científico en las ciencias sociales y la producción de noticias en el periodismo. Explicamos la infraestructura electrónica necesaria para el análisis de big data con el uso de la computación en la nube y evaluamos el uso de los principales paquetes y bibliotecas para la recuperación y análisis de información en software comercial y lenguajes de programación como Python o R. [10]

J. METODOLOGÍA

La presente investigación se encuentra dentro del aspecto cualitativo debido a que se analizará el porcentaje de datos recopilados de BIG DATA de los patrones de comportamiento emocional ante la emergencia sanitaria mundial, esta es la importancia de este trabajo porque se obtendrán datos reales.

Según el nivel, la investigación este trabajo investigativo se encuentra dentro de:

Investigación Descriptiva: Se detallará los datos que demuestren la proporción de la población que presentan el comportamiento emocional en las redes sociales esperando con esta información generar un impacto en la vida de las personas.

DISEÑO DE INVESTIGACIÓN

El trabajo de investigación está enmarcado dentro de los siguientes diseños:

Investigación Documental: esto es fundamental para la elaboración de bases conceptuales que fundamentara el trabajo investigativo la misma que se realiza basándonos en una amplia búsqueda de información en diferentes fuentes bibliográficas.

INSTRUMENTOS DE RECOLECCIÓN DE DATOS

La investigación corresponde a la fundamentación práctica, teórica o bibliográficas de los subtemas que se encuentran en relación al tema de investigación.

Por otras partes también está presente la observación, que es un método científico el cual me permite obtener información y conocimientos acerca del objeto de estudio, de esta manera permite obtener información de forma directa e inmediata sobre el objeto que se está investigando.

Se utilizará la metodología de desarrollo de proyectos de BIG DATA que más se adapte a las necesidades de la presente investigación y que además nos permita asegurarnos del éxito de nuestro proyecto.

HERRAMIENTA PARA EL ANÁLISIS DE LA INFORMACIÓN

El análisis de los datos se fundamentará en la utilización de herramientas de BIG DATA que permitirá el procesamiento y el análisis de los datos para el cumplimiento de las metas propuestas.

K. CRONOGRAMA DE ACTIVIDADES

N°	ACTIVIDAD	MES						MEDIOS DE VERIFICACIÓN
		I	II	III	IV	V	VI	
1	Fundamentación Teórica	x						Primer capítulo de la Tesis (Conceptos Relacionados y Trabajos Relacionados).
2	Diagnóstico Situacional		x					Segundo capítulo de la Tesis (Problema, objetivos, justificación, alcance y aplicación de la metodología propuesta).
3	Desarrollo de la propuesta			X	x			Tercer capítulo de la Tesis.
4	Validación de la propuesta				x	x		Cuarto capítulo de la Tesis.
5	Conclusiones y recomendaciones						x	Sección de conclusiones y recomendaciones de la Tesis.

L. DECLARACIÓN FINAL

Los abajo firmantes declaramos bajo juramento que el proyecto descrito en este documento no ha sido presentado a otra institución nacional o internacional para su financiamiento, no causa perjuicio al ambiente, es de nuestra autoría y no transgrede norma ética alguna.

M. PARTICIPANTES

DIRECTOR:	
ESTUDIANTE 1	Willan Alfredo Llivicota Buñay
ESTUDIANTE 2	

N. FIRMAS DE RESPONSABILIDAD	
Lugar:	
Fecha:	

Firmas:

Nombre: CC: **Director del Proyecto**	Nombre: C.C.: **Estudiante / Egresado**

O. APROBACIÓN

Firmas:

Nombre: CC: **Primer Par Revisor**	Nombre: C.C.: **Segundo Par Revisor**

P. REFERENCIAS

BIBLIOGRAFÍA

[1] IBM, «¿Qué es Big Data?,» 9 12 2013. [En línea]. Available: http://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/.

[2] Moreno, «UNED Facultad de Derecho.,» 2014. [En línea]. Available: http://espacio.uned.es/revistasuned/index.php/RDUNED/article/view/13303/12174.

[3] B. Marr, «The basic idea behind the What is Big Data?,» 2014. [En línea]. Available: https://www.slideshare.net/BernardMarr/140228-big-data-slide-share/3-The_basic_idea_behind_the.

[4] M. Gonzalez S, «Patrones de comportamiento,» *el campamento de DIOS,* pp. 13,14.

[5] A. García-Herrera, «Los sentimientos y las emociones en el proceso de mediación,» *Revista de mediacion ,* vol. 10, nº 1, 2017.

[6] L. Pang, «Opinion mining and sentiment analysis,» Fundation and trends in information retrieval, 2008, pp. 2(1-2),1-135.

[7] S. B. Seidman., «Network structure and minimum degree,» *Social Networks,* vol. 5, nº 3, p. 269–287, 1983.

[8] R. Padilla, «Las Redes Sociales Más Populares,» Raquel Padilla, 2018. [En línea]. Available: https://www.genwords.com/blog/redes-sociales-mas-populares.

[9] Cisco, « Internet será cuatro veces más grande en 2016,» 9 12 2013. [En línea]. Available: http://www.cisco.com/web/ES/about/press/2012/2012-05-30-internet-sera-cuatro-veces-mas-grande-en-2016--informe-vini-de-cisco.html].

[10] E. B.-C. y. F. c.-l. Carlos Arcila-Calderón, «TÉCNICAS BIG DATA,» *ANÁLISIS DETEXTOS A GRAN ESCALA PARA LA INVESTIGACIÓN CIENTÍFICA Y PERIODÍSTICA ,* pp. 623-630, 2016.

[11] M. Campoverde-Molina y L. Valverde, «Accessibility analysis of the web portals of the educational institutions in Cuenca, Ecuador,» *Revista Cátedra,* vol. 2, nº 2, pp. 55-75, 2019.

[12] V. Simbaña-Gallardo y S. Luján-Mora, «Instructions about the manuscript structureof Revista Cátedra,» *Revista Cátedra,* vol. 1, nº 1, pp. 36-52, 2018.

[13] Universidad Católica de Cuenca, «Directrices para autores/as,» 2020. [En línea]. Available: https://killkana.ucacue.edu.ec/index.php/killkana_tecnico/about/submissions.

ANEXO 2

INICIAR SESION EN TWITTER

```
> appname <- "MineriaCovid"

> key <- "pWE6Tg739JAB7xfTUqMKlndnX"

> secret <-"KPjCxkEkE7lb9aTIHJPGSeQeSrjLAomrvfACAIzwIf5paMXMey"

> token<-"202868066-v6mOD2MrqiKw9DazJxZxYogCTyVF2pxS0nVvoVHc"

> tokensecreto<-"pn2zJt1IIjwCdHrlbojCddumYpqr5T7nDI9onEb7QdRKz"

> options(httr_oauth_cache=TRUE

> setup_twitter_oauth(consumer_key = key, consumer_secret = secret,access_token= token, access_secret
= tokensecreto)

[1] "Using direct authentication"

Adding .httr-oauth to .gitignore
```

RECUPERAR TWEETS

```
> covid <- searchTwitter("#covid exclude:retweets", n=3200)

> covid_df <- as_tibble(map_df(covid, as.data.frame))

> write.csv(covid_df, "covid.csv")



> covid <- searchTwitter("#PlanVacunación9100 exclude:retweets", n=3200)

> covid <- searchTwitter("#Lasso exclude:retweets", n=3200)

> lasso_df <- as_tibble(map_df(covid, as.data.frame))

> write.csv(lasso_df, "lasso.csv")
```

```
> tweetslenin<- searchTwitter("@lenin exclude:retweets", n=3200)
> covid <- searchTwitter("#EconomíaEcuador exclude:retweets", n=3200)
> covid <- searchTwitter("#EconomiaEcuador exclude:retweets", n=3200)
> covid <- searchTwitter("EconomiaEcuador exclude:retweets", n=3200)
> covid <- searchTwitter("Economia Ecuador exclude:retweets", n=3200)
> economia_df <- as_tibble(map_df(covid, as.data.frame))
> write.csv(economia_df, "economia.csv")
> lenin_df <- as_tibble(map_df(tweetslenin, as.data.frame))
> write.csv(lenin_df, "lenin.csv")
```

SELECCIÓN DE VARIABLES

```
> tweets <- covid_df %>% select(screenName, created, id, text)
> tweets <- tweets %>% rename(autor = screenName, fecha = created,
+                             texto = text, tweet_id = id)
> head(tweets)
# A tibble: 6 x 4
  autor      fecha               tweet_id     texto
  <chr>      <dttm>              <chr>        <chr>
1 fiona96fm~ 2021-06-17 21:22:58 14056371148~ "Great atmosphere in the @HQ_Cork tod~
2 MPRHCanada 2021-06-17 21:22:56 14056371090~ "News story on some of the potential ~
```

```
3 pwguru65   2021-06-17 21:22:45 14056370601~ "@SKWrestling_ Do I think #AEW will s~

4 RonaldRHa~ 2021-06-17 21:22:44 14056370551~ "New Jersey Rep. Donald Payne  @Donal~

5 Kimberley~ 2021-06-17 21:22:34 14056370148~ "#BREAKING: Health warning issued for~

6 McEwanMor~ 2021-06-17 21:22:26 14056369826~ "Previous #Covid #infection may not o~
```

LIMPIEZA DE LOS DATOS

```
Función para limpiar los datos

> limpiar_tokenizar <- function(texto)

{

+    nuevo_texto <- tolower(texto)

+    nuevo_texto <- str_replace_all(nuevo_texto,"http\\S*", "")

+    # Eliminación de signos de puntuación

+    nuevo_texto <- str_replace_all(nuevo_texto,"[[:punct:]]", " ")

+    # Eliminación de números

+    nuevo_texto <- str_replace_all(nuevo_texto,"[[:digit:]]", " ")

+    # Eliminación de espacios en blanco múltiples

+    nuevo_texto <- str_replace_all(nuevo_texto,"[\\s]+", " ")

+    # Tokenización por palabras individuales

+    nuevo_texto <- str_split(nuevo_texto, " ")[[1]]

+    # Eliminación de tokens con una longitud < 2

+    nuevo_texto <- keep(.x = nuevo_texto, .p = function(x){str_length(x) > 1})

+    return(nuevo_texto)
```

Función de limpieza y tokenización a cada tweet

```
tweets <- tweets %>% mutate(texto_tokenizado = map(.x = texto,
                                                   .f = limpiar_tokenizar))

tweets %>% select(texto_tokenizado) %>% head()
```

```
tweets %>% slice(1) %>% select(texto_tokenizado) %>% pull()
```

Análisis exploratorio

```
> tweets_tidy <- tweets %>% select(-texto) %>% unnest(cols = c(texto_tokenizado))

> tweets_tidy <- tweets_tidy %>% rename(token = texto_tokenizado)

> head(tweets_tidy)
```

Distribución temporal de los tweets

```
> ggplot(tweets, aes(x = as.Date(fecha), fill = autor)) +

+     geom_histogram(position = "identity", bins = 20, show.legend = FALSE) +

+     scale_x_date(date_labels = "%m-%Y", date_breaks = "5 month") +

+     labs(x = "fecha de publicación", y = "número de tweets") +

+     facet_wrap(~ autor, ncol = 1) +

+     theme_bw() +

+     theme(axis.text.x = element_text(angle = 90))
```

ANALISIS DE LOS DATOS

```
> lista_stopwords <- c('me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves',
+                      'you','your', 'yours', 'yourself', 'yourselves', 'he', 'him','his',
+                      'himself', 'she', 'her', 'hers', 'herself', 'it', 'its', 'itself',
+                      'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which',
+                      'who', 'whom', 'this', 'that', 'these', 'those', 'am', 'is', 'are',
+                      'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had',
+                      'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and',
+                      'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at',
+                      'by', 'for', 'with', 'about', 'against', 'between', 'into',
+                      'through', 'during', 'before', 'after', 'above', 'below', 'to',
+                      'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under',
+                      'again', 'further', 'then', 'once', 'here', 'there', 'when',
+                      'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more',
+                      'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own',
+                      'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will',
+                      'just', 'don', 'should', 'now', 'd', 'll', 'm', 'o', 're', 've',
+                      'y', 'ain', 'aren', 'couldn', 'didn', 'doesn', 'hadn', 'hasn',
+                      'haven', 'isn', 'ma', 'mightn', 'mustn', 'needn', 'shan',
+'shouldn', 'wasn', 'weren', 'won',
'wouldn','i','el','los','son','la','ella','también','si','de','en','que','para','las','un','por','con','del','new','se','le','il','di',
'les','us','get','una','al','get','da','des')
```

```
> tweets_tidy_mes_anyo <- tweets_tidy_mes_anyo %>% filter(!(token %in% lista_stopwords))
>    tweets_tidy_mes_anyo %>% group_by(mes_anyo, token) %>% count(token) %>%
group_by(mes_anyo) %>%
+    top_n(20, n) %>% arrange(mes_anyo, desc(n)) %>%
+    ggplot(aes(x = reorder(token,n), y = n, fill = mes_anyo)) +
+    geom_col() +
+    theme_bw() +
+    labs(y = "", x = "") +
+    theme(legend.position = "none") +
+    coord_flip() +
+    facet_wrap(~mes_anyo,scales = "free", ncol = 1, drop = TRUE)
```

```
library(wordcloud)
library(RColorBrewer)


wordcloud_custom <- function(grupo, df){
  print(grupo)
  wordcloud(words = df$token, freq = df$frecuencia,
            max.words = 400, random.order = FALSE, rot.per = 0.35,
            colors = brewer.pal(8, "Dark2"))
```

```
}


df_grouped <- tweets_tidy_mes_anyo %>% group_by(mes_anyo, token) %>% count(token) %>%
        group_by(mes_anyo) %>% mutate(frecuencia = n / n()) %>%
        arrange(mes_anyo, desc(frecuencia)) %>% nest()


walk2(.x = df_grouped$autor, .y = df_grouped$data, .f = wordcloud_custom)
```

BIGRAMAS

```
library(tidytext)
limpiar <- function(texto){
  # El orden de la limpieza no es arbitrario
    # Se convierte todo el texto a minúsculas
    nuevo_texto <- tolower(texto)
    # Eliminación de páginas web (palabras que empiezan por "http." seguidas
    # de cualquier cosa que no sea un espacio)
    nuevo_texto <- str_replace_all(nuevo_texto,"http\\S*", "")
    # Eliminación de signos de puntuación
    nuevo_texto <- str_replace_all(nuevo_texto,"[[:punct:]]", " ")
    # Eliminación de números
    nuevo_texto <- str_replace_all(nuevo_texto,"[[:digit:]]", " ")
    # Eliminación de espacios en blanco múltiples
```

```
  nuevo_texto <- str_replace_all(nuevo_texto,"[\\s]+", " ")

  return(nuevo_texto)

}

bigramas <- tweets %>% mutate(texto = limpiar(texto)) %>%

        select(texto) %>%

        unnest_tokens(input = texto, output = "bigrama",

                token = "ngrams",n = 2, drop = TRUE)



# Contaje de ocurrencias de cada bigrama

bigramas  %>% count(bigrama, sort = TRUE)
```

```
# A tibble: 34,439 x 2

  bigrama          n

  <chr>        <int>

 1 covid covid     135

 2 of the          84

 3 de la           83

 4 in the          83

 5 de covid        78

 6 of covid        77

 7 the covid       77
```

```
 8 covid vaccine    75

 9 el covid         67

10 covid en         60

# ... with 34,429 more rows
```

```
# Separación de los bigramas

bigrams_separados <- bigramas %>% separate(bigrama, c("palabra1", "palabra2"),

                                           sep = " ")

head(bigrams_separados)
```

```
# A tibble: 6 x 2

  palabra1   palabra2

  <chr>      <chr>

1 great      atmosphere

2 atmosphere in

3 in         the

4 the        hq

5 hq         cork

6 cork       today
```

```
# Filtrado de los bigramas que contienen alguna stopword
```

```
bigrams_separados <- bigrams_separados  %>%
                filter(!palabra1 %in% lista_stopwords) %>%
                filter(!palabra2 %in% lista_stopwords)


# Unión de las palabras para formar de nuevo los bigramas
bigramas <- bigrams_separados %>%
          unite(bigrama, palabra1, palabra2, sep = " ")


# Nuevo contaje para identificar los bigramas más frecuentes
bigramas  %>% count(bigrama, sort = TRUE) %>% print(n = 20)
```

```
  bigrama              n
  <chr>            <int>
 1 covid covid        135
 2 covid vaccine       75
 3 covid coronavirus   43
 4 coronavirus covid   40
 5 covid cases         40
 6 covid vaccines      37
 7 covid pandemic      35
 8 contra covid        34
 9 covid vaccination   28
```

```
10 covid jun            27

11 siguientes medidas   27

12 covid lockdown       25

13 covid corona         22

14 nuevos casos         22

15 covid awareness      21

16 buy domain           17

17 covid patients       17

18 domain covid         17

19 domains buy          17

20 blood covid          16

# ... with 17,365 more rows
```

GRÁFICO DE LA RELACIÓN DE PALABRAS

```
library(igraph)

library(ggraph)

graph <- bigramas %>%

    separate(bigrama, c("palabra1", "palabra2"), sep = " ") %>%

    count(palabra1, palabra2, sort = TRUE) %>%

    filter(n > 18) %>% graph_from_data_frame(directed = FALSE)

set.seed(123)



plot(graph, vertex.label.font = 2,
```

```
    vertex.label.color = "black",
    vertex.label.cex = 0.7, edge.color = "gray85")
```

```
ggraph(graph = graph) +
  geom_edge_link(colour = "gray70")  +
  geom_node_text(aes(label = name), size = 4) +
  theme_bw()
```

FRECUENCIA DE UN TERMINO

```
# Número de veces que aparece cada término por tweet
tweets_tf <- tweets_tidy_mes_anyo %>% group_by(tweet_id, token) %>% summarise(n = n())


# Se añade una columna con el total de términos por tweet
tweets_tf <- tweets_tf %>% mutate(total_n = sum(n))


# Se calcula el tf
tweets_tf <- tweets_tf %>% mutate(tf = n / total_n )
head(tweets_tf)
```

RESULTADOS

```
# A tibble: 6 x 5
# Groups:   tweet_id [1]
  tweet_id           token         n total_n    tf
  <chr>              <chr>     <int>   <int> <dbl>
1 1405531169265840135 class         1       8 0.125
2 1405531169265840135 closed        1       8 0.125
3 1405531169265840135 computed      1       8 0.125
4 1405531169265840135 entire        1       8 0.125
5 1405531169265840135 justasking    1       8 0.125
6 1405531169265840135 marks         1       8 0.125
```

Frecuencia de documento inversa

```
total_documentos = tweets_tidy_mes_anyo$tweet_id %>% unique() %>% length()
total_documentos
[1] 3198
```

```
# Número de documentos en los que aparece cada término
tweets_idf <- tweets_tidy %>% distinct(token, tweet_id) %>% group_by(token) %>%
summarise(n_documentos = n())
```

```
# Cálculo del idf

tweets_idf <- tweets_idf %>% mutate(idf = n_documentos/ total_documentos) %>%arrange(desc(idf))

head(tweets_idf)
```

```
# A tibble: 6 x 3

  token       n_documentos    idf

  <chr>              <int>  <dbl>

1 coronavirus          143 0.0447

2 vaccine              136 0.0425

3 people                94 0.0294

4 contra                89 0.0278

5 pandemic              88 0.0275

6 cases                 81 0.0253
```

Término Frecuencia - Frecuencia inversa del documento

```
token      |  n| total_n| n_documentos|       idf|    tf_idf|

|:----------|--:|-------:|------------:|---------:|---------:|

|class      |  1|       8|            4| 0.0012508| 0.0001563|

|closed     |  1|       8|            4| 0.0012508| 0.0001563|
```

|computed | 1| 8| 1| 0.0003127| 0.0000391|

|entire | 1| 8| 2| 0.0006254| 0.0000782|

|justasking | 1| 8| 2| 0.0006254| 0.0000782|

|marks | 1| 8| 2| 0.0006254| 0.0000782|

|schools | 1| 8| 11| 0.0034396| 0.0004300|

Willan Alfredo Llivicota Buñay portador(a) de la cédula de ciudadanía Nº **0302907068**. En calidad de autor/a y titular de los derechos patrimoniales del trabajo de titulación **"Título del trabajo"** de conformidad a lo establecido en el artículo 114 Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación, reconozco a favor de la Universidad Católica de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos y no comerciales. Autorizo además a la Universidad Católica de Cuenca, para que realice la publicación de éste trabajo de titulación en el Repositorio Institucional de conformidad a lo dispuesto en el artículo 144 de la Ley Orgánica de Educación Superior.

Cañar, **15 de octubre de 2021**

F: 

……..

Willan Alfredo Llivicota Buñay

C.I. **0302907068**