UNIVERSIDAD LAICA ELOY ALFARO DE MANABÍ

FACULTAD DE CIENCIAS INFORMÁTICAS

TRABAJO DE TITULACIÓN MODALIDAD PROYECTO INTEGRADOR, PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

INGENIERO EN SISTEMAS

TEMA:

"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"

AUTORES:

PINARGOTE MENDOZA WENDY JAHAYRA

VÉLEZ FLORES BRYAN FERNANDO

DIRECTOR:

ING. FABRICIO RIVADENEIRA

MANTA – MANABI – ECUADOR

2018

CERTIFICACIÓN

En calidad de docente tutor(a) de la Facultad de Ciencias Informáticas de la Universidad Laica "Eloy Alfaro" de Manabí, certifico:

Haber dirigido y revisado el trabajo de titulación, cumpliendo el total de 71 horas, bajo la modalidad de Proyecto Integrador, cuyo tema del proyecto es "**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**", el mismo que ha sido desarrollado de acuerdo a los lineamientos internos de la modalidad en mención y en apego al cumplimiento de los requisitos exigidos por el Reglamento de Régimen Académico, por tal motivo CERTIFICO, que el mencionado proyecto reúne los méritos académicos, científicos y formales, suficientes para ser sometido a la evaluación del tribunal de titulación que designe la autoridad competente.

La autoría del tema desarrollado, corresponde a los señores **PINARGOTE MENDOZA WENDY JAHAYRA** y **VÉLEZ FLORES BRYAN FERNANDO**, estudiantes de la carrera de Ingeniería en Sistemas, período académico 2018-2019, quienes se encuentran aptos para la sustentación de su trabajo de titulación.

Particular que certifico para los fines consiguientes, salvo disposición de Ley en contrario.

Manta, 05 de septiembre del 2018.

Los certifico,

Ing. Fabricio Rivadeneira

Docente Tutor(a)

UNIVERSIDAD LAICA "ELOY ALFARO" DE MANABÍ
Creada el 13 de noviembre de 1985 mediante Decreto Ley No.10, publicado en el Registro Oficial No. 313
FACULTAD DE CIENCIAS INFORMÁTICAS
Creada, Resolución H. Consejo Universitario del 11 de Julio del 2001

TRABAJO DE TITULACIÓN MODALIDAD PROYECTO INTEGRADOR, PREVIO A LA OBTENCIÓN DEL TÍTULO DE: INGENIERO/A EN SISTEMAS

"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"

Tribunal examinador que declara APROBADO el Grado de INGENIERO/A EN SISTEMAS, del, la, las o los señor/ita:

WENDY JAHAYRA PINARGOTE MENDOZA
BRYAN FERNANDO VÉLEZ FLORES

Lic. Dolores Muñoz Verduga PhD. _____

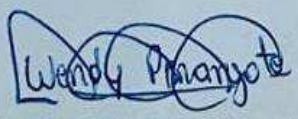
Ing. Oscar González López Mg. _____

Ing. Robert Moreira Centeno Mg. _____

Manta, 11 / 12 de septiembre de 2018

DECLARACIÓN DE AUTORÍA

La Srta. Wendy Jahayra Pinargote Mendoza con cédula de ciudadanía N° 1315606630 y el Sr. Bryan Fernando Vélez Flores con cédula de ciudadanía N° 1311434961, declaran ser autores de este trabajo de titulación "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" que ha sido desarrollado respetando los derechos intelectuales de terceros considerándose en las citas bibliográficas.



Wendy Pinargote Mendoza

Bryan Vélez Flores

DEDICATORIA

A Dios por permitirme culminar esta nueva etapa en mi vida., por darme a mis padres maravillosos, quienes me inculcaron valores, me enseñaron a luchar para conseguir mis metas y por su apoyo incondicional en todo momento. A la señora Jesenia Borja por brindarme sus consejos y apoyo constante, y sobre todo dedicado con mucho amor y cariño a mi hija porque es mi vida entera, mi mayor motivación e inspiración para superarme día tras día y así poder brindarle un futuro mejor.

Wendy Pinargote

Dedico este proyecto a Dios con mucho amor y gratitud, por darme a mis padres como pilar fundamental y apoyo en mi formación académica, dándome todo lo que soy como persona, mis valores, principios, perseverancia, empeño y deseo de superación, les estaré eternamente agradecido.

Bryan Vélez

AGRADECIMIENTOS

En primera instancia agradecemos a Dios que hace que cada circunstancia y vivencia tenga su razón de ser.

Agradecemos también a la institución, docentes y padres, personas de gran sabiduría, por sus esfuerzos en guiarnos en la formación como profesionales. A nuestro tutor el Ing. Fabricio Rivadeneira por encaminarnos y guiarnos en el desarrollo y culminación de nuestro proyecto, expresamos nuestro más sincero agradecimiento.

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE GRÁFICOS

RESUMEN

El presente proyecto aplica técnicas de minería de texto a la extracción de comentarios en la red social Facebook de los perfiles oficiales de las empresas Movistar, Claro y CNT a nivel nacional, desde enero hasta mayo del 2018, sobre el servicio de telefonía móvil, con la finalidad de analizar el criterio positivo, negativo o neutro que expresan los internautas mediante un análisis de sentimientos, implementando la metodología CRISP-DM para la extracción y manipulación de la información y aplicando gráficos estadísticos. Donde se estableció que el gráfico de pirámides es el más adecuado en base a las comparaciones realizadas de los términos comunes más frecuentes entre empresas, determinando que la empresa Movistar cuenta con el mejor servicio de telefonía móvil a nivel nacional mientras que la empresa Claro se encuentra en un nivel intermedio dejando a CNT Ecuador como la empresa que posee un bajo nivel de aceptación por parte de los internautas.

Palabras claves: análisis de sentimientos, CRISP-DM, gráficos estadísticos, minería de datos y minería de texto.

INTRODUCCIÓN

El análisis de sentimientos o sentiment analysis es el estudio por el cual se determina la opinión de las personas en Internet sobre algún tema en específico en los que se incluyen debates sobre críticas positivas, negativas o neutras, abarcando temas que van desde productos, películas, servicios a intereses socio culturales.

En el caso particular de este proyecto de titulación, su estudio se basa sobre el servicio de telefonía móvil del país para realizar un análisis de sentimientos aplicando técnicas de minería de datos con sus respectivos resultados. La plataforma o red social escogida fue Facebook, debido a su alto nivel de usabilidad y dependencia de los usuarios al momento de interactuar y de emitir comentarios públicos en torno al ámbito del servicio de telefonía móvil, haciendo que el análisis de datos que se generan de los comentarios en dicha red social se vea como una oportunidad de indagar sobre el comportamiento de la opinión de los usuarios acerca de las diferentes compañías telefónicas, sus respectivos comentarios o críticas no solo son base fundamental que permite el crecimiento o desuso del servicio por parte de los clientes sino también un pilar de visión sobre posibles mejoras a futuro ya que a nivel nacional se debate mucho sobre la preferencia de los clientes al momento de obtener el mejor servicio de alguna telefonía móvil en particular. Lo que ha hecho que la opinión en línea se convierta en una especie de divisa virtual para los negocios que buscan comercializar y dar conocer sus servicios.

Es por esta razón que se planteó el estudio y análisis del presente proyecto aplicando la metodología CRISP-DM en todas sus fases, mediante la cual, se

analiza y se extrae la información más relevante al aplicar tres técnicas de minería de datos y poder determinar cuál es la más eficaz al momento de realizar un análisis y a su vez establecer que compañía telefónica tiene mejores criterios en su estatus sobre su servicio a nivel nacional.

Ubicación y contextualización de la problemática

Facebook es una red social donde millones de personas interactúan entre sí, difunden información de todo tipo, el mismo que se ha convertido en un medio donde las empresas ponen en oferta sus productos y servicios, siendo esta una oportunidad para la propagación de los mismos que ofrecen las diversas telefonías móviles del país, por medio de sus páginas oficiales: Movistar, Claro y Cnt Ecuador, donde difunden información por medio de publicaciones sobre sus productos y servicios, el mismo que cuenta con miles de comentarios de todo tipo por parte de los usuarios seguidores siendo estos positivos, negativos o neutros.

Planteamiento del problema

Según el Ministerio de Telecomunicaciones y Seguridad de la Información (2012), menciona que hoy en día el teléfono celular se encuentra presente en todos los sitios en los que se pasa mayor tiempo como el lugar de trabajo, calle, casa, etc. el mismo que se ha convertido en un factor de integración social, lo que ha permitido que los ciudadanos puedan comunicarse con sus familias que se encuentran en otras ciudades o en otros países. El mercado de telefonía móvil en Ecuador ha experimentado un aumento grande en toda su historia, como ningún otro servicio de telecomunicaciones, uno de los mecanismos más efectivos en la creación de competencia en el mercado de telefonía móvil es la portabilidad, es

decir la posibilidad que tienen los consumidores de cambiar de operador conservando su número. Los usuarios señalan que las causas del cambio responden a las promociones que ofrecen las otras operadoras, la cobertura o la gama de equipos disponibles.

En la actualidad millones de personas expresan opiniones, reseñas y comentarios acerca de diferentes temáticas a través de las redes sociales como Facebook donde se pueden incrementar sus publicaciones, dando a conocer las bondades y beneficios sobre los servicios y promociones que ofrecen las diferentes telefonías móviles en el país, cuyas expresiones van desde una crítica social hasta la recomendación o juicio de un producto o servicio, el problema surge cuando se emite el marketing publicitario de estas empresas telefónicas en la red social pero no se cuenta con la opinión o el criterio con el cual se recibe el mensaje de los seguidores, mensajes o post que generan grandes volúmenes de información, dificultando el análisis de las diversas opiniones de cada uno de sus seguidores con respecto al servicio que se ofrece motivando a que los usuarios cambien constantemente de una operadora a otra. Varios autores señalan que es más costoso conseguir un nuevo cliente que mantener uno antiguo debido a la constante competencia que existe entre las telefonías.

Hoy en día Facebook se ha transformado en un medio, donde las empresas telefónicas optan por aplicar diferentes estrategias de marketing, compitiendo de una manera agresiva entre operadoras, con el único fin de captar nuevos clientes y lograr una mayor relación con el consumidor con respecto a las experiencias que tienen con las operadoras. Experiencias que no son analizadas y que podrían generar un aporte positivo, neutro o negativo en el futuro de estas compañías.

Objetivos

Objetivo General

Analizar la extracción de opiniones positivas, negativas y neutras en comentarios de usuarios de la red social Facebook aplicando técnicas de minería de datos para determinar la aceptación del servicio de telefonía móvil en Ecuador.

Objetivos Específicos de investigación y resolución del problema

- Identificar conceptos de la minería de texto y sus respectivas etapas.
- Extraer información de las empresas telefónicas: Movistar, Claro y CNT Ecuador de la red social Facebook.
- Aplicar la metodología CRISP-DM para el análisis y el desarrollo de esta propuesta.
- Aplicar gráficos estadísticos a los reportes de resultados.
- Determinar por medio del análisis de sentimientos de la red social Facebook que empresa telefónica posee un mejor estatus a nivel nacional.

Justificación

Hoy en día la red social Facebook se ha transformado en un medio, donde las empresas pueden optar por aplicar sus estrategias de marketing, con el fin de lograr una mayor relación entre los usuarios que se benefician del servicio de telefonía móvil y los seguidores de dicha red social, algo que se debate mucho en el país es la diversa competitividad que existe entre las diferentes empresas telefónicas, siendo una constante interrogante sobre cual tiende a tener mejores comentarios sobre el servicio ofertado dependiendo de los criterios de los usuarios en Facebook, para esto es necesario saber que opinan las personas al momento de

tener una experiencia, si este es positivo, darán paso a una relación constante y tendrá buenas críticas al momento de recomendar sobre el servicio ofertado en la red social en el que se le permita expresar este sentimiento.

Debido a que los datos extraídos de la red social Facebook son información pública, ya que la persona que los publica expresa su opinión de forma libre y voluntaria en las redes sociales, siendo una información valiosa la cual puede ser extraída y utilizada para fines positivos. Así uno de los usos es el empleo del análisis de sentimientos en cuestión con el propósito de conseguir, a través de los posts y comentarios de usuarios de las diferentes empresas de telefonía móvil del Ecuador en la red social Facebook, la extracción de información por medio de herramientas como Facepager para poder procesar y analizar el sentimiento u opinión: positiva, negativa y neutra.

Los resultados de este análisis de sentimientos aplicando técnicas de minería de datos servirán para poder determinar qué compañía telefónica brinda un mejor servicio en base a los comentarios y posts de los usuarios seguidores de la red social Facebook que han tenido diversas experiencias del servicio en el Ecuador.

CAPÍTULO I

MARCO TEÓRICO

Introducción

Este capítulo define la minería de datos y la importancia que genera los grandes volúmenes de información en las redes sociales, su estructura y como se analizan estos datos a partir de técnicas de estudio basados en análisis de texto.

Hoy en día las redes sociales se han vuelto parte de nuestra vida diaria, con un avance tecnológico que está en constante crecimiento y este a su vez genera grandes volúmenes de datos de manera automática y de forma continua, las cuales pueden tener estructuras creadas por base de datos o en forma de texto, teniendo en cuenta que en los datos almacenados existe una gran posibilidad de develar información que es de mucha importancia en la que se puede adquirir lo más relevante mediante el análisis de datos. Existe una gran complejidad para realizar estos análisis de forma manual por el hecho de contener grandes volúmenes de información, para ello se han desarrollado tecnologías especializadas en el análisis de estas y una de ellas es la minería de texto dentro del campo de la minería de datos, contando con métodos útiles para la extracción de texto, siendo considerada como una herramienta útil para el análisis de grandes volúmenes de información.

1.1 Antecedentes de investigaciones relacionadas al tema presentado

1.1.1 Análisis de opiniones

Para Jacobo (2016) en su publicación de análisis automático de opiniones de productos en redes sociales, expone sobre el análisis de opiniones de usuarios a cerca de productos o servicios que proporciona una empresa, siendo una actividad que se ha realizado tradicionalmente de diferentes formas, en distintas etapas y contextos, ejemplo de ello son los estudios de mercado, estudios de impacto de un producto en el mercado, análisis de resultados, etc. Las herramientas utilizadas van desde los formularios en papel, entrevistas y sondeos, hasta formularios electrónicos; aunque gracias a los avances en el procesamiento del lenguaje natural, se abren nuevas posibilidades para el análisis de opinión mediante las redes sociales. El problema se resuelve mediante una metodología y se desarrolla una aplicación que permite el análisis de textos cortos de opinión, clasificándolos en muy positivos, positivos, neutros, negativos, muy negativos y sin opinión o sentimiento.

1.1.2 Análisis de sentimientos en la red social Twitter

Los sistemas de análisis de sentimientos y la minería de opiniones han resultado ser de gran utilidad en los últimos años, con la introducción de las redes sociales, su principal objetivo es identificar opiniones positivas o negativas en textos generados por usuarios y sobre qué entidad o aspecto de esta se han realizado. Según Becerra (2016) en su publicación sobre el análisis de sentimientos en Twitter de lo bueno y lo malo, se analiza grandes volúmenes de opiniones generadas por usuarios en dicha red social, permitiendo al analista procesarlas de forma rápida y efectiva.

En su tesis de análisis de sentimientos en la red social Twitter Viteri (2016), determinó que se pueden obtener nuevos socios que aporten al crecimiento económico del club deportivo Barcelona S.C. del país, aplicando un modelo de análisis en el que se extraen los comentarios emitidos por los socios del club en la red social Twitter, con respecto a los planes presentados para mejoras del club deportivo y en base a ello determinar mediante respuestas positivas o negativas, los aportes necesarios para tener la mayor captación de personas y obtener nuevos socios interesados en el club.

Para García (2014) predecir eventos en base a las opiniones de personas en internet es de vital importancia, es por ello que realiza un análisis de sentimientos en el que emplea parte de estas técnicas escogiendo la red social Twitter, realizando un estudio predictivo sobre las elecciones presidenciales, buscando predecir resultados a favor por los candidatos postulantes y quien tendría una mayor ventaja con respecto de los que estén en contra mediante tweets, determinando la polaridad del mensaje pudiendo ser positiva, negativa o neutra.

1.1.3 Análisis de sentimientos en la red social Facebook

Actualmente las publicaciones en Facebook escritas en español pueden contener información semántica relevante, para determinar un sentimiento de manera automatizada mediante la herramienta WordNet-Affect y un clasificador Naive Bayes, que identifica en las publicaciones emociones de alegría, tristeza y enojo, de esta manera lo han explicado Acevedo et al. (2014) en su artículo Arquitectura Web para el análisis de sentimientos en Facebook con enfoque semántico, donde los resultados experimentales muestran una precisión del 63%,

en palabras relacionadas con el concepto de amor que estaban implícitamente relacionadas con la alegría.

Para la marca de fideos Cayambe (Narvaez, 2017) es importante conocer la opinión de las personas con respecto a sus productos, es por ello que aplicaron técnicas de minería de datos para realizar un análisis de sentimientos en la red social Facebook en el que determinaron una percepción positiva con respecto a las diversas opiniones de sus productos, en la cual aplicando varias técnicas de minería, se determinó que la más idónea es la técnica de árboles de decisión para el análisis y procesamiento de la información, debido a que cuenta con el mayor porcentaje de correlación entre las variables.

1.2 Definiciones conceptuales

1.2.1 Minería de datos

La minería de datos surge para ayudar a indagar el contenido de una base de datos; para lo cual hace uso de experiencias estadísticas y de algoritmos de investigación parecidos a la inteligencia artificial y a las redes neuronales; cuyo objetivo consisten en la interconexión de neuronas y la colaboración entre ellas, dando como resultado un estímulo de salida, lo que permitirá realizar un estudio de los resultados obtenido a fin de determinar el patrón o tendencia de los datos (Jácome, 2017).

Dentro del contexto de minería de datos se tienen las siguientes definiciones:

- La minería de datos es el proceso de detectar la información procesable de grandes conjuntos de datos, utilizando el análisis matemático para deducir

patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiados datos (Microsoft, 2018).

- Minería de datos es la extracción de conocimiento (patrones, tendencias, modelos) en bancos de datos, enfocado a un análisis de tipo predictivo. El concepto de Knowledge Discovery in Databases (KDD) comprende un área similar. En muchas ocasiones se usan indistintamente, aunque también se usa el término Data Mining para referirse específicamente a la etapa analítica dentro del KDD (Mikel et al., 2017).
- Es una disciplina de las ciencias e ingenierías de la computación que intenta hallar patrones significativos en conjuntos de datos para producir modelos descriptivos, predictivos y clasificadores apoyándose en técnicas de manejo y programación de bases de datos, en estadística y aprendizaje automático (Coria, 2016).
- Se define como un conjunto de técnicas y métodos que permiten investigar grandes volúmenes de información utilizando herramientas para análisis de datos como RapidMiner, donde el objetivo radica en encontrar patrones de datos iguales (Jácome, 2017).

Analizando las definiciones anteriormente citadas, se concluye que la minería de datos es un conjunto de técnicas de análisis, en la cual se extrae y analiza la información de grandes volúmenes de datos, por medio de la identificación de patrones, aplicando técnicas de aprendizaje automático y estadístico, se detectan patrones que no son visibles o fáciles de percibir mediante la exploración tradicional por los grandes volúmenes de información, en la cual

aplicando técnicas de minería de datos se puede producir modelos descriptivos, predictivos y clasificadores.

1.2.2 Técnicas de minería de datos

Existen dos clasificaciones o técnicas en la minería de datos tales como: supervisados y no supervisados. En la tabla 1 se analiza la clasificación de las técnicas de minería de datos, en la cual determina que los supervisados son algoritmos que tienen entradas de datos cambiantes por lo cual el resultado va en función de dichas entradas, lo que significa que estos deben adaptarse a las nuevas variables para obtener nuevos resultados, mientras que los no supervisados son algoritmos que no necesitan de la intervención humana para interpretar resultados, al momento de pasar por un proceso riguroso de evaluación su respuesta siempre será satisfactoria.

Tabla 1. Clasificación de las técnicas de minería de datos

Supervisados	**No supervisados**
Red neuronal	Agrupamiento (“clustering”)
Árboles de decisión	Reglas de asociación
Regresión	Minería de texto (“text mining”)

Siguiendo la clasificación de las técnicas de minería de datos divididos en supervisados y no supervisados, se detallan algunos algoritmos de c/u de ellos:

- *Modelo de red neuronal, e*s una técnica de minería de datos predictiva que imita las redes de inteligencia del cerebro, las cuales resultaban difíciles o

imposibles para las maquinas lógicas secuenciales. Su interés radica en que son herramientas útiles que permiten realizar predicciones, por lo que son usadas por muchas aplicaciones para el procesamiento de un gran número de elementos altamente interconectados.

- *Modelo de árboles de decisión,* es un modelo de predicción que representa de forma gráfica una serie de reglas que sirven para la toma de decisiones en la asignación de un valor de salida a un determinado registro, permitiendo una clasificación controlada en la cual los nodos intermedios son los atributos de entrada, las ramas representan valores de dichos atributos y los nodos finales valores de la clase.
- *Modelo de regresión lineal, e*s una técnica predictiva que estudia el cambio de la variable independiente en relación de la variable dependiente, buscando patrones para determinar un valor único.
- *Modelo agrupamiento (K-Means), e*sta técnica de minería de datos no supervisada detecta agrupamientos o estructuras intrínsecas en el conjunto de datos, la cual se divide en subgrupos llamados clases que a su vez se diferencian por su máxima distancia de separación entre ellas, identificando grupos homogéneos de individuos parecidos que tengan características comunes.
- *Reglas de asociación,* permite organizar los atributos según las relaciones, expresando afinidades entre elementos; por ejemplo, ventas de mercancías que son especialmente interesantes en las rutinas de compra de los consumidores y constituye la relación entre lo que se oferta asociado con las ventas.

- *Minería de texto,* es una disciplina englobada dentro de las técnicas de acceso, recuperación y organización de información, en la cual se obtiene información nueva a partir de grandes cantidades de texto, donde la información suele no estar estructurada.

1.2.3 Minería de texto

La minería de texto (text mining) es una disciplina englobada dentro de las técnicas de acceso, recuperación y organización de información y consiste en un conjunto de técnicas que nos permiten extraer información relevante y desconocida de manera automática dentro de grandes volúmenes de información textual, normalmente en lenguaje natural y generalmente no estructurado (Belinchón, 2015).

La minería de texto es una aplicación de la lingüística computacional y del procesamiento de texto que pretende facilitar la identificación y extracción de nuevo conocimiento a partir de colecciones de documentos textuales (Moreno, 2017). Cabe recalcar que es un proceso encargado del descubrimiento de información que no se puede obtener de forma evidente al revisar los documentos y se aplica a cualquier tipo de texto ya sea en la web, correos electrónicos o en un texto simple.

1.2.4 Etapas de la minería de texto

La minería de texto tiene como objetivo extraer el significado, los conceptos, la relación de conceptos, patrones ocultos, opiniones, para presentarlo de una manera comprensible para el usuario.

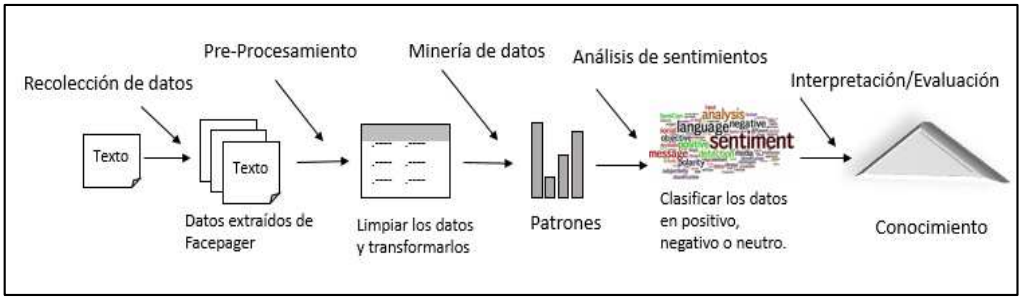


Gráfico 1. Etapas del proceso de minería de texto.

En el gráfico 1, se define el proceso de minería del texto, que consiste en las siguientes etapas:

- *Recolección de datos,* en esta etapa se realiza una recolección de datos específicamente comentarios de usuarios en la red social Facebook de las empresas telefónicas de Movistar, Claro y CNT Ecuador, usando la herramienta Facepager.
- *Preprocesamiento, e*n esta etapa los textos se transforman a algún tipo de representación estructurada o semiestructurada, se eliminan todas las etiquetas innecesarias y caracteres especiales, los párrafos se dividen en forma de oraciones para hacer simetría de oraciones, luego los datos no estructurados se transforman en forma estructurada para separar las características juntos con sus parámetros concernientes que facilite su posterior análisis.
- *Minería de datos,* en esta etapa las representaciones internas se analizan con el objetivo de descubrir patrones o nueva información, esto se puede realizar mediante técnica de minería de datos como la clasificación, clustering, machine learning y procesamiento del lenguaje natural se aplican a los atributos seleccionados para extraer el conocimiento concerniente.

- *Análisis de sentimientos,* en esta etapa se realizan las respectivas clasificaciones de los resultados obtenidos de la etapa anterior para conocer si son negativas, positivas o neutras las opiniones de los usuarios.
- *Interpretación/Evaluación,* en esta etapa se evalúan los resultados obtenidos en el proceso de descubrimiento de un nuevo conocimiento.

1.2.5 Aplicaciones de minería de texto en diferentes campos

Minería de texto es una variante de la minería de datos, donde adopta técnicas de aprendizaje automático para el reconocimiento de patrones y la comprensión de la nueva información. Tiene aplicaciones en diferentes campos, como en la medicina, biología, gestión documental, análisis de sentimientos o minería de opiniones, extracción de información, elaboración de resúmenes y extracción del conocimiento.

- *Medicina y biología,* las enfermedades respiratorias se consideran un gran reto para la salud y para los sistemas sanitario, debido al gran coste económico y social que representan. Por otra parte, la minería de texto permite analizar grandes cantidades de información mediante sistemas optimizados relacionados con la salud, enfermedad, diagnósticos, gravedad del caso, resultados analíticos, pruebas funcionales y medicación.
- *Gestión documental,* la mayor parte de información que manejan las empresas corresponde a información no estructurada; ejemplo, contratos, presupuestos, correos electrónicos, informes, órdenes de compras, facturas recibos, materiales de marketing, artículos de prensa, etc. Esta información no estructurada representa un alto valor en la toma de decisiones en una

empresa, para ello existe la minería de texto que permite extraer conocimientos a partir de grandes volúmenes de documentos, para posteriormente poder crear repositorios de documentos, controlar el acceso, realizar seguimiento, etc.

- *Análisis de sentimientos,* es una de las técnicas de mayor interés para evaluar la opinión de usuarios en redes sociales con el fin de conocer si los mensajes contienen emociones positivas, negativas o neutras. Un ejemplo en la red social Twitter mediante el streaming se puede clasificar el tono de los mensajes publicados por los usuarios para determinar el impacto en distintas ciudades dentro de un país. Se aplican diccionarios de sentimientos que brinden la valoración a cada palabra y mediante el análisis de cada una de ellas se conocen si las ciudades tienen un sentimiento positivo, negativo o neutro.

De esta manera se puede observar que la minería de texto es muy útil dentro de las organizaciones, empresas y administraciones en general, dentro de ellas se generan gran cantidad de documentos, que les sirve para obtener información a partir de todo ese volumen de datos. Esto le puede servir para conocer mejor a sus clientes y así poder brindar un mejor servicio.

1.2.6 Análisis de sentimientos

El análisis de sentimientos también conocido como minería de opiniones tiene la tarea de clasificar automáticamente un texto escrito en un lenguaje natural, en un sentimiento positivo o negativo, opinión o subjetividad (Villena, 2015). Además, se refiere a los diferentes métodos de lingüística computacional que

ayudan a identificar y extraer información subjetiva del contenido existente en redes sociales, foros, web, etc.

El análisis de sentimientos trata de una clasificación masiva de documentos de una forma automática, que se centra en detallar los documentos en función del vínculo positivo o negativo del lenguaje ocupado en el mismo.

Hoy en día con el uso de las redes sociales se permite a los usuarios con facilidad mostrar sus opiniones sobre cualquier tema, producto o servicio y con esto las empresas pueden medir su impacto de forma inmediata y poder obtener ventajas competitivas en diferentes ámbitos y aplicar estrategias para mejorar su imagen.

Para Intelligent (2017) la cantidad de datos que se generan actualmente en las empresas mediante las redes sociales está creciendo a un ritmo impresionante, y obtener información útil y valiosa de ellos supone una ventaja competitiva muy importante respecto a los competidores. Para poder obtener un resultado de los comentarios que realizan los seguidores en las diversas redes sociales, se aplica lo siguiente:

1. Filtración de datos, en primer lugar, se utilizan las palabras claves para descartar contenido no deseado, y posteriormente se establecen palabras para obtener categorías según su polaridad o su procedencia.
2. Extracción del contenido, una vez que pasen el filtro, se elimina el contenido no deseado y se comenzará a trabajar con el contenido de calidad.
3. Análisis de contenido, este proceso lo puede realizar el algoritmo o una persona física en sí, aquí el contenido útil y de calidad quedará encuadrado en la categoría que le corresponda.

4. Limpieza del contenido, quizás se haya añadido contenido erróneamente, y este es el momento de enviarlo a su categoría correcta o descartarlo directamente.
5. Revisión, se gestionarán en este apartado todos los posibles aspectos a mejorar, tal vez se encuentre una nueva palabra a incluir para descartar contenido, o se observe alguna palabra considerada positiva se utiliza a modo negativo en determinados momentos.

1.2.7 Redes Sociales

Las redes sociales son la forma de comunicación e interacción entre diferentes usuarios en un contexto social, a partir de la cual se generan conversaciones, comentarios, likes y se puedan compartir publicaciones. Hoy en día el uso de las plataformas es cada vez más intuitivo y accesible con gran influencia en los seres humanos de cualquier género y edad.

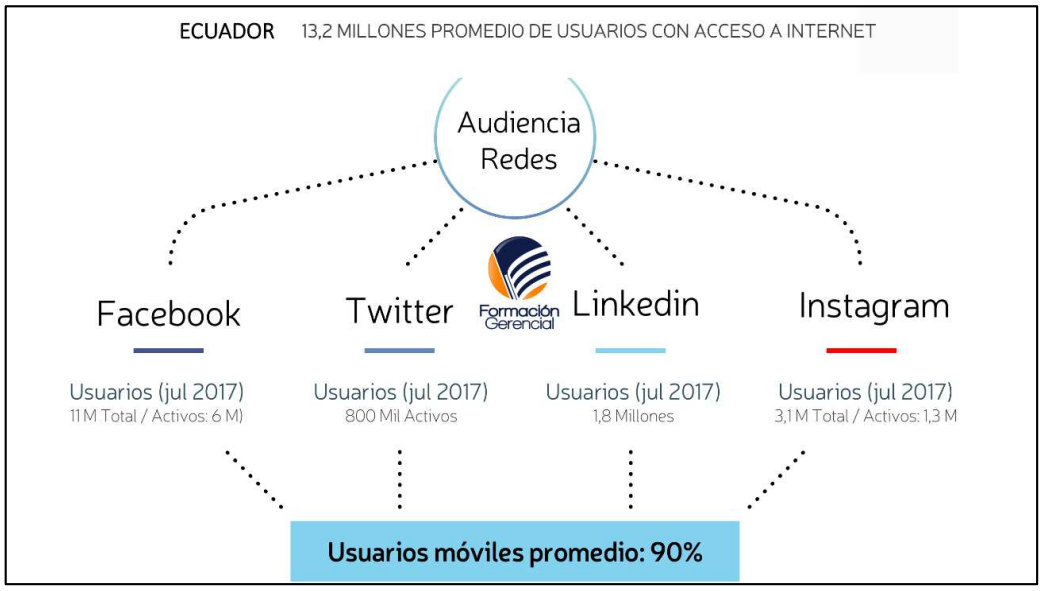


Gráfico 2. Usuarios de Internet y Redes Sociales en Ecuador Julio 2017, (Alcazar, 2017).

Como se muestra en el gráfico 2, en Ecuador el ranking de usuarios conectados a internet y a las redes sociales en el año 2017 es de más de 13 millones

de usuarios que acceden regularmente a Internet tanto desde dispositivos móviles como de escritorio (incluyendo cifras de espacios públicos, "Escuelas del Milenio"). En términos de redes sociales, Facebook mantiene el liderazgo absoluto con 11 millones de usuarios registrados en Ecuador, de los cuales, un promedio de 6,4 millones está activos mensualmente (acceden a la plataforma), seguido por Instagram, LinkedIn y Twitter. Todas estas redes son utilizadas principalmente desde dispositivos móviles (Alcazar, 2017).

Hoy en día la red social Facebook es la más utilizada por los ecuatorianos, debido que permite hacer amigos de diferentes partes del mundo, posee aplicaciones y juegos, crear paginas o grupos para vender productos u ofrecer sus servicios, un ejemplo son las empresas telefónicas, Movistar, Claro y CNT Ecuador que a través de páginas en Facebook pueden explotar esta ventaja para influenciar de manera positiva con sus servicios a sus seguidores y así ellos pueden obtener un comentario positivo, negativo o neutro acerca de los servicios que oferten las empresas telefónicas.

1.2.8 Herramienta Facepager

Facepager fue creado para buscar datos públicos disponibles de Facebook, Twitter y otra API basada en JSON. Todos los datos se almacenan en una base de datos SQLite y pueden exportarse a csv (Strohne, 2018). La última versión 3.9.1 fue lanzada el 08 de enero, sus nuevas funciones son la siguientes:

- Módulo de YouTube. Hay muy pocos estudios de YouTube, especialmente en comparación con los Estudios de Twitter.

- Sistema preestablecido reelaborado. Los pre-ajustes se cargan directamente desde GitHub. Los pre-ajustes básicos se actualizan para reflejar los cambios en la API de Facebook.
- Campos de encabezado en módulo genérico y módulo de archivos, de esta forma, puede interactuar con las API de Googles Cloud, debido que necesita autorización por campos de encabezado (token de portador).
- Interfaz de configuración unificada en todos los módulos.
- Barra de estado que se puede hacer clic: abrir la carpeta predefinida o la carpeta de la base de datos
- Tarea y manejo de errores mejorados.
- Corrección de errores y mejoras menores.

1.2.9 RStudio

RStudio es un entorno de desarrollo integrado (IDE) para R. Incluye una consola, editor de resaltado de sintaxis que admite la ejecución directa de código, así como herramientas para trazado, historial, depuración y gestión del espacio de trabajo.

RStudio está disponible en ediciones comerciales siendo este de código abierto y se ejecuta en el escritorio (Windows, Mac y Linux) o en un navegador conectado a RStudio Server o RStudio Server Pro (Debian / Ubuntu, RedHat / CentOS y SUSE Linux) (RStudio, 2018).

RStudio es una herramienta potente que soporta procedimientos y técnicas requeridas para análisis de calidad y dignos de confianza. Al mismo tiempo, pretende ser sencillo e intuitivo como sea posible, proporcionando un entorno

amigable. A continuación, se detallan algunas de las ventajas que presenta RStudio:

- Es una herramienta que respeta la tradicional consola en R.
- Es multiplataforma y se puede ejecutar en el escritorio (Linux, Mac, Windows) o incluso a través de internet mediante RStudio Server.
- Permite abrir varios scripts a la vez.
- Ejecuta pedazos de código con sólo marcarlo en los scripts.
- Dispone de autocompletado de código.
- Presenta facilidades para codificar: extract function, coment/uncoment lines, reindent, etc.

1.2.10 Metodología CRISP-DM

La metodología CRISP-DM es considerado un estándar en los proyectos de minería de datos. La metodología se encuentra estructurada en seis fases: comprensión del negocio, comprensión de datos, preparación de datos, modelado, evaluación e implantación. Incluye una serie de bucles de retroalimentación entre las fases, esto con el objetivo de obtener resultados fiables y consistentes (Buenaño et al., 2016).

Según Piatetsky (2014) en base a datos de la última encuesta reflejada en el gráfico 3, sobre la comparación de la metodología principal que se está utilizando para proyectos de análisis, extracción de datos o ciencias de datos, se determinó que la metodología más utilizada es CRISP-DM.

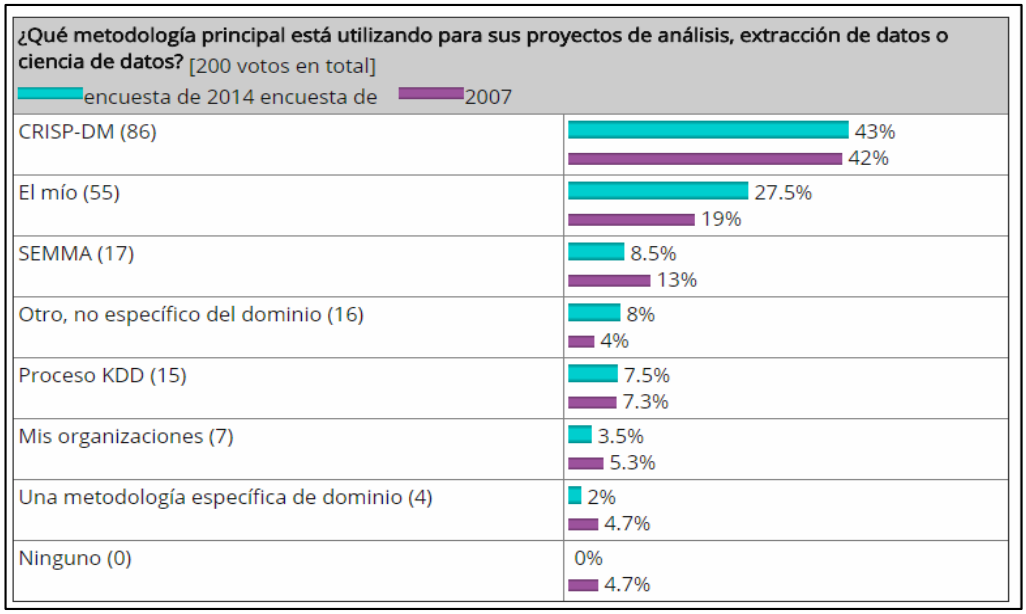


Gráfico 3. Metodologías utilizadas en Data Mining, (Kdnuggets, 2014).

CRISP-DM sigue siendo la mejor metodología para proyectos de minería de datos, como está reflejado en el gráfico 3 con un porcentaje del 43% del año 2014 en comparación del 2007 que fue del 42% con respecto a la principal metodología que se está utilizando para proyectos de análisis, extracción de datos o ciencia de datos.

CRISP-DM es una metodología estándar basada en una forma de expresar la compresión del problema modelando las decisiones en base al análisis del diseño, componiéndose en 6 fases de alto nivel de manera bidireccional, en la cual indica la secuencia que se debe realizar para alcanzar los objetivos. Es bidireccional porque permite avanzar o regresar hacia una fase en caso de que esté inconcreta o se hayan definido mal los objetivos como se refleja en el gráfico 4 representando las 6 fases de la metodología.

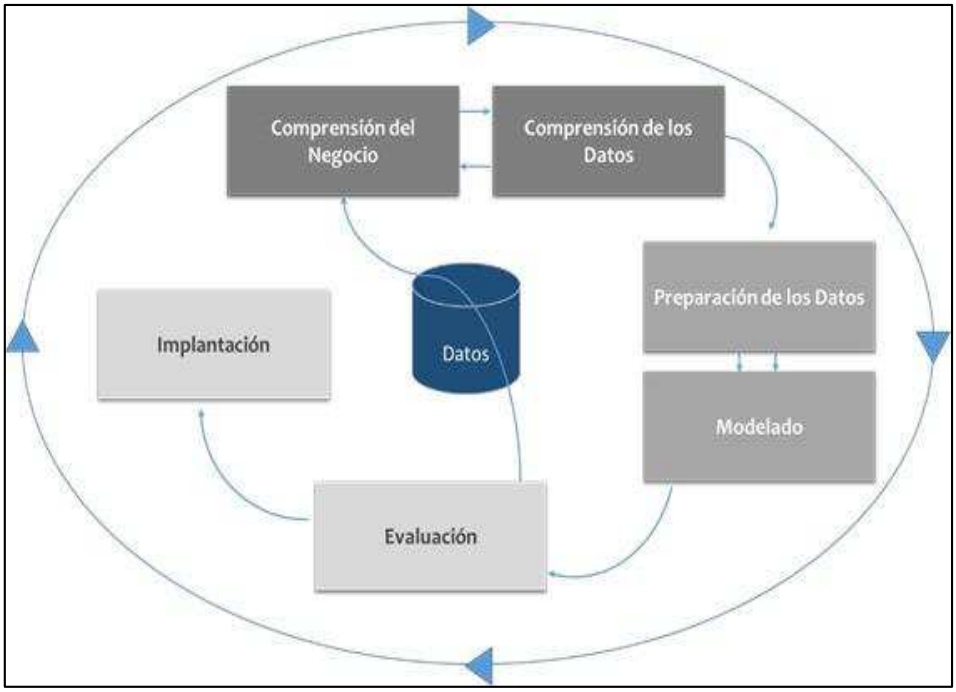


Gráfico 4. Fases del modelo de referencia CRISP-DM, (CRISP-DM, 2000).

Siguiendo las fases del modelo de referencia CRISP-DM como se detalla en el gráfico 4. A continuación, se analiza cada una de ellas:

- *Comprensión del negocio*, en esta fase se establecen los objetivos y requerimientos de la minería de datos, realizando la recopilación inicial de datos, descripción de los datos, exploración y verificación de la calidad de los datos de una perspectiva no técnica.
- *Preparación de los datos,* para la preparación de los datos se detectan problemas de calidad de estos y en base a ello se obtienen los primeros insights, subconjunto de datos que a su vez vienen siendo las primeras hipótesis.
- *Modelado,* para la fase de modelado se realiza la selección de la técnica del modelado correcto para aplicar las técnicas de minería de datos.
- *Evaluación,* en esta fase se evalúan los resultados, se revisa el ciclo del proceso y si existen problemas o errores que muestren resultados no

deseados se vuelve a la primera fase del modelo para corregir, por ello es considerado un modelo bidireccional.

- *Implantación,* esta fase permite realizar la planificación de la monitorización y del mantenimiento, generando un informe final y a su vez se procede a la revisión del proyecto.
- *Etapa de Pruebas del Sistema,* define las pruebas a realizar y a ejecutar.

Las fases ya mencionadas en el gráfico 4, son de vital importancia en el momento de aplicar la metodología CRISP-DM para la extracción, análisis y limpieza de los datos, recordando que los procesos son bidireccionales permitiendo avanzar o regresar a una fase en concreto según los cambios o mejoras que vayan surgiendo.

CAPÍTULO II

MARCO INVESTIGATIVO

2.1 Método de Investigación Teórico-Analítico

El método utilizado en esta investigación es teórico analítico siendo un método que consiste en la desmembración de un todo, descomponiendo en partes o elementos para observar detalladamente las causas que lo provocan. Para este caso se utilizará como prueba de análisis de sentimientos datos públicos de la telefonía móvil Tuenti Ecuador en su perfil oficial de la red social Facebook, para poder tener una visión previa al análisis original de esta investigación y entender su comportamiento con ideas y teorías que aporten al análisis original.

2.2 Herramientas de recolección de datos

2.2.1 Find your Facebook ID

Esta herramienta ayuda a encontrar fácilmente el ID personal de Facebook para las operaciones de la API gráfica (interfaz de programación de aplicaciones), complementos sociales e integraciones de ciertos de ellos, como el botón "me gusta" y "cuadro de me gusta", entre otros, Facebook requiere que conozca su ID de usuario. Desafortunadamente hace que esto sea muy difícil de encontrar especialmente si existen URL de perfil personalizado y dependiendo de las necesidades, la obtención de un identificador único permitirá facilidad para la extracción de datos, en este caso como prueba de estudio sobre el funcionamiento de la herramienta, se obtendrá un identificador del perfil oficial de Tuenti Ecuador de la red social Facebook como se refleja en el gráfico 5.

Gráfico 5. URL de la página oficial de Tuenti Ecuador.

Gráfico 6. Prueba del buscador y resultado generado de Facebook ID personal numérico.

Se ingresa en la página https://findmyfbid.com/ donde se extrae el identificador único de la página oficial como se muestra en el gráfico 6.

2.2.2 Facepager

Facepager fue creado para buscar datos públicos disponibles de Facebook, Twitter y otra API basada en JSON. Todos los datos se almacenan en una base de datos SQLite y pueden exportarse a csv (Jünger et al., 2017).

Por medio del ID que se obtuvo de la página del perfil oficial de Tuenti Ecuador, se procede a utilizar el identificador en la herramienta Facepager como se ve reflejado en el gráfico 7.

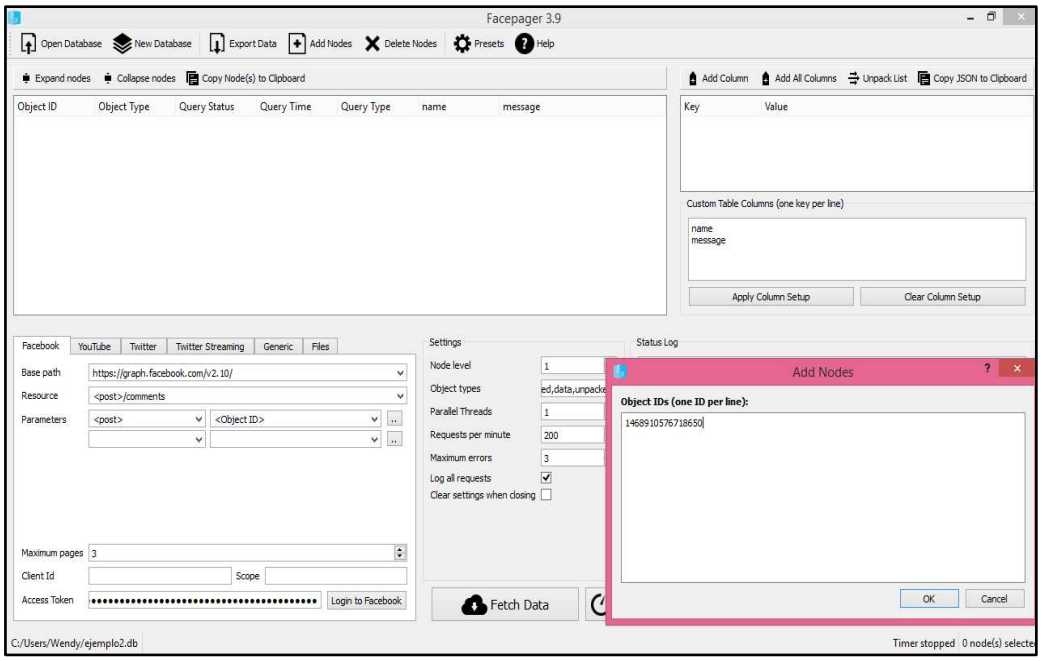


Gráfico 7. Ingreso del ID del perfil oficial de Tuenti Ecuador en la herramienta Facepager.

Una vez ingresado el identificador del perfil de Facebook oficial de la telefonía móvil Tuenti Ecuador, se procede a la obtención de posts (publicaciones) y comentarios de usuarios en la red social como se muestra en el gráfico 8, desde el 1 de junio del 2018 hasta el 30 de junio del 2018, donde se encontró un aproximado de 658 nodos entre posts (publicaciones) y comentarios, probando así la eficiencia y eficacia de la herramienta y la facilidad de obtención de información para poder realizar el análisis de sentimientos.

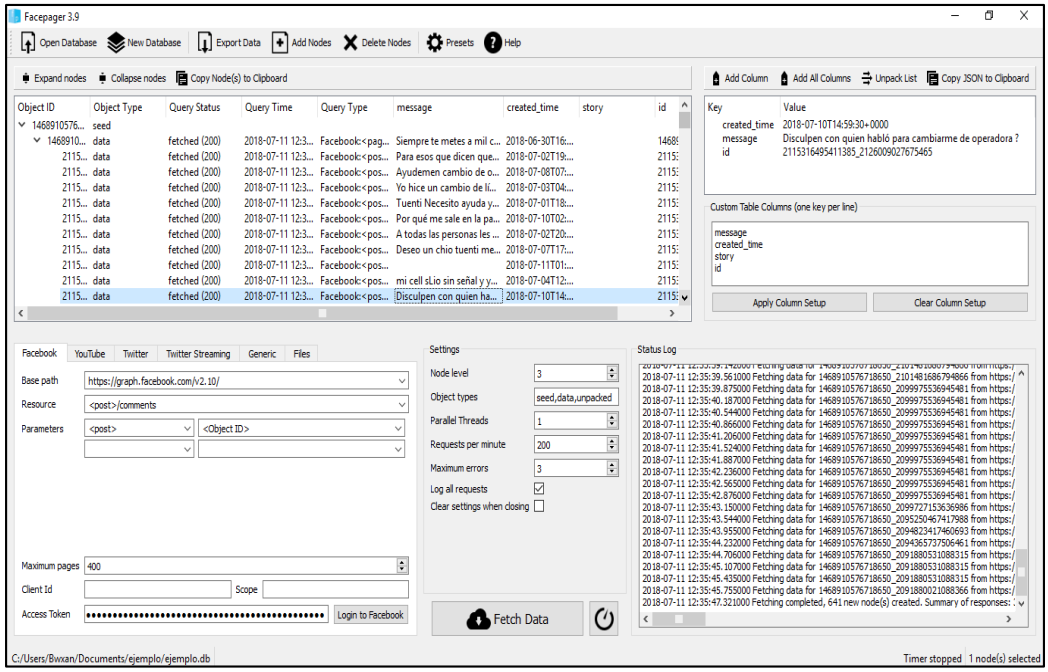


Gráfico 8. Extracción de posts y comentarios de Tuenti Ecuador a en la herramienta Facepager.

2.3 Fuentes de información de datos de prueba

Por medio de la red social Facebook desde el perfil oficial de Tuenti Ecuador se analizaron los posts (publicaciones) y comentarios de manera breve, desde el 1 de junio del 2018 hasta el 30 de junio del 2018 con un aproximado de 658 opiniones públicas.

2.3.1 Tuenti Ecuador

Tuenti, la nueva marca de la concesionaria Otecel, dueña también de Movistar, busca consolidarse en el mercado de los servicios móviles ofreciendo navegación y acceso a redes sociales al público juvenil.

El servicio a los clientes se ofrece mediante un chip y hay tres opciones de paquetes a escala nacional y Ecuador es el quinto país a donde llega la marca de

telefonía móvil que en la modalidad prepago permite navegación, llamadas y mensajería (ElUniverso, 2015).

Desde el perfil oficial de Tuenti Ecuador en la red social Facebook, se extrae la información de posts (publicaciones) y comentarios de los usuarios seguidores del perfil @TuentiEC que brinda información de servicios y promociones y tratan de solucionar problemas en los que los usuarios reflejan su malestar con comentarios de todo tipo y que a su vez no son analizados por la gran cantidad de información constante que emiten los usuarios seguidores.

2.4 Recolección y tabulación de datos previo análisis

Mediante la herramienta Facepager como ejemplo se obtuvieron datos del perfil oficial de la telefonía móvil Tuenti Ecuador en la red social Facebook para probar la efectividad de esta.

2.4.1 Tabulación de prueba Tuenti Ecuador

En la tabla 2, se muestran los datos obtenidos del perfil en Facebook @TuentiEC por medio de la herramienta Facepager, el cual consta con un total de 658 datos entre publicaciones y comentarios obtenidos del 1 de junio del 2018 al 30 de junio del 2018 como prueba de un previo análisis.

Tabla 2. Prueba de análisis Tuenti Ecuador.

Telefonía móvil Tuenti Ecuador (junio 2018)			
Entradas	1 al 15 de junio	16 al 30 de junio	Total
Publicaciones	18	9	27
Comentarios	423	218	641
Total	431	227	658

2.5 Análisis e interpretación de los datos de prueba

En la tabla 2, se refleja el número de datos de opinión pública obtenidos como prueba de una representación del funcionamiento de la herramienta Facepager aplicando el método teórico – analítico, el cual consta con un total de 658 datos, destacando los días 15 primeros días del mes de junio del 2018 como los días que hubo mayor actividad en el perfil @TuentiEC entre posts y comentarios de los datos obtenidos del 1 de junio del 2018 al 30 de junio del 2018.

Previo a la interpretación de los datos de prueba se procede a utilizar la herramienta Facepager con los datos de opinión pública de las diferentes empresas telefónicas del país: Movistar, Claro y CNT Ecuador de los perfiles oficiales que se encuentran en la red social de Facebook aplicando la metodología CRISP-DM siguiendo su estructura para el análisis de sentimientos.

CAPÍTULO III

MARCO PROPOSITIVO

Para la investigación presente en la aplicación de técnicas de minería de datos para el análisis de sentimientos de las diferentes empresas telefónicas del país: Movistar, Claro y CNT Ecuador de la red social Facebook, se aplicará cada una de las fases de la metodología CRISP-DM detallando su desarrollo de inicio a fin.

3.1 Comprensión del negocio

Facebook es un medio, donde las empresas telefónicas optan por aplicar diferentes estrategias de marketing, compitiendo de una manera agresiva entre operadoras, con el único fin de captar nuevos clientes y lograr una mayor relación con el consumidor en base a las experiencias que tienen con el servicio que ofertan las diferentes telefonías móviles del país.

En la actualidad, el teléfono móvil está presente en todos los lugares en los que se pasa la mayor parte del tiempo, es decir, trabajo, calle, domicilio, etc., convirtiéndose en algo de índole cultural, incluso puede actuar como factor de integración social. Esta tecnología se encuentra en constante evolución, ofreciendo no solo nuevas versiones de software destinadas a un mayor rendimiento, sino que da la oportunidad de cambiarlo permanentemente, superando la transmisión de datos, video, fotos y audio.

El Ministerio de Telecomunicaciones y de la Sociedad de la Información, junto a sus entidades adscritas y relacionadas, continúa impulsando proyectos para

difundir y hacer más accesibles los servicios de telefonía en todas las provincias del Ecuador, con la finalidad de ser un país solidario e inclusivo (MINTELZ/RLBA, 2013).

3.1.1 Determinar los objetivos del negocio

Garantizar la masificación de las telefonías móviles del país, permitiendo mejorar la infraestructura de telecomunicaciones para brindar productos y servicios de la más avanzada tecnología a precios asequibles, con el fin de implementar iniciativas a nivel comercial que acerque más a las personas al uso de los dispositivos móviles.

3.1.2 Evaluación del negocio

El Ecuador se encuentra conformado por 14’483.499 millones de habitantes (Censo, 2010), los cuales a la hora de comunicarse pueden decidirse por distintos medios entre los que se encuentran: telefonía fija, telefonía móvil o por internet. El medio más utilizado y preferido por los ecuatorianos para comunicarse es a través del servicio móvil como una vía de acceso desde cualquier lugar y todo al alcance de sus manos. Para corroborar esta información se tiene que al cierre mayo del 2018 la penetración de la telefonía móvil fue del 89,54% según datos de la Agencia de Regulación y Control de las Telecomunicaciones (ARCOTEL, 2018). Mientras que el Servicio de Telefonía Fija (STF) alcanzó una densidad del 14,19% al cierre de mayo del mismo año con un total de 2.395.577 millones de abonados + prestación a través de terminales de telecomunicaciones de uso público (TTUP), notando una clara diferencia con respecto a la telefonía movil como el medio más utilizado y preferido por los ecuatorianos. (ARCOTEL, 2018)

Con respecto a la presencia de estas empresas de telecomunicación móvil en el mercado ecuatoriano: Movistar, Claro y CNT, en el 2014 Claro tenía un posicionamiento del 67% mientras que, Movistar era del 29% y CNT del 4% (Rosado, 2016). Pero con la reducción de líneas, Claro pasó del 67% del 2014 al 53,17% de mayo del 2018 en su posición en el mercado de telefonía móvil, mientras que, Movistar pasó al 30,32% y CNT al 16,51% del mismo año, esto como consecuencia de la demora en la asignación de espectro adicional para 4G. Donde el 53,17% representa un total de 8.03 millones de líneas activas en Claro (CONECEL), 4.58 millones de líneas en Movistar (OTECEL) y 2.49 millones en CNT telefonía móvil (ARCOTEL, 2018), como se muestra en el gráfico 9.

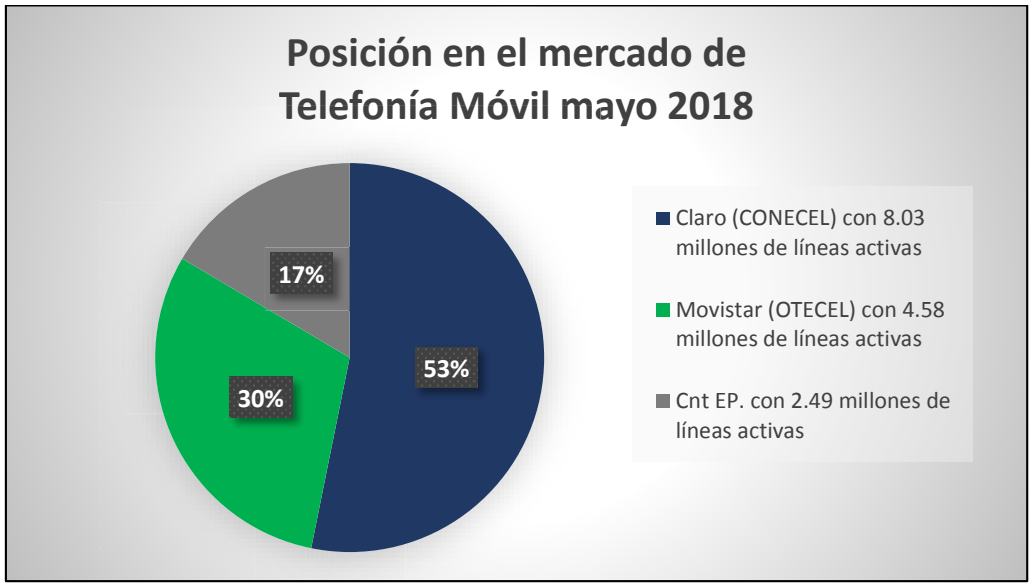


Gráfico 9. Posicionamiento de las diferentes telefonías móviles en el mercado mayo 2018.

Hasta julio del 2017, las empresas prestadoras del Servicio Móvil Avanzado: Claro, Movistar y CNT reportaron a la Agencia de Control y Regulación de las Telecomunicaciones, ARCOTEL, 15’055.240 líneas activas (ARCOTEL, 2017). El crecimiento de líneas en los últimos años es evidente, en

especial si se compara los datos del 2017 con los de 2008: en ese periodo se incrementaron 3’362.992 líneas móviles como se refleja en el gráfico 10.

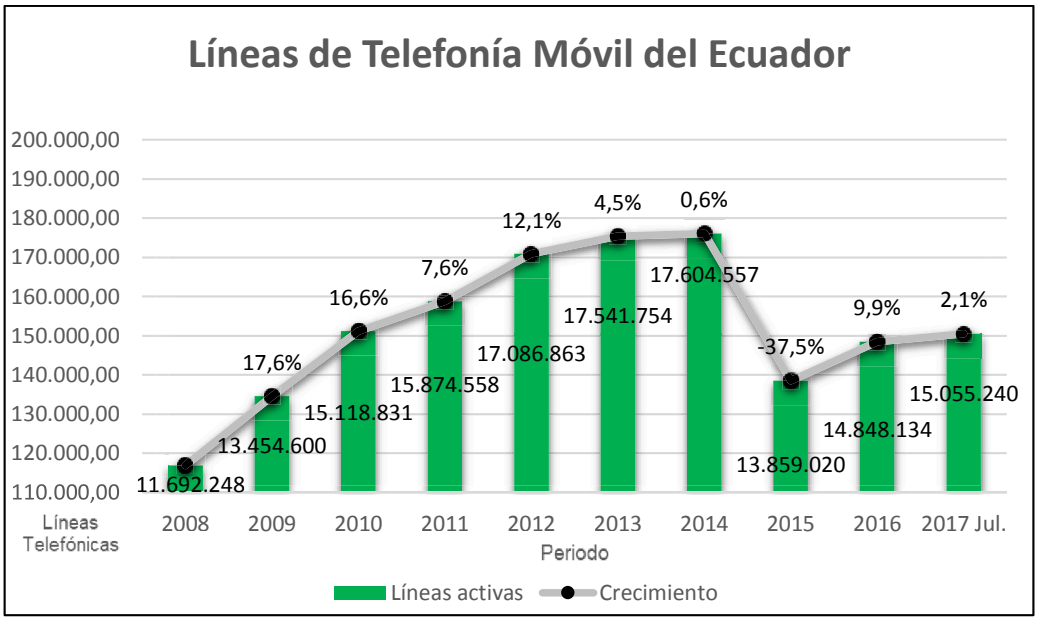


Gráfico 10. Crecimiento de líneas telefónicas en los últimos años.

En la tabla 3, se muestra de manera general el número de usuarios seguidores de las empresas Movistar, Claro y CNT Ecuador de la red social Facebook.

Tabla 3. Número de usuarios seguidores en la red social Facebook.

Cantidad de seguidores de las empresas de telefonía móvil en Facebook (mayo 2018)			
Empresas	Movistar	Claro	CNT Ecuador
Seguidores	1’430.529	1’589.374	572.772

3.1.3 Objetivos de la minería de datos para el análisis de sentimientos

- Clasificar las opiniones de los usuarios seguidores en términos positivos, negativos y neutros en base al servicio que ofrecen las telefonías móviles Movistar, Claro y CNT Ecuador.

- Realizar una comparación de los términos comunes positivos y negativos con mayor frecuencia existente en las empresas telefónicas Movistar, Claro y CNT Ecuador.

3.1.4 Realizar el plan de proyecto

El proyecto se divide en las siguientes etapas para facilitar su organización de realización de este:

- Etapa 1: Análisis de la estructura de los datos y la información de la base de datos.
- Etapa 2: Exploración y calidad de datos.
- Etapa 3: Preparación de los datos (selección, limpieza, construcción, conversión y formateo si el caso es necesario) para facilitar la minería de datos sobre estos.
- Etapa 4: Elección de las técnicas de minería de datos y ejecución sobre los datos.
- Etapa 5: Análisis de los resultados obtenidos en la etapa anterior, en caso de ser necesario se repite la etapa 4.
- Etapa 6: Evaluación de los resultados obtenidos en función de los objetivos del negocio y criterios establecidos.
- Etapa 7: Presentación de resultados.

3.1.5 Evaluación inicial de las herramientas y técnicas

La herramienta que se va a utilizar para llevar a cabo este proyecto de aplicación de técnicas de minería de datos para el análisis de sentimientos es RStudio, como se menciona en el apartado 1.2.8, esta herramienta es potente,

soporta procedimientos y técnicas requeridas para análisis de calidad y al mismo tiempo, pretende ser sencillo e intuitivo como sea posible. Posterior a ello se utiliza la herramienta Facepager como se menciona en el apartado 1.2.7 que permite buscar y extraer datos públicos de las diversas redes sociales, los mismos que se almacenan en una base de datos SQLite y pueden exportarse a csv.

3.2 Comprensión de los datos

En esta fase de la metodología CRISP-DM se realiza la recolección inicial de los datos para poder establecer un primer contacto con el problema; analizar, explorar y medir la calidad de las opiniones públicas, así como identificar las relaciones existentes entre empresas telefónicas y su popularidad en la red social Facebook.

3.2.1 Recolectar los datos iniciales

Los datos utilizados en este proyecto son datos referentes a las opiniones públicas de los usuarios seguidores de las diferentes empresas telefónicas en la red social Facebook, reflejado en el gráfico 11, sobre el servicio que brindan las mismas. Utilizando la herramienta Facepager se recolectan todos los posts (publicaciones) y comentarios de personas que han participado en cada una de las publicaciones realizadas en los perfiles oficiales de Movistar, Claro y CNT Ecuador en la red social en el último periodo enero del 2018 hasta mayo del 2018 con un aproximado de 14.000 opiniones públicas. La herramienta extrae y recolecta la información separando por nodos cada post (publicación) y comentario de la red social diferenciando por nodos y con un ID permitiendo identificar cada uno de ellos como se menciona en el apartado 2.2.2.

Gráfico 11. URL de las páginas oficiales de las diferentes empresas telefónicas del país.

Se ingresa en la página https://findmyfbid.com/ donde se extrae el identificador único de cada página oficial como se muestra en el gráfico 12.

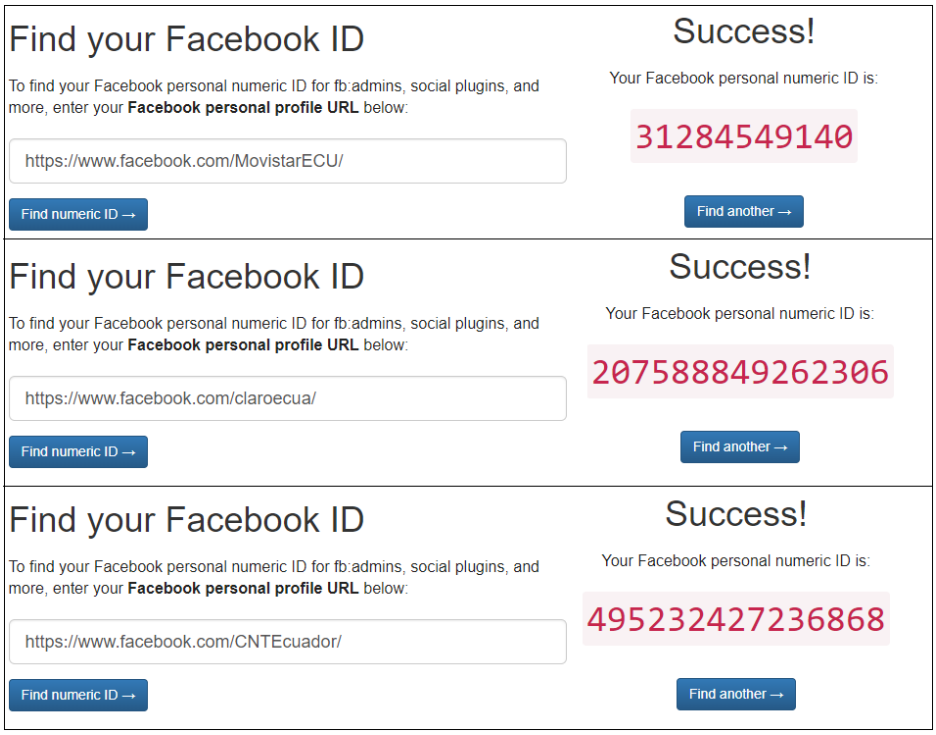


Gráfico 12. Buscador y resultado generado de Facebook ID personal numérico.

Una vez ingresados los ID de cada página oficial de las diferentes telefonías móviles del país, obtenemos los posts (publicaciones) realizados desde enero del 2018 hasta mayo del 2018 como se muestra en el gráfico 13, seguido de esto se procede a obtener los comentarios de cada publicación de las fechas presentadas para la obtención de los datos en nuestro caso de estudio.

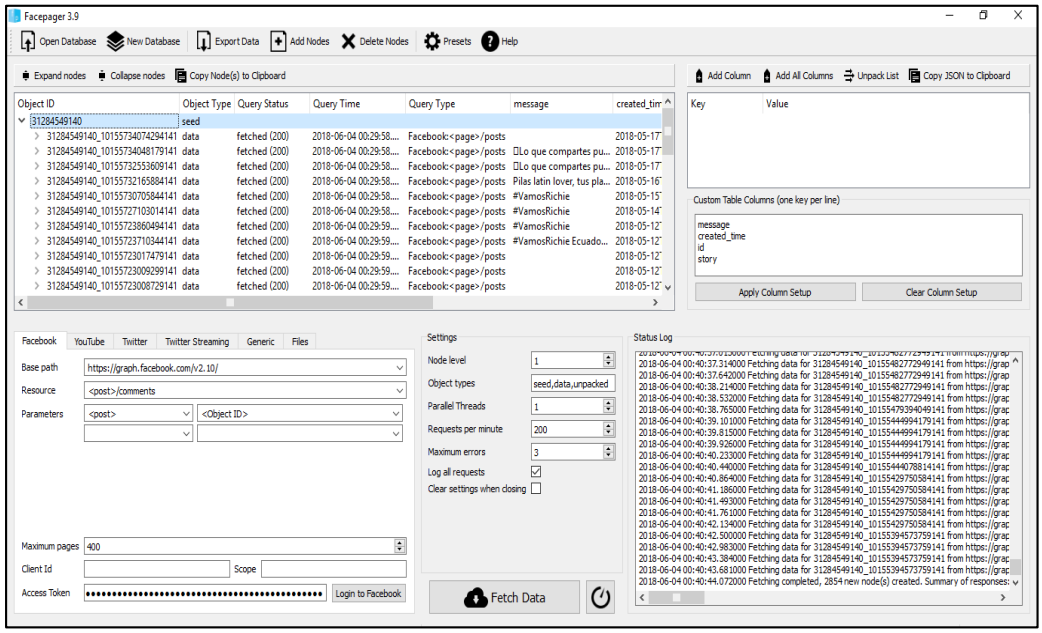


Gráfico 13. Extracción de comentarios en la herramienta Facepager.

3.2.2 Descripción de los datos

Los parámetros acerca de las publicaciones extraídas de las diferentes telefonías móviles del país se muestran en la tabla 4:

Tabla 4. Parámetros de Facepager

Columnas	Descripción
Object ID	Es la llave única que identifica el objeto sea post o comentario
Object type	Tipo de objeto
Query status	Estado del query ejecutado
Query time	Fecha de descarga de los datos
Query type	Tipo de query sea post o comentario, videos, etc.
Message	Mensaje sea publicación o comentario
Created_time	Fecha de creación de la publicación o comentario

Los datos extraídos se encuentran almacenados y separados por empresas telefónicas en archivos planos csv específicamente creadas para este fin generