



Impacto de la red social Twitter en la imagen de Colfondos S.A.

Leidy Jovana Aragón Chaverria

William Andrés Vargas Pérez

Corporación Universitaria Minuto de Dios

Rectoría Cundinamarca

Sede Madrid (Cundinamarca)

Programa Especialización en Big Data

Junio de 2022

Impacto de la red social Twitter en la imagen de Colfondos S.A.

Leidy Jovana Aragón Chaverria

William Andrés Vargas Pérez

Trabajo de Grado presentado como requisito para optar al título de Especialista en Big Data

Tutor

Efrén Eduardo Rojas Burgos

Máster en Investigación de Mercados

Corporación Universitaria Minuto de Dios

Rectoría Cundinamarca

Sede Madrid (Cundinamarca)

Programa Especialización en Big Data

junio de 2022

## **Dedicatoria**

*Este trabajo lo dedicamos principalmente a Dios por permitirnos emprender y culminar de manera satisfactoria y sobresaliente este desafío, a nuestros Padres por su apoyo incondicional y motivación constante para lograr alcanzar este reto en nuestras vidas académicas. A nuestra hija que es nuestro motor, por su paciencia y comprensión durante este proceso y a todas aquellas personas que formaron parte y aportaron un grano de arena para alcanzar esta meta.*

*Leidy Jovana Aragón Chaverria*

*William Andrés Vargas Pérez*

## Contenido

Lista de figuras.....	6
Resumen.....	7
Abstract.....	8
Introducción .....	9
Capítulo I. Generalidades del estudio .....	10
1.1 Planteamiento del Problema .....	10
1.2 Formulación del Problema .....	11
1.3 Justificación.....	12
1.4 Objetivos .....	13
1.4.1 Objetivo General.....	13
1.4.2 Objetivos Específicos .....	13
Capítulo II. Marco de Referencia.....	14
2.1 Estado del Arte .....	14
2.2 Marco Teórico .....	16
2.3 Marco conceptual .....	18
2.4 Marco legal.....	19
Capítulo III. Marco Metodológico .....	22
3.1 Alcance de la investigación.....	22
3.2 Enfoque metodológico .....	22
3.3 Población y Muestra.....	23
3.4 Instrumentos .....	23
3.5 Análisis de Información .....	24

3.6 Condiciones Éticas .....	24
3.7 Desarrollo .....	25
3.9 Análisis de sentimientos.....	25
3.10 Exploración de datos .....	32
3.11 Métricas generales de Twitter .....	40
Conclusiones .....	42
Referencias.....	43

## Lista de figuras

Figura 1 Cambio en el tiempo diario dedicado al uso de las redes sociales enero de 2022	
DataReportal, años 2013 al 2021 (26/01/2022) .....	11
Figura 2 Pasos para el análisis de sentimientos .....	26
Figura 3 Metodología de recolección de información .....	27
Figura 4 Proceso de recolección de tweets en Google Drive.....	28
Figura 5 Almacenamiento de datos en Mongo DB.....	29
Figura 6 Preparando Datos (campo fecha).....	30
Figura 7 Convertir tweets en palabras .....	31
Figura 8 Palabras por Categoría.....	32
Figura 9 Nube de Palabras .....	33
Figura 10 Palabras Positivas por Categoría .....	34
Figura 11 Palabras Negativas por Categoría.....	35
Figura 12 Tendencia por categoría .....	36
Figura 13 Proporción negativa y positiva de los tweets .....	37
Figura 14 Visual Tablero de control por Tweet.....	38
Figura 15 Visual Tablero de control Análisis de sentimientos por Palabras .....	39
Figura 16 Visual Tablero de control Nube de Palabras .....	40
Figura 17 Metricas de Twitter.....	41

## **Resumen**

En la última década se ha incrementado el interés por el manejo de grandes cantidades de información y la extracción de conocimiento por medio de las redes sociales, las cuales han jugado un papel muy importante en la sociedad debido a la fácil interacción e intercambio de opiniones acerca de un tema determinado. Esto hace de estas plataformas una gran fuente de información para las empresas, aprovechar dichos datos e interpretarlos de una manera ágil y eficiente; todo ello con el fin de mejorar sus estrategias de mercadeo y optimizar la toma de decisiones.

Este proyecto se encuentra enmarcado dentro de la minería de texto y el procesamiento del lenguaje natural; cuyo objetivo será desarrollar un modelo de clasificación para el análisis de sentimientos en Twitter, permitiendo conocer la percepción de los usuarios de esta red acerca del sistema pensional colombiano y los fondos privados de pensiones del país.

Finalmente se muestran los resultados del análisis por medio de un tablero de control donde se complementa con métricas utilizadas para el seguimiento regular de la red social.

***Palabras Clave:*** Twitter, Análisis de sentimientos, Aprendizaje automático, Pensiones, Big Data.

### **Abstract**

In the last decade, there has been an increase in interest in managing large amounts of information and extracting knowledge through social networks, which have played a very important role in society due to the easy interaction and exchange of opinions about of a certain topic. This makes these platforms a great source of information for companies, leading them to take advantage of data and interpret it in an agile and efficient manner, in order to improve their marketing strategies and optimize decision-making.

This project is framed within text mining, natural language processing and machine learning; whose objective will be to develop a classification model that allows an analysis of sentiments on Twitter, allowing to know the perception of the users of this network about the Colombian pension system and the country's private pension funds.

Finally, the results of the analysis are shown through a control panel where they are complemented with metrics used for regular monitoring of the social network.

***Keywords:*** *Twitter, Sentiment analysis, Machine learning, Pensions, Big Data.*



## **Introducción**

Actualmente las redes sociales juegan un papel importante en la sociedad, según el Global Digital Reports en enero de 2022 existen 4.620 Millones de usuarios de redes sociales en todo el mundo, esta cifra corresponde al 58.4% del total de la población mundial, igualmente señalan que los usuarios globales de redes han crecido más del 10% en el último año, con 424 millones de nuevos usuarios que empezaron a utilizar las redes sociales durante 2021, (Hall, 2022). Las redes se han convertido en la principal fuente de información lo que hace fundamental por parte de las compañías tomar estos datos y analizar todas las opiniones tanto positivas como negativas que puedan llegar a influenciar en la decisión de un consumidor y en el posicionamiento de la marca, (Barrio, 2017).

Según el informe de gestión enero-diciembre 2020 Colfondos S.A. Pensiones y Cesantías destaca la profundización en Instagram, en donde se alcanzaron casi 2.000 seguidores, explorando un público más joven, con temas relacionados con educación financiera; así mismo resalta la gestión en Facebook, en donde aumentó la interacción orgánica en un 285% con respecto al año anterior; finalmente se presenta una mejora en los niveles de engagement de redes como Facebook, Instagram y Twitter.

Este proyecto de basa principalmente en el análisis de sentimiento de la red Social Twitter, de la Administradora de Fondos de Pensiones privados Colfondos y el sistema pensional colombiano, información que permite identificar la percepción de los usuarios tanto positiva como negativa. Dicha información se visualiza mediante un tablero de control permitiendo realizar un seguimiento a nivel estratégico enfocado al mejoramiento de la imagen de la compañía.

## **Capítulo I. Generalidades del estudio**

### **1.1 Planteamiento del Problema**

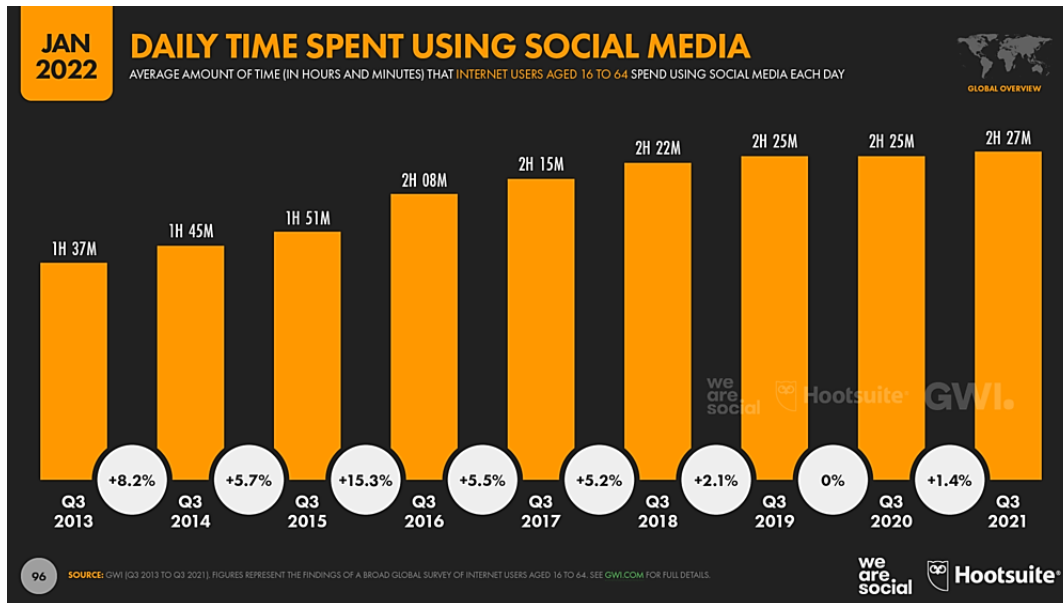
Las sociedades administradoras de pensiones y cesantías tienen una gran particularidad y es que dependen netamente del entorno en el que se mueven.

Un acto de cualquier tipo puede llevar a que el mercado se mueva abruptamente, lo que hace que este sea complejo. Externalidades como las elecciones políticas, los efectos del invierno, el terrorismo, las malas expectativas sobre un mercado, país o divisa, entre otras, pueden llegar a mover el mercado y, por lo tanto, las compañías del sector se pueden ver altamente afectadas. La completa y constante incertidumbre sobre los mercados hace que el terreno en el que se desenvuelve una AFP sea hostil. Cada segundo se realizan cambios que deben ser empleados de la mejor forma por los encargados del área para cumplir con los requerimientos de la compañía. (Jiménez et al., 2011, p.54)

Estos cambios también deben estar enfocados en las redes sociales, las cuales se han convertido en las herramientas más utilizadas por las personas consumiendo la mayor parte de su tiempo libre según el Global Digital Reports de 2022 revela que el tiempo promedio que pasamos usando las “redes sociales es de 2 horas y 27 minutos por día, presentando un incremento de 2 minutos por día (+1.4%)” (Hall, 2022, p.1).

## Figura 1

Cambio en el tiempo diario dedicado al uso de las redes sociales enero de 2022 DataReportal, años 2013 al 2021 (26/01/2022).



Nota. Nos muestra los tiempos dedicados a las redes sociales. Fuente. Simón Kemp. Global Digital Reports

A raíz de este escenario nace la necesidad de aprovechar la información de las redes y realizar un seguimiento y análisis a la opinión de los usuarios de twitter. El problema de la investigación se origina del poco conocimiento acerca del alcance que puedan llegar a tener las redes sociales en el fondo de pensiones y cesantías Colfondos, así como la reputación de la marca y de la competencia (Fondos de pensiones privados).

### 1.2 Formulación del Problema

¿Cuál es el nivel de percepción de los usuarios de la red social Twitter acerca del fondo de pensiones y cesantías Colfondos y el sistema pensional en Colombia?

### **1.3 Justificación**

Las oportunidades presentadas en las redes sociales para generar información son múltiples y representan maneras eficaces para que los fondos de pensiones privados en Colombia mejoren su imagen corporativa y promocionen sus productos y servicios (Guzmán, 2013). Las AFP pueden encontrar en las redes sociales como Twitter, grandes vías de comunicación con los usuarios de una manera clara y precisa, recolectando información relacionada con sus necesidades y expectativas.

Durante el año 2020 la Administradora de Fondo de Pensiones Colfondos adelantó campañas de comunicación digital buscando orientar a los usuarios de acuerdo con la etapa de vida en la que se encuentran. Igualmente, en conjunto con ASOFONDOS se desarrolla la campaña “Mata el Mito” la cual atreves de la figura de Mitos y verdades sobre las pensiones, busca explicar de manera fácil y practica las inquietudes frente a temas pensionales en Colombia, mitigando la desinformación que se genera en redes afectando la toma de decisiones financieras y pensionales de manera objetiva, (COLFONDOS, 2020).

Debido a que la empresa no cuenta con un seguimiento y control de una red social como Twitter, se hace necesario la creación de una herramienta de visualización que permita analizar el comportamiento de los usuarios, generar indicadores de control y conocer la percepción general de la compañía, la competencia y el sistema pensional en Colombia; brindando así información que permita identificar oportunidades de mejora en el manejo de las redes sociales y la toma de decisiones que puedan afectar el área comercial de COLFONDOS.

## **1.4 Objetivos**

### ***1.4.1 Objetivo General***

Analizar patrones y tendencias de las publicaciones realizadas en Twitter donde se mencionen las AFP privadas en Colombia, con el fin de aportar herramientas a las estrategias de mercadeo y fidelización de Colfondos.

### ***1.4.2 Objetivos Específicos***

- Evaluar los tweets donde sea mencionada la AFP Colfondos.
- Observar el contenido de la industria y los competidores directos (AFP Privadas) en Twitter por medio de la escucha social.

## Capítulo II. Marco de Referencia

### 2.1 Estado del Arte

Los antecedentes referidos en este estudio van orientados al uso del Big Data aplicado a las redes sociales y el manejo que le pueden llegar a dar las organizaciones; según la bibliografía revisada se destacan:

En el artículo de Galeano y Peña (2019) “las nuevas oportunidades del Big Data para las instituciones financieras” concluyen que las instituciones financieras alcanzan mejorar el entorno diseñando las estrategias para la recolección y análisis de información que generan sus clientes y así utilizarla para ajustar más sus políticas comerciales a las necesidades de estos.

Por ejemplo, el análisis detallado ingresos y gastos en la cuenta de cada cliente puede anticipar cambios en sus comportamientos de pago e incluso puede establecer su probabilidad de abandono permitiendo proponer estrategias para mejorar su comportamiento financiero. Además, esta información puede enriquecerse con los datos que se encuentran disponibles sobre los clientes en las redes sociales, mejorando su segmentación y permitiendo mejores predicciones sobre su actividad. Un uso inteligente y controlado de toda esta información puede aumentar la fidelidad de los clientes con una institución, mejorar su imagen y ubicarlo en una posición más sólida ante la competencia. (Galeano y Peña, 2019, p.18)

En el artículo de Romero et al. (2020) denominado “Técnicas para la Clasificación de Sentimientos en Redes Sociales como Apoyo en el Marketing Digital” resumen que:

Las redes sociales han cambiado la manera de comunicarnos, se han convertido en una fuente significativa de información y de expresión. Se trata de una comunicación rápida, libre y universal; que permite a sus usuarios una participación según el tema que les

interese. El contenido que se genera en las redes sociales es extenso y ya que se ha convertido en una importante fuente de información, requiere ser analizada y explorada mediante la aplicación de técnicas para la clasificación de sentimientos; con el propósito de encontrar tendencias en el comportamiento de los usuarios sirviendo de apoyo a las organizaciones para el fortalecimiento de las acciones relacionadas con el marketing digital. (Romero et al., 2020, p.1)

Sobrino (2018) en su trabajo de maestría “Análisis de sentimientos en Twitter” dice que El PNL es una disciplina donde su finalidad es mejorar y hacer eficaz la comunicación entre maquina y personas, por lo que cualquier área asociada al lenguaje humano y las relaciones entre humanos y computadores se puede ver afectada de manera positiva por el PNL. Aunque esta disciplina tiene una gran cantidad de aplicaciones y el único límite es la imaginación en esta ocasión se hablará del Análisis de sentimientos; el cual ofrece una manera de conocer automáticamente la opinión que una persona tiene expresada en un texto acerca de un tema determinado, este sentimiento es una valoración cualitativa o cuantitativa acerca de un producto o servicio.

Por su parte Selva y Caro (2016) en su artículo “Uso de datos en la creatividad publicitaria: el caso de Art, Copy & Code de Google” indican que:

Las posibilidades de captar datos en tiempo real de un Usuario en redes permiten una personalización de productos o servicios que podría mejorar significativamente la eficacia de las campañas, tanto de captura como de fidelización de clientes; algo que hace algunos años habría resultado poco viable, todo esto se puede llevar a cabo mediante la cooperación entre diversos productos y tecnología. (p.9)

Finalmente, López (2012-2013) en su trabajo de maestría “Análisis de las posibilidades de uso de Big Data en las organizaciones” menciona que:

Compañías como Facebook, Twitter, Shazam o Spotify se han convertido imbatibles gracias la tecnología del Big Data, igualmente se han mostrado muchas técnicas relacionadas con dicha tecnología, Igualmente señala que métodos relacionados con Business Intelligence, Data Warehouse y Data Mining no deben de ser tratados como tecnologías diferentes, sino como una evolución necesaria para la “era de la información”. Por último, a pesar de que hoy en día se generan enormes cantidades de datos, estas cantidades pronto se convertirán en ínfimas. Es por esto por lo que la tecnología Big Data debe seguir evolucionando para de este modo aprovechar el gran océano de datos, sin olvidar que no todos son útiles, teniendo en cuenta que existe mucho ruido entre ellos, siendo necesario destacar la importancia de captar únicamente aquellos que puedan ser transformados en información y conocimiento. (p.69)

## **2.2 Marco Teórico**

Se puede decir que el termino Big data no es nuevo, manejar grandes conjuntos de datos se remonta a las décadas de los 60s y 70s, cuando se empezaron a crear los famosos centros de datos y las bases relacionales (SQL), pero es en 2005 cuando el mundo comienza a darse cuenta de la cantidad de datos que los usuarios generaban con el uso de las redes sociales, YouTube y otros servicios en línea. Según la página de ORACLE (s.f):

El desarrollo aplicaciones tales como Hadoop y, más recientemente, Spark fueron esenciales para el crecimiento del Big Data, pues estos hacían que su uso fuera más fácil y con un almacenamiento más económico. En los últimos años, el volumen de Big data se ha disparado, los usuarios continúan generando grandes cantidades de datos, pero ahora



los usuarios no son los únicos que lo hacen. Con la llegada del Internet de las Cosas (IoT), hay una gran cantidad de objetos y dispositivos conectados a Internet que generan datos sobre el uso de los clientes y el rendimiento de los productos. (pp.2-3)

Igualmente es importante mencionar el nacimiento de la web 2.0 o también conocida como la web participativa, podemos decir que es el sistema que permite a los usuarios interactuar y colaborar entre sí; es entonces cuando los usuarios comienzan a buscar opiniones de otras personas, se presenta la participación en foros de opinión, se comparten experiencias de compra y se realizan recomendaciones a los demás usuarios. Esto hace que la participación de los usuarios sea más activa tanto en el proceso de búsqueda de información como en la recomendación a otros clientes causando un mayor poder de influencia, ya que pueden criticar o recomendar cualquier producto o servicio siendo esta una de las grandes ventajas de este sistema permitiendo generar valoraciones tanto positivas como negativas, (Sicilia et al., 2021).

Sin embargo, es con el nacimiento de las redes sociales que se presenta una verdadera revolución en el comercio electrónico, ya que se encuentran desarrolladas en torno a los usuarios buscando fomentar el contacto con ellos aprovechando la información de sus perfiles, como textos fotos o videos, al igual que los amigos y contactos con los que interactúan. El uso del Big data en las redes sociales es una gran oportunidad para las empresas de adaptarse a una nueva forma de hacer negocios, se puede utilizar para saber lo que los clientes opinan acerca de la marca e igualmente este resultado sirve para averiguar cómo esta información influye en las decisiones y la forma en que una empresa se comporta; es decir se puede determinar qué factores están impactando en ventas, aceptación de campañas y exactitud del marketing. Es importante llegar a la base de porque los clientes se comportan de una manera determinada, por ejemplo, el

análisis de los tweets es un indicador revelador sobre el impacto del sentimiento del cliente acerca del producto o la empresa. (Pérez, 2015).

### **2.3 Marco conceptual**

**Big Data:** Este campo científico es el que nos da las pautas metodológicas para gestionar grandes volúmenes de datos con el fin de crear valor mediante una serie de procesos y procedimientos. Pero, además, hemos de tener en cuenta que debemos contar con la tecnología, para capturar datos, procesarlos, analizarlos e interpretarlos de manera rápida y eficiente (Gómez y Conesa, 2015 como se citó en García, 2017, p.9).

**Sociedad administradora de fondos de pensiones y de cesantía – AFP:** instituciones financieras de carácter previsional, vigiladas por la Superintendencia Financiera de Colombia, cuyo objeto exclusivo es la administración y manejo de fondos y planes de pensiones del Régimen de Ahorro Individual con Solidaridad y de fondos de cesantía. (Secretaría Distrital de Planeación, 2022, p.1)

**Monitorización de medios:** Hablando de medios sociales, escuchar y supervisar lo que se dice sobre tu marca y sector tanto en canales propios (tu blog, página de Facebook, perfil de Twitter...) como en ajenos, es monitorizar, (Ibáñez, 2014, p.35)

**Twitter:** Creada en 2006 por Jack Dorsey, es lo que se denomina una plataforma de microblogging, donde los usuarios publican mensajes de no más de 280 caracteres denominados tuits. Esta red se caracteriza por la inmediatez, y el etiquetado de palabras claves mediante el símbolo # nos da una imagen de los temas del momento, (Ibáñez, 2014, p. 92).

**El aprendizaje automático:** Es una rama de la inteligencia artificial dedicada al desarrollo de técnicas que permitan a los ordenadores aprender. Se entiende por aprender la

capacidad de generalizar comportamientos a partir de información no estructurada suministrada como ejemplo. Es decir, se pretende que un ordenador sea capaz de reconocer patrones similares dentro de un conjunto determinado, (Bishop, 2010 como se citó en Rodríguez y Camacho, 2018, p.7).

Tipos de algoritmos y enfoques: A partir de una serie de datos, utilizando algoritmos de aprendizaje automático, es posible interpretar la información para transformarla en conocimiento. Algunas de las formas de representación del conocimiento pueden ser las tablas de decisión, los árboles de decisión, las reglas de clasificación o los clusters, entre otros (Witten & Frank, 2005, pp.61-82 como se citó en Rodríguez y Camacho, 2018, p.7).

#### **2.4 Marco legal**

En el ámbito nacional el gobierno y diferentes entes de control han creado leyes y normatividad para el uso de datos personales y protección de la información, a continuación, se presenta una recopilación de algunos referentes normativos alineados al objeto del proyecto.

Constitución Política de Colombia 1991. Artículo 15: Todas las personas tienen derecho a su intimidad personal y familiar y a su buen nombre, y el Estado debe respetarlos y hacerlos respetar. De igual modo, tienen derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en los bancos de datos y en archivos de entidades públicas y privadas. En la recolección, tratamiento y circulación de datos se respetarán la libertad y demás garantías consagradas en la Constitución. La correspondencia y demás formas de comunicación privada son inviolables. Sólo pueden ser interceptados o registrados mediante orden judicial, en los casos y con las formalidades que establezca la ley, (Constitución Política Nacional, 1991, p.1).

La Ley 1266 de 2008, también conocida como Ley de Habeas Data, se aplica a todos los datos personales financieros, crediticios, comerciales y de servicios registrados en un banco de datos. En este sentido, la aplicación de la Ley 1266 de 2008 está encaminada a regular el uso de esa información y por tanto otro tipo de datos (por ejemplo, aquellos mantenidos en un ámbito exclusivamente personal o doméstico o los que se incluyen en una historia clínica) se encuentran excluidos de la aplicación de esta norma, (Superintendencia de Industria y Comercio, 202, p.1)

Ley 1273 de 2009: Por medio de la cual se modifica el Código Penal, se crea un nuevo bien jurídico tutelado - denominado “de la protección de la información y de los datos”- y se preservan integralmente los sistemas que utilicen las tecnologías de la información y las comunicaciones, entre otras disposiciones, (Policía Nacional de Colombia, 2022, p.1)

Ley 1581 de 2012: Desarrolla el derecho constitucional que tienen todas las personas a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos o archivos, definiendo su ámbito de aplicación, sus principios rectores, las categorías especiales de datos, el procedimiento para el tratamiento de los datos, los derechos de los titulares de la información, las situaciones en que es y no es necesaria la autorización del titular, el deber de información que debe ejercerse frente al titular, el proceso para la corrección de información cuando sea necesario, los deberes de los responsables y/o encargados del tratamiento, así mismo la autoridad que velará por la vigilancia y sanción de los temas reglamentados en la presente Ley será la Superintendencia de Industria y Comercio, finalmente determina como operará la transferencia de datos a terceros países, (Alcaldía de Bogotá, s.f, p.1).

Ley 1712 de 2014: El derecho al acceso a la información pública, es una prerrogativa de orden constitucional dispuesta en los artículos 20, 23 y 74 de la Constitución Política, en la que se establece de manera expresa que todo ciudadano tiene derecho a recibir información pública, presentar peticiones respetuosas y acceder a documentos públicos salvo las excepciones previstas en la ley, (Secretaría de transparencia, s.f, p.1).

### **Capítulo III. Marco Metodológico**

#### **3.1 Alcance de la investigación**

El presente estudio tendrá un alcance de tipo descriptivo, teniendo en cuenta que los datos recolectados de redes sociales como Twitter son una fuente importante de información mediante la cual es posible identificar tendencias y percepción de la población en general acerca de una compañía como Colfondos S.A; este análisis proporcionara una herramienta que permita a la empresa objeto de este estudio tener una visual frente al mercado, logrando así generar estrategias para una comunicación eficaz y asertiva.

Según Hernández et al. (2017):

Con los estudios descriptivos se busca especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis. Es decir, únicamente pretenden medir o recoger información de manera independiente o conjunta sobre los conceptos o las variables a las que se refieren, esto es, su objetivo no es indicar cómo se relacionan éstas. (p. 6)

Apoyando esta teoría Cognodata (2019) indica que:

Este tipo de metodología proporciona un enfoque por el que se confecciona un resumen de información que dan los datos de una muestra. Es decir, su meta es hacer síntesis de la información para arrojar precisión, sencillez y aclarar y ordenar los datos. (p.1)

#### **3.2 Enfoque metodológico**

El presente trabajo será diseñado bajo un enfoque mixto, dado que es el que mejor se adapta a las características y necesidades de la investigación. Esto debido a que se buscará identificar patrones en los diferentes contextos de comunicación e intercambio de opiniones tomando como unidad de análisis los usuarios de la red social Twitter, lo cual será clave para

entender lo que pueden llegar a sentir las diferentes audiencias a las que llegan las publicaciones realizadas y el entorno social, (Otero, 2018).

El enfoque mixto se caracteriza porque favorece una perspectiva amplia y profunda del fenómeno que se estudia, así mismo, esto le permite al investigador plantear el problema con claridad y formular objetivos apropiados para el estudio junto a una conceptualización teórica justa al fenómeno que se investiga. Para Hernández et al. (2017) “los métodos mixtos no nos proveen de soluciones perfectas, sin embargo, hasta hoy, son la mejor alternativa para indagar científicamente cualquier problema de investigación. Conjuntan información cuantitativa y cualitativa, y la convierten en conocimiento sustantivo y profundo”, (p. 16).

### **3.3 Población y Muestra**

La población con la que se realizó la investigación son los Tweets publicados entre marzo y abril de 2021 de los usuarios que han tenido experiencia con las AFP'S en Colombia, que realizan comentarios, opiniones e ideas acerca de los fondos de pensiones en el país y el servicio que ofrecen. Igualmente se tiene en cuenta la coyuntura política por la que atraviesa el país al momento de esta investigación (Elecciones presidenciales 2022-2026); cifras que determinan que la población a ser estudiada es finita. Teniendo en cuenta el tipo de población no se realiza muestreo ya que se tomará el 100% de los datos existentes en la red objeto de estudio.

### **3.4 Instrumentos**

El instrumento utilizado para el levantamiento y recolección de información es el complemento Tweet archiver de Google Drive, adicionalmente por medio del lenguaje computacional R se clasifica la información y se realiza un análisis de sentimientos.

Se tiene como base de información los Tweets recolectados entre marzo y abril de 2022 que contienen las palabras clave “Pensiones Colombia”, “Colfondos”, “Porvenir” y “Asfondos”;

dado que dentro de estas palabras podemos delimitar que la población corresponda al contexto colombiano, que se tenga un enfoque de la empresa objeto de estudio y del competidor más grande de los fondos de pensiones privadas y finalmente el ente que las regula.

Con esta información se obtienen las características de los usuarios entre las que se encuentran el origen del Tweet (Iphone, Android, Web App, etc.), fecha de creación del Usuario, Localización geográfica entre otras, posteriormente por medio del lenguaje computacional R se clasifica y se crea un modelo que realiza un análisis de sentimientos.

### **3.5 Análisis de Información**

Al realizar la recolección y transformación de la información los datos serán llevados a Power BI donde se creará un tablero de control que permita identificar el perfil de las personas que mencionan alguna de las palabras clave seleccionadas, adicionalmente se mostrará un análisis de sentimientos de estos Tweets clasificándolos en positivos, negativos o neutros. Finalmente, el tablero de control también permitirá evidenciar cantidad de seguidores de la red de Colfondos y una comparación con la competencia puntualmente Porvenir que es la AFP de mayor tamaño. De esta forma se logrará comunicar de manera descriptiva el comportamiento de los usuarios de la red y su opinión acerca de los fondos de pensiones en Colombia.

### **3.6 Condiciones Éticas**

La información recolectada en el presente proyecto será utilizada únicamente con fines académicos y no se utilizara de manera comercial, de igual manera la información plasmada en este documento será confidencial y tendrá como fin estudiar el comportamiento de los usuarios en la red social Twitter y su opinión acerca de las administradoras de fondos de pensiones en Colombia y el sistema pensional en general, se respetará y se dará manejo justo de los datos recolectados con los principios y valores inculcados por la universidad.



### **3.7 Desarrollo**

Para el desarrollo de este proyecto se realiza un Dash Board o tablero de control donde se visualizan indicadores de sentimiento y las métricas tanto de las publicaciones como las interacciones con respecto a los fondos de pensiones Privados en Colombia, esta información permitirá a Colfondos tomar decisiones estratégicas acerca de sus contenidos.

La construcción tablero de control se basa en 2 procesos. Un proceso de extracción de tweets para posterior limpieza y almacenamiento en una base de datos, dicha información pasa por un proceso de clasificación según el sentimiento (Análisis de sentimientos), proceso realizado en lenguaje R. El segundo proceso, es la construcción de la visualización de dicha información en Power BI donde se incluirán datos como la tasa de engagement, seguidores entre otras; que será entregada de manera gráfica y comprensible para la toma de decisiones.

### **3.9 Análisis de sentimientos**

Por medio del análisis de sentimientos se puede entender la percepción pública de las personas acerca de una empresa o producto, reaccionar a cambios en el sentimiento, determinar temas de interés o encontrar actores generadores de opinión. Para este proyecto se toma la información de Twitter por medio de la API de esta red social.

El análisis de sentimientos se centra en 3 pasos fundamentales. (Ver figura 2)

**Figura 2**

*Pasos para el análisis de sentimientos*

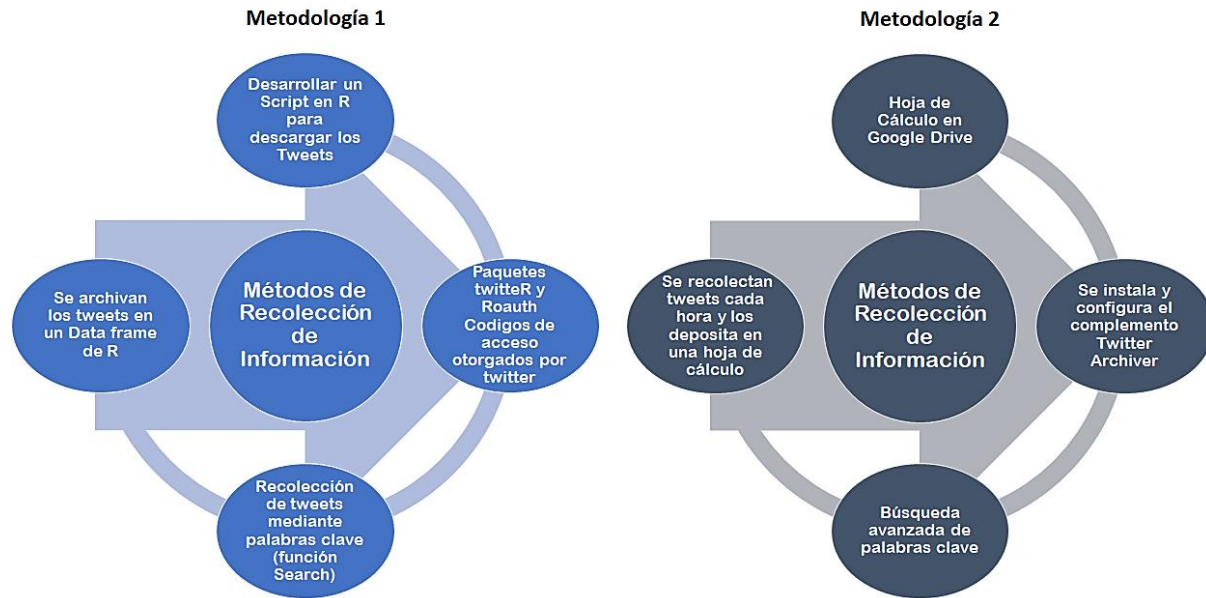


Nota. Nos muestra los 3 pasos fundamentales de análisis de sentimientos. Elaboración propia a partir de BD Guidance (Video)

- **Métodos de Recolección de datos**

Existen dos metodologías que utiliza la API de Twitter y que se explican a continuación:

- a) La primera consiste en desarrollar un Scrip en R que permita descargar en tiempo real los tweets donde se hace uso de los paquetes TwitteR y ROauth utilizados para activar las credenciales de la API, mediante la función Search se descargan los tweets que se necesitan para ser archivados en un data frame de R.
- b) La segunda metodología consta de un complemento de Google Drive llamado Tweet Archiver que permite configurar y exportar una búsqueda específica de tweets en una hoja de cálculo. Esta herramienta cuenta con una versión gratuita y otra paga.

**Figura 3***Metodología de recolección de información*

Nota. Elaboración propia a partir de BD Guidance (Video)

Para obtener los tweets en el presente proyecto la metodología utilizada para la minería de datos es Tweet Archiver, ya que esta no es necesaria la instalación de software adicional, su conectividad es integrada con Twitter y cuenta con una interfaz amigable al usuario.

En esta etapa se tuvieron en cuenta los tweets y ReTweets en español que contenían palabras como “Colfondos”, “Pensiones Colombia”, “Porvenir” y “Asofondos” utilizando el complemento ya mencionado de Google Drive, en este proceso se lograron recolectar aproximadamente 3,900 tweets; es importante mencionar que la herramienta también nos permite establecer criterios de búsqueda como Frases exactas, que el tweet contenga o no alguna palabra en específico, un hashtag, el idioma de los tweets entre otras como se puede ver en la Figura 4.

**Figura 4**

*Proceso de recolección de tweets en Google Drive*

**Update Twitter Rule** ×

All of these words	<input type="text" value="Colfondos"/>	This exact phrase	<input type="text"/>
Any of these words	<input type="text"/>	None of these words	<input type="text"/>
These #hashtags	<input type="text"/>	Written in	<input type="text" value="Spanish (español)"/>
Near This Place	<input type="text" value="Premium feature"/>	Advanced Rules	<input type="text" value="-filter:retweets -filter:repl"/>
<b>People</b>			
To these accounts	<input type="text"/>	Mentioning accounts	<input type="text"/>
From these accounts	<input type="text"/>		

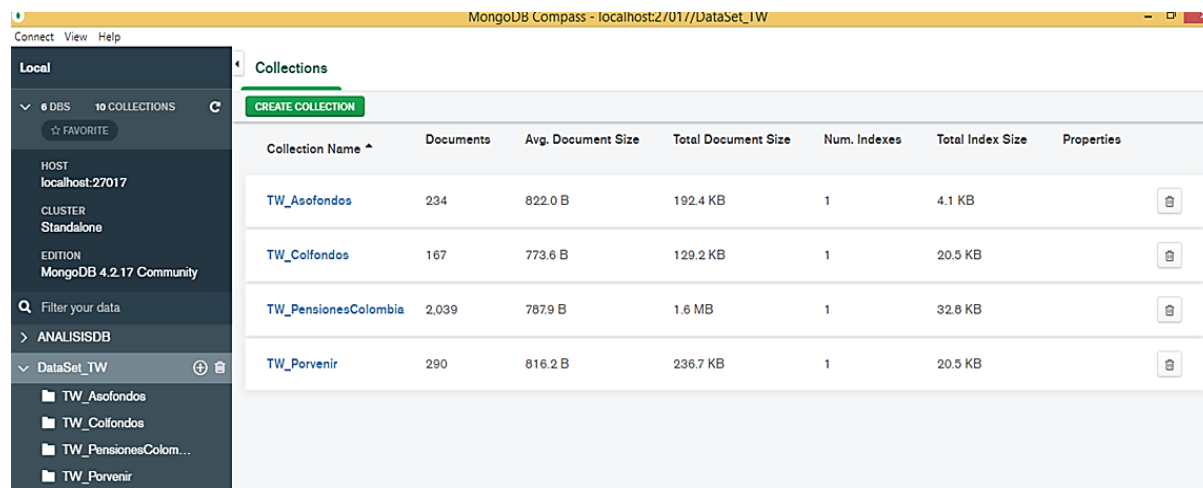
Twitter Search Query: Colfondos lang:es -filter:retweets -filter:replies

Nota. Google Drive

Posterior a la recolección de información, se formó un Dataset utilizando el motor de base de datos no relacionales MongoDB, se crearon 4 colecciones donde se alojan los tweets de cada uno de los filtros aplicados, dichas colecciones son el insumo principal para las siguientes etapas, estos datos se encuentran sin tratar, por ello aun no es posible aplicar un modelo para el análisis de sentimientos.

## Figura 5

### Almacenamiento de datos en Mongo DB



The screenshot shows the MongoDB Compass interface. The main window displays a table of collections under the 'DataSet\_TW' database. The table has columns for Collection Name, Documents, Avg. Document Size, Total Document Size, Num. Indexes, Total Index Size, and Properties. The collections listed are TW\_Asofondos, TW\_Colfondos, TW\_PensionesColombia, and TW\_Porvenir.

Collection Name	Documents	Avg. Document Size	Total Document Size	Num. Indexes	Total Index Size	Properties
TW_Asofondos	234	822.0 B	192.4 KB	1	4.1 KB	
TW_Colfondos	167	773.6 B	129.2 KB	1	20.5 KB	
TW_PensionesColombia	2,039	787.9 B	1.6 MB	1	32.8 KB	
TW_Porvenir	290	816.2 B	236.7 KB	1	20.5 KB	

Nota. Mongo DB

- **Pre-Procesamiento de Datos y Análisis de Sentimientos**

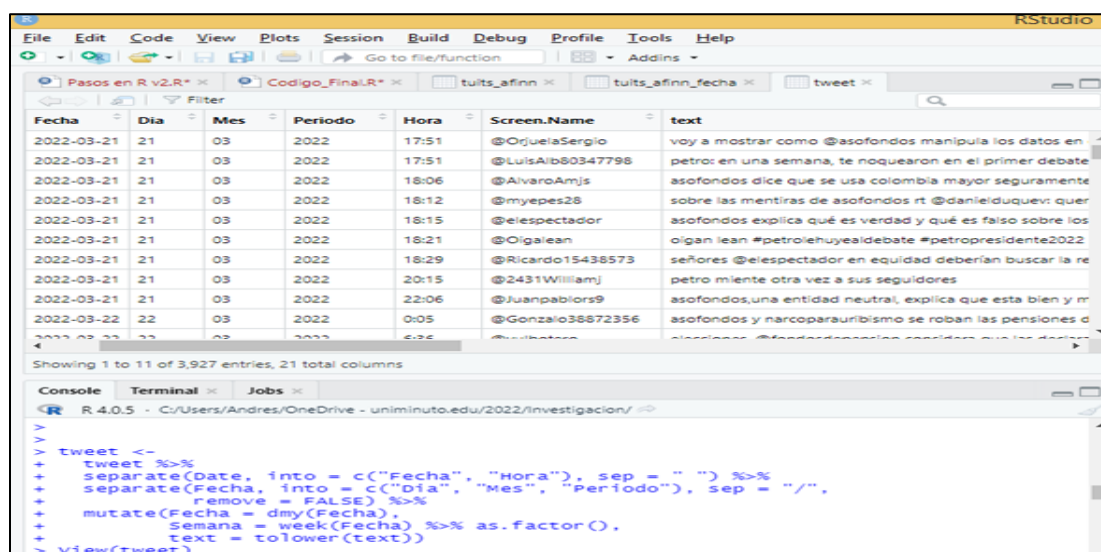
Para realizar el análisis de los tweets obtenidos se necesita realizar un proceso de limpieza, preprocesamiento y transformación al total de los tweets, este proceso se realiza en lenguaje R, para el cual es necesario instalar los siguientes paquetes:

- **Textdata:** Este paquete permite descargar, analizar y almacenar los datos en el disco y cargarlos cuando sea necesario.
- **Tidytext:** Contiene funciones y conjuntos de datos de apoyo para la conversión de texto hacia y desde formatos ordenados, permite cambiar entre las herramientas ordenadas y los paquetes de minería de texto existentes.
- **Tidyr:** El objetivo es crear datos ordenados donde cada columna es variable, cada fila es una observación y cada celda es un valor único.
- **Lubridate:** Sirve para convertir una columna a tipo de dato fecha.

Después de instaladas las librerías el paso inicial es convertir el campo Date de la base consolidada con la función `separate()` de la librería `tidyr`, esta columna queda separada en una fecha y hora del día, para después separar la fecha en día, mes y año, posteriormente la función `ymd()` de la librería `lubridate` es utilizada para convertir la nueva columna Fecha a tipo de dato fecha, posteriormente con la función `tolower` se convierten todas las palabras de los tweets en minúscula y se eliminan signos de puntuación.

## Figura 6

*Preparando Datos (campo fecha)*



The screenshot displays the RStudio interface. The top pane shows a data table with the following columns: Fecha, Dia, Mes, Periodo, Hora, Screen.Name, and text. The bottom pane shows the R console with the following code:

```

> tweet <-
+ tweet %>%
+ separate(Date, into = c("Fecha", "Hora"), sep = " ") %>%
+ separate(Fecha, into = c("Dia", "Mes", "Periodo"), sep = "/",
+ remove = FALSE) %>%
+ mutate(Fecha = dmy(Fecha),
+ Semana = week(Fecha) %>% as.factor(),
+ text = tolower(text))
> view(tweet)

```

Nota. Proceso en R

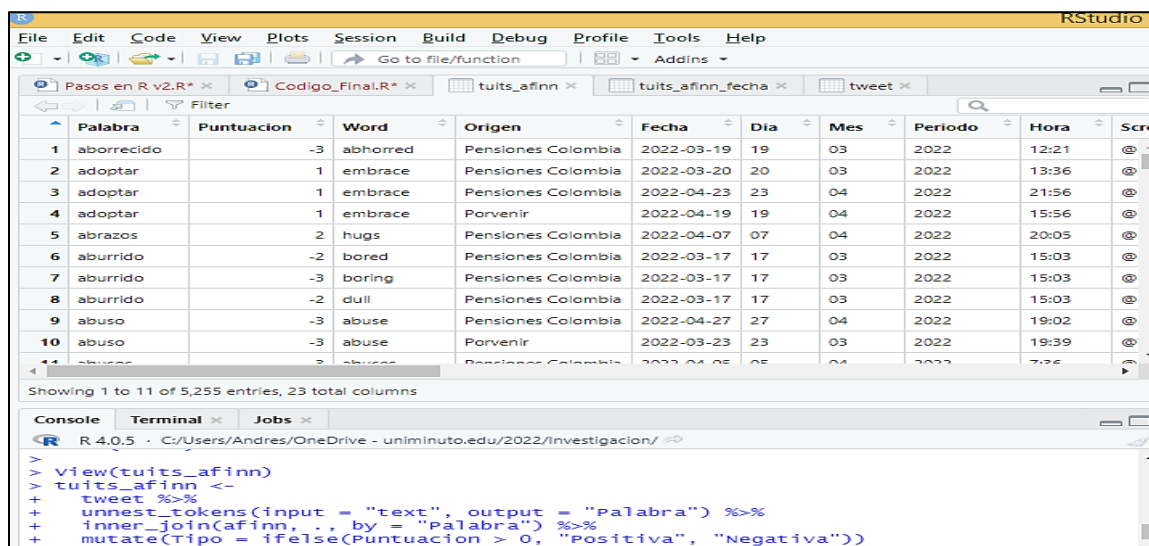
Con el fin de clasificar los tweets según el sentimiento es necesario separar cada uno en palabras y asignarle una puntuación de sentimiento a cada una, para esto se utiliza el Léxico Afinn; el cual corresponde a un conjunto de palabras puntuadas dependiendo que tan positiva o negativamente son percibidas; para determinar si son positivas se asigna una puntuación de 1 a 5 y para las negativas su puntuación es entre -5 y -1. La versión utilizada corresponde a una traducción automática, de inglés a español la cual fue descargada directamente de la siguiente

dirección [https://raw.githubusercontent.com/jboscomendoza/rpubs/master/sentimientos\\_afinn/exico\\_afinn.en.es.csv](https://raw.githubusercontent.com/jboscomendoza/rpubs/master/sentimientos_afinn/exico_afinn.en.es.csv); la cual contiene 2,476 palabras, 3 columnas, una con las palabras en español, su puntuación y una tercera con las palabras en inglés.

Con este proceso se obtiene un data frame que contiene los tweets con palabras presentes en el léxico Afinn, una columna Tipo y la columna Origen que será utilizada para los análisis posteriores.

## Figura 7

### *Convertir tweets en palabras*



	Palabra	Puntuacion	Word	Origen	Fecha	Dia	Mes	Periodo	Hora	Screenname
1	abhorrecido	-3	abhorred	Pensiones Colombia	2022-03-19	19	03	2022	12:21	@
2	adoptar	1	embrace	Pensiones Colombia	2022-03-20	20	03	2022	13:36	@
3	adoptar	1	embrace	Pensiones Colombia	2022-04-23	23	04	2022	21:56	@
4	adoptar	1	embrace	Porvenir	2022-04-19	19	04	2022	15:56	@
5	abrazos	2	hugs	Pensiones Colombia	2022-04-07	07	04	2022	20:05	@
6	aburrido	-2	bored	Pensiones Colombia	2022-03-17	17	03	2022	15:03	@
7	aburrido	-3	boring	Pensiones Colombia	2022-03-17	17	03	2022	15:03	@
8	aburrido	-2	dull	Pensiones Colombia	2022-03-17	17	03	2022	15:03	@
9	abuso	-3	abuse	Pensiones Colombia	2022-04-27	27	04	2022	19:02	@
10	abuso	-3	abuse	Porvenir	2022-03-23	23	03	2022	19:39	@
11	abuso	-3	abuse	Pensiones Colombia	2022-04-05	05	04	2022	7:26	@

```

R 4.0.5 · C:/Users/Andres/OneDrive - uniminuto.edu/2022/Investigacion/
>
> view(tuits_afinn)
> tuits_afinn <-
+   tweet %>%
+   unnest_tokens(input = "text", output = "Palabra") %>%
+   inner_join(afinn, ., by = "Palabra") %>%
+   mutate(Tipo = ifelse(Puntuacion > 0, "Positiva", "Negativa"))

```

Nota. Proceso en R

El paso final del análisis de sentimientos es asignar una puntuación a cada uno de los tweets, utilizando un promedio del puntaje de las palabras que compone el tweet; también se les asigna a los tweets q no tienen puntuación positiva o negativa un valor de cero (0) que indica neutralidad.

### 3.10 Exploración de datos

Después de obtener el modelo de sentimientos se pasa a realizar la exploración de los datos encontrados en el proceso:

- En la siguiente figura se muestra la cantidad de palabras utilizadas en cada categoría; para Colfondos (237), Porvenir (525), Asofondos (736) y Pensiones Colombia (3,757).

Después de agrupar las palabras utilizadas estas se reducen a 249 palabras únicas para Asofondos, 117 para Colfondos, 547 Pensiones Colombia y 175 para Porvenir.

#### Figura 8

##### *Palabras por Categoría*

```
> view(tuits_afinn)
> write.csv(tuits_afinn, "base_tuit_afinn.csv")
> view(tuits_afinn)
> view(afinn)
> tuits_afinn %>% count(Origen)
# A tibble: 4 x 2
  Origen          n
  <chr>         <int>
1 Asofondos       736
2 Colfondos       237
3 Pensiones Colombia 3757
4 Porvenir         525
>
> tuits_afinn %>% group_by(Origen) %>% distinct(Palabra) %>% count()
# A tibble: 4 x 2
# Groups:   Origen [4]
  Origen          n
  <chr>         <int>
1 Asofondos       249
2 Colfondos       117
3 Pensiones Colombia 547
4 Porvenir         175
```

Nota. R Studio

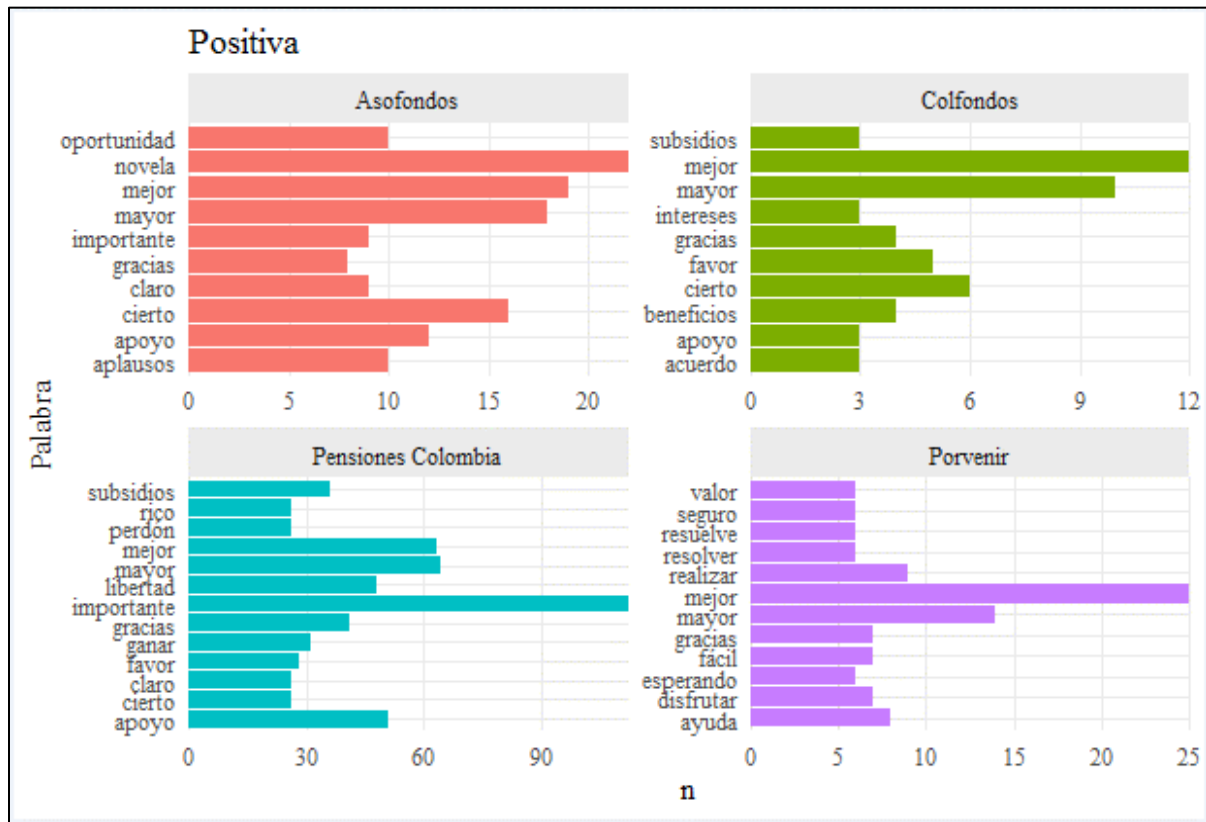
- La nube de palabras nos permite identificar e interpretar de manera ágil y sencilla las palabras con mayor relevancia en los tweets objeto de estudio.





**Figura 10**

*Palabras positivas por Categoría.*

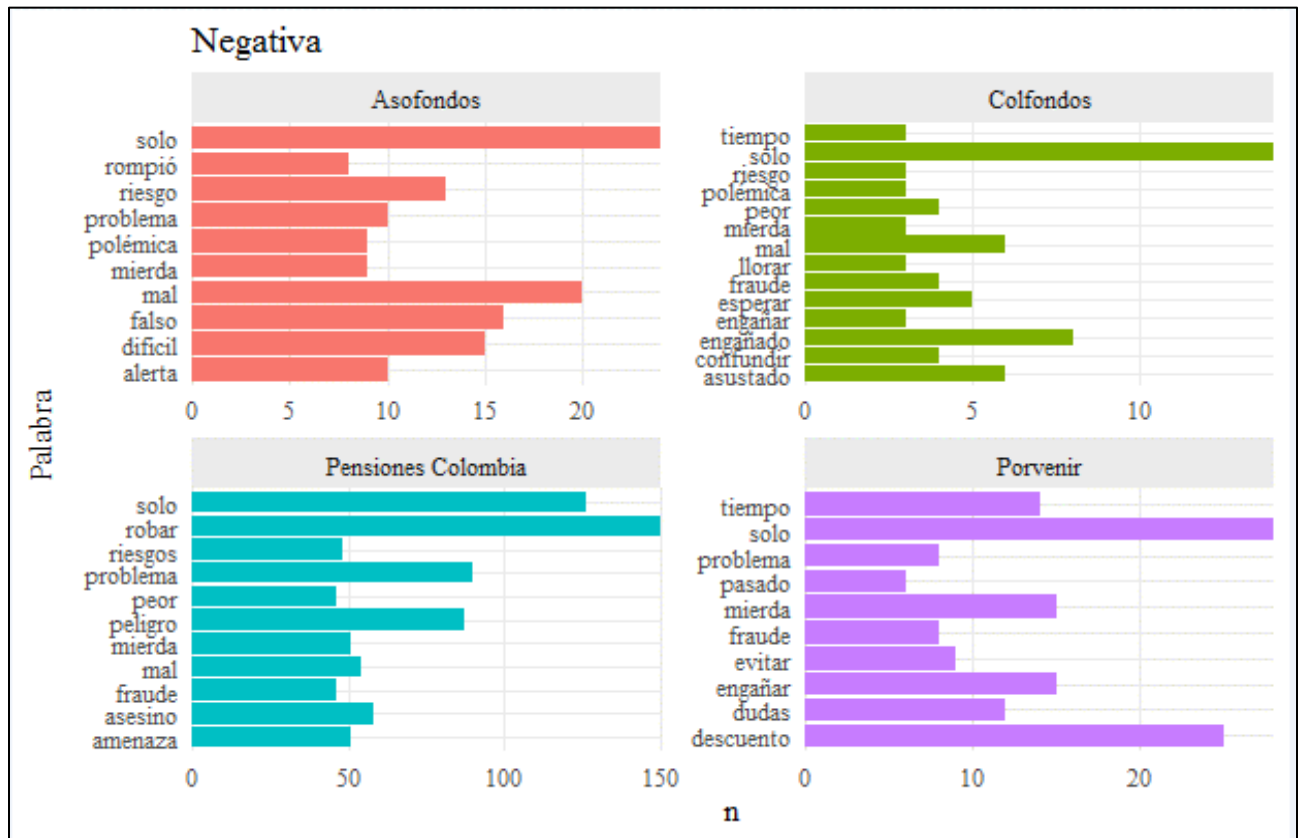


Nota. Elaboración propia en R Studio

- La figura 10 Corresponde al top de las palabras negativas más utilizadas; Asofondos, Colfondos y Porvenir se utiliza la palabra Solo, para Pensiones Colombia la palabra es Robar.

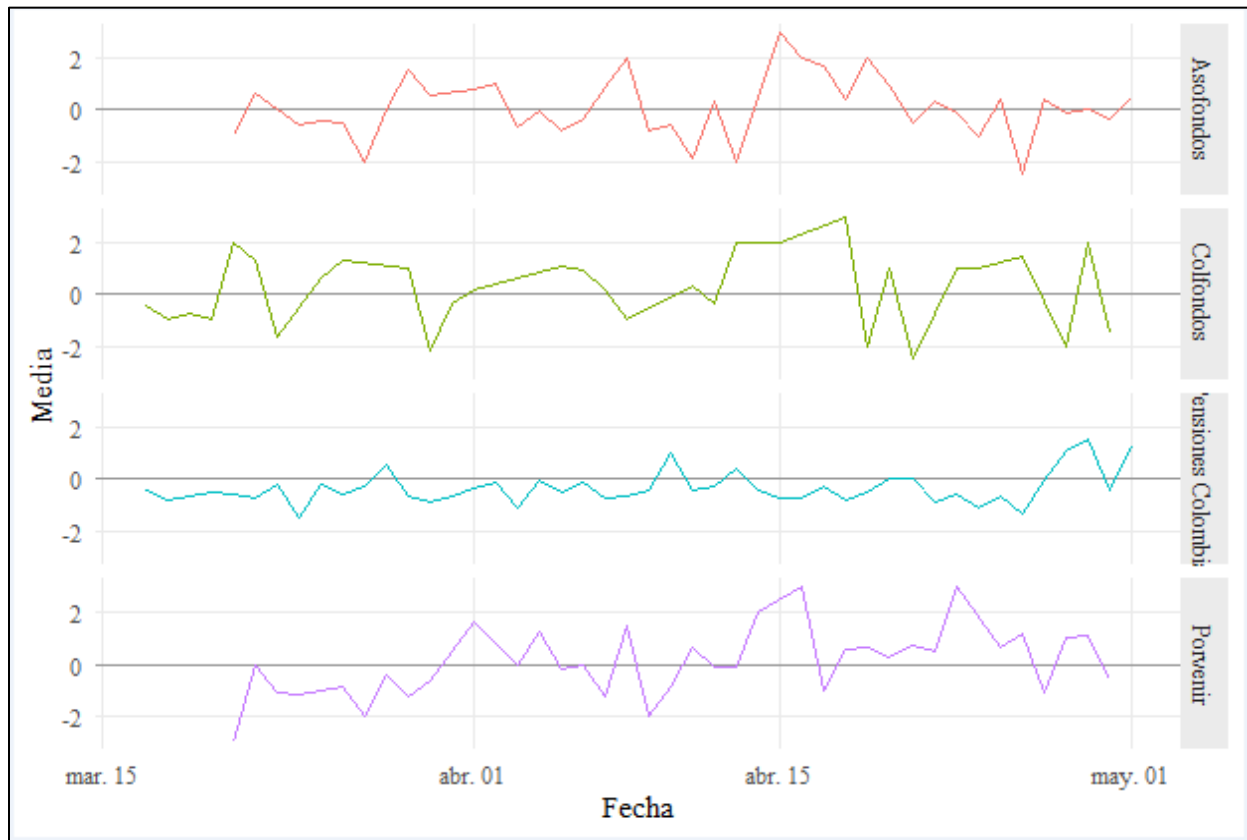
**Figura 11**

*Palabras Negativas por Categoría*



Nota. Elaboración propia en R Studio

- Es importante observar la tendencia de sentimientos de las palabras utilizadas en el tiempo, por categoría.

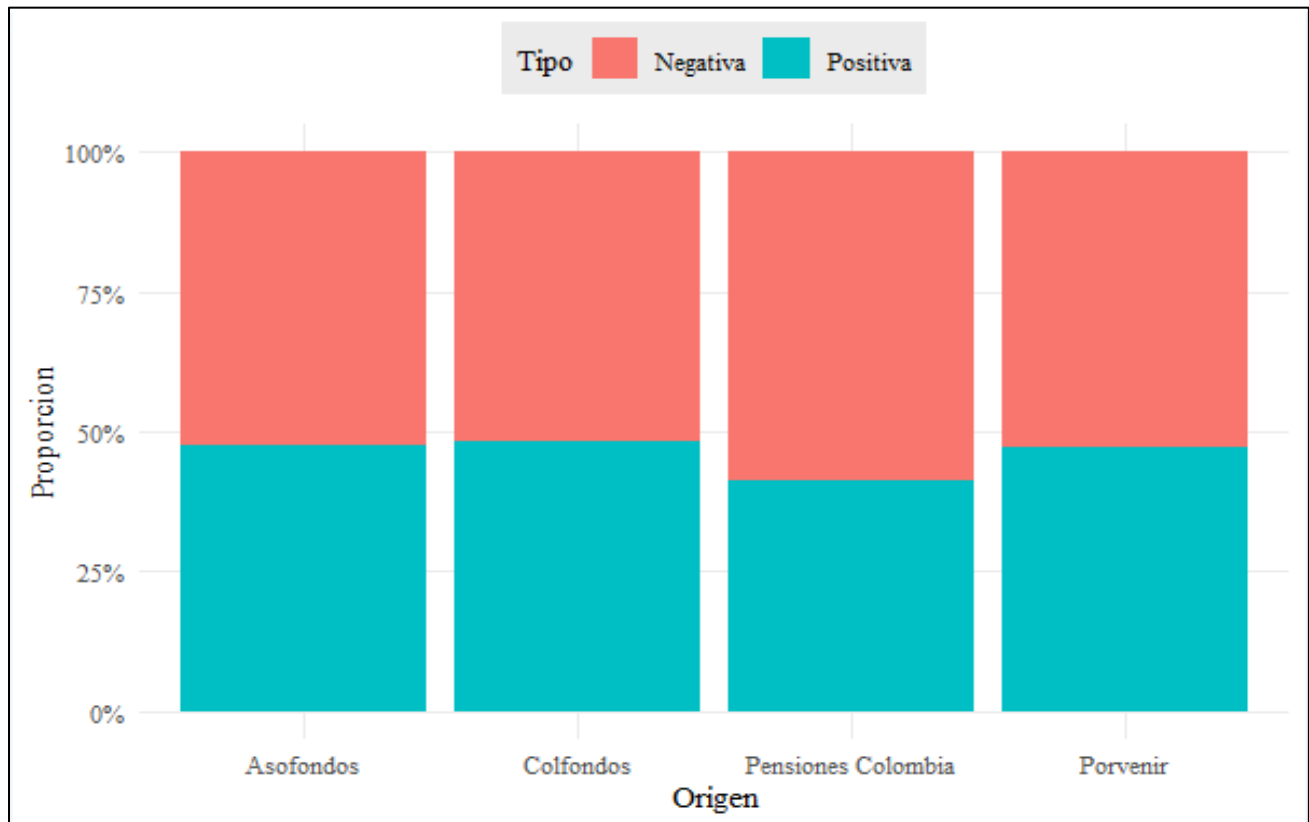
**Figura 12***Tendencia por categoría*

Nota. Elaboración propia en R Studio

- En la siguiente figura se presenta la proporción de tweets positivos y negativos por categoría, encontrando gran neutralidad en 3 de las 4 categorías analizadas, para el caso de Pensiones Colombia la proporción de tweets negativos es mayor.

**Figura 13**

*Proporción negativa y positiva de los tweets.*



Nota. Elaboración propia en R Studio

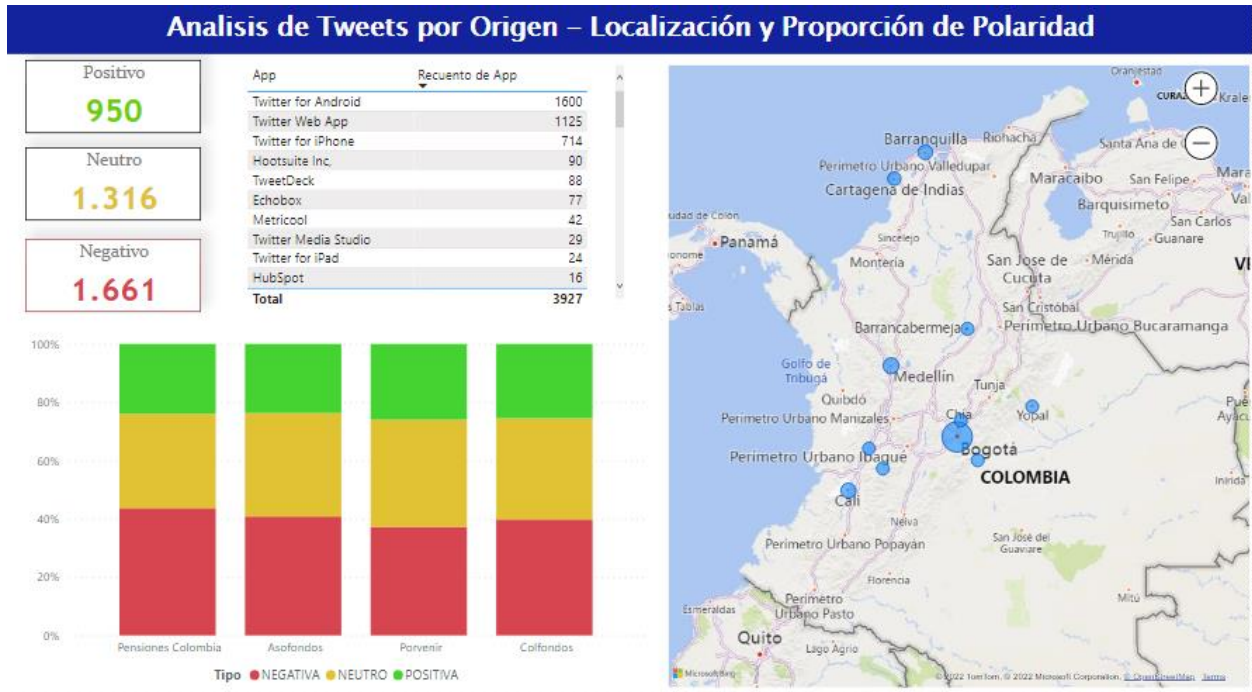
**Tablero de Control**

Como ya fue mencionado el tablero de control es diseñado en el sistema de información Power BI, herramienta de Microsoft, dicha herramienta es una solución de análisis que permite visualizar los datos de una manera dinámica y ordenada. Se construyeron gráficas que plasmaron la información recolectada en los tweets para facilitar la toma de decisiones (Gráficos de Barras, Nube de palabras, Tablas, Gráficos de torta, medidores, etc.). El tablero incluye Análisis por tweets y análisis por palabras totales en los tweets recolectados.

En la figura 14 se presenta una vista de los Tweets por Origen (Andriod, Web, Iphone, entre otros), localización y la correspondiente proporción de polaridad.

**Figura 14**

*Visual Tablero de control por Tweet.*



Nota. Elaboración propia en Power BI

Igualmente se podrá evidenciar un análisis de sentimientos por palabras, así como las palabras más utilizadas en los Tweets tanto en el sector como un comparativo con la competencia, en la siguiente figura se muestra un ejemplo.

**Figura 15**

*Visual Tablero de control Análisis de sentimientos por Palabras*



Nota. Elaboración propia en Power BI

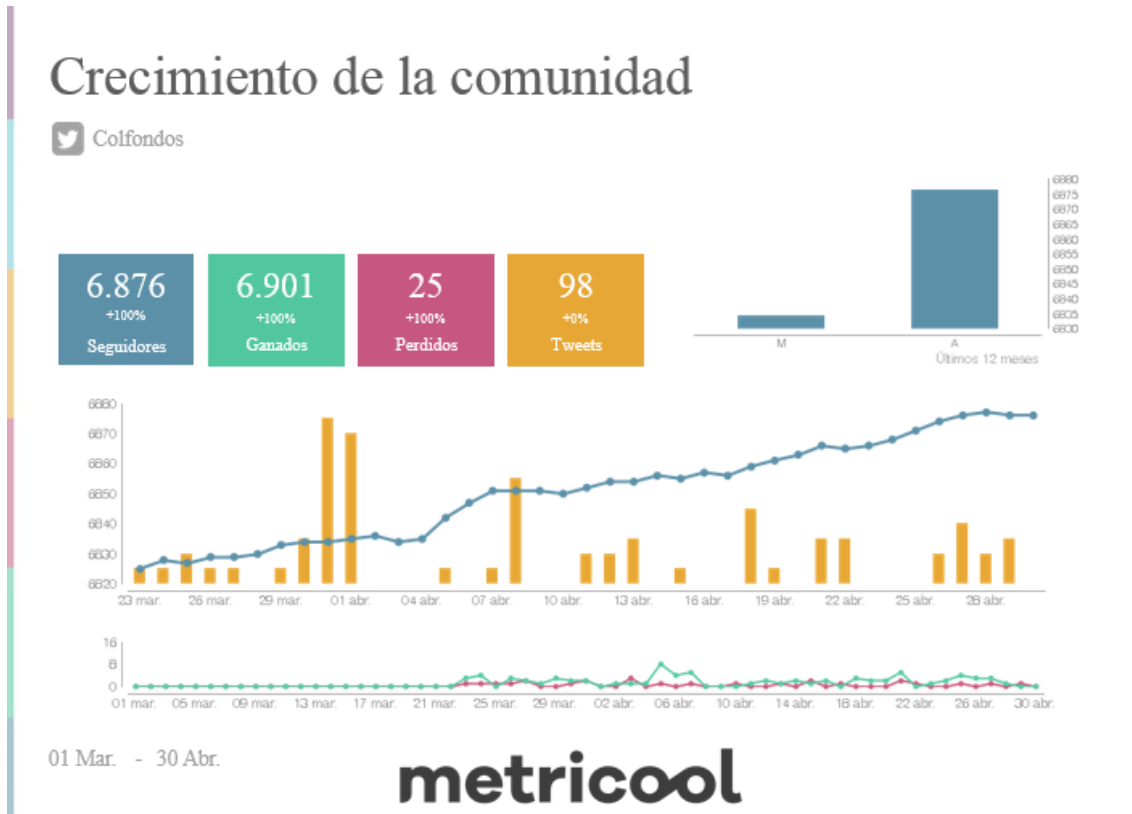
Finalmente podemos ver un ejemplo de las gráficas que muestran un panorama general de la nube de palabras y su relación con diferentes variables encontradas en la data recolectada.





**Figura 17**

*Métricas de Twitter*



Nota. Muestra un seguimiento a las redes sociales. App Metricool

## Conclusiones

Con el desarrollo de la presente investigación se encontró que mediante las técnicas de PNL (procesamiento de lenguaje Natural) es posible realizar un análisis de los sentimientos presentes en una cadena de texto, para este caso aplicado a los tweets que dentro del texto del mismo contengan “Colfondos”, “Porvenir”, “Asofondos” o “Pensiones Colombia”, permitiendo identificar la percepción de los usuarios de la red convirtiendo estos en un insumo importante para la creación de planes de acción a nivel estratégico y comercial.

Entre las herramientas más importantes de minería de texto se encuentra R, el cual a través de diferentes librerías, permite realizar un preprocesamiento de datos (limpieza y normalización) y aplicar técnicas de análisis de sentimientos

Durante el análisis se identificó la importancia que tiene el contexto político-social al momento de su desarrollo, de manera específica Elecciones presidenciales 2022. Teniendo en cuenta que el lenguaje natural es complejo, se hace difícil enseñar a una máquina a entender como el contexto puede llegar a afectar la intención real o significado de alguna frase. Enseñar sarcasmo, expresiones propias de una cultura o jergas a una máquina es uno de los mayores desafíos en este tipo de análisis.

Finalmente se logra identificar una mayor proporción de negatividad en los tweets recolectados, tanto para Colfondos como para el sistema pensional en Colombia, es importante para la compañía generar estrategias de seguimiento y control a las redes sociales, teniendo en cuenta la importancia que tiene la opinión de los usuarios de este tipo de redes. Esta investigación proporciona datos del sector pensional en Colombia, el análisis de tweets en el sector podrá ser usado como estrategias de marketing en diferentes compañías

## Referencias

Alcaldía de Bogotá. (s.f). Total: 23 documentos encontrados para datos personales Protección.

<https://www.alcaldiabogota.gov.co/sisjur/listados/tematica2.jsp?subtema=27077>

Alcaldía de Bogotá. (s.f). Ley 1581 de 2012. Desarrolla el derecho constitucional

Barrio Carrasco, J. (2017). *La influencia de los medios sociales digitales en el consumo. La función prescriptiva de los medios sociales en la decisión de compra de bebidas refrescantes en España*. (tesis). Universidad Complutense de Madrid. Facultad de Ciencias de la Información. <https://eprints.ucm.es/id/eprint/42339/1/T38702.pdf>

Colfondos de Grupo Habitat. (2020). Informe de gestión.

<https://www.colfondos.com.co/dxp/documents/20143/36286755/INFORME-DE-GESTION-VF-2021.pdf/>

Cognodata. (2019). ¿Qué es el análisis descriptivo? <https://www.cognodata.com/que-es-analisis-descriptivo/>

Constitución Política de Colombia. (1991). Artículo 15.

<https://www.mincit.gov.co/ministerio/normograma-sig/procesos-estrategicos/gestion-de-informacion-y-comunicacion/constitucion-politica/derechos/articulo-15.aspx>

Galeano, P. y Peña, D. (2019). Las Nuevas Oportunidades Del Big Data Para Las Instituciones Financieras. Papeles de economía española, N.º 162. [https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS\\_PEE/162art07.pdf](https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PEE/162art07.pdf)

García Alsina, M. (2017). Big data: gestión y explotación de grandes volúmenes de datos. Barcelona, Editorial UOC. <https://elibro.net/es/ereader/uniminuto/116303?page=9>.

Hall. (2022). Digital report 2022: el informe sobre las tendencias digitales, redes sociales y mobile.

Guzmán Duque, A. P., del Moral Pérez, M. E., González Ladrón de Guevara, F., & Gil Gómez,

H. (2013). Impacto de twitter en la comunicación y promoción institucional de las universidades. *Pixel-Bit. Revista De Medios Y Educación*, (43), 139-153.

<https://doi.org/10.12795/pixelbit.2013.i43.10>

Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2017). *Metodología de la investigación*. 6ed. Mc Graw Hill. <https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf>

Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2017). *Capítulo 12. Ampliación y Fundamentación de los métodos Mixtos. Metodología de la investigación*  
<https://es.slideshare.net/conyas16/sampieri-metodos-mixtos>

Ibáñez San Millán, M. D. (2014). *Redes sociales para PYMES: Introducción al Community Management*. Madrid, Spain: Ministerio de Educación y Formación Profesional de España. <https://elibro.net/es/ereader/uniminuto/49378?page=35>

Jiménez P, C.C., Piraquive G, R.R., Malaver R, N. y Rivera R, H. A. (2011). *Análisis estratégico sector fondos de pensiones en Colombia*. Universidad del Rosario.  
<https://repository.urosario.edu.co/bitstream/handle/10336/3317/Fasc%C3%ADculo109.pdf?sequence=1&isAllowed=y>

Kemp. S. (2022). *Digital 2022: global overview Report*. <https://datareportal.com/reports/digital-2022-global-overview-report>

Ley 1266 de 2008. por la cual se dictan las disposiciones generales del hábeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones.

<https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=34488#:~:text=por%20la%20cual%20se%20dictan,y%20se%20dictan%20otras%20disposiciones.>

López García, D. (2012-2013). *Análisis de las posibilidades de uso de Big Data en las organizaciones*. (Maestría) Universidad de Cantabria.

<https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADaS.pdf?sequence=1>

Oracle. (2022). ¿Qué es el big data?. <https://www.oracle.com/co/big-data/what-is-big-data/>

Otero Ortega, A. (2018). *Enfoques de Investigación*.

[https://www.researchgate.net/profile/Alfredo-Otero-Ortega/publication/326905435\\_\\_ENFOQUES\\_DE\\_INVESTIGACION/links/5b6b7f9992851ca650526dfd/ENFOQUES-DE-INVESTIGACION.pdf](https://www.researchgate.net/profile/Alfredo-Otero-Ortega/publication/326905435__ENFOQUES_DE_INVESTIGACION/links/5b6b7f9992851ca650526dfd/ENFOQUES-DE-INVESTIGACION.pdf)

Pérez Marqués, M. (2015). *Big Data Técnicas, herramientas y aplicaciones*.

[http://rclibros.es/wp-content/uploads/2015/04/capitulo\\_9788494305559.pdf](http://rclibros.es/wp-content/uploads/2015/04/capitulo_9788494305559.pdf).

Policía Nacional de Colombia. (2022). Ley 1273 de 2009. Normatividad sobre delitos informáticos. <https://www.policia.gov.co/denuncia-virtual/normatividad-delitos-informaticos>

Secretaría de transparencia. (2021). *Ley de Transparencia y Derecho de Acceso a la Información Pública*. Ley 1712 de 2014.

<http://www.secretariatransparencia.gov.co/politica-publica/ley-de-transparencia>

Romero Moreno, F. Y., Sánchez Martelo, C. A., Alfonso Corredor, B. Y., Sánchez Cifuentes, J. F. y Ospina López, J. P. (2020). Técnicas para la Clasificación de Sentimientos en Redes Sociales como Apoyo en el Marketing Digital. *Revista Ibérica de Sistemas e Tecnologías de Informação RISTI*, N.º E35, 167.

<https://www.proquest.com/openview/10505af13f0740e170a77dec671c59e9/1?pq-origsite=gscholar&cbl=1006393>

Rodríguez Tapia, S., y Camacho Cañamón., J. (2018). Los métodos de aprendizaje automático supervisado en la clasificación textual según el grado de especialización *Tonos Digital; Murcia N.º 35, 1-28.*

<https://www.proquest.com/docview/2091668229/fulltext/6B484245CDA54DF8PQ/1?accountid=48797>

Secretaria Distrital de Planeación. (2022). Sociedad administradora de fondos de pensiones y de cesantía – AFP. <https://www.sdp.gov.co/transparencia/informacion-interes/glosario/sociedad-administradora-de-fondos-de-pensiones-y-de-cesantia-afp>

Selva Ruiz, D. y Caro Castaño, L. (2016). Uso de datos en la creatividad publicitaria: el caso de Art, Copy & Code de Google. *El profesional de la información, V. 25, N.º. 4.*

<http://profesionaldelainformacion.com/contenidos/2016/jul/14.pdf>

Sicilia, M., Palazón, M., López, I., y López, M. (2021). Marketing en redes sociales.

<https://books.google.es/books?hl=es&lr=&id=TZoZEAAAQBAJ&oi=fnd&pg=PA7&dq>

Superintendencia de Industria y Comercio. (2021). Sobre el Habeas Data Financiero.

<https://www.sic.gov.co/sobre-el-habeas-data-financiero>

Sobrino Sande, J (2018). Análisis de Sentimientos en Twitter.

<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81435/6/jsobrinosTFM0618memoria.pdf>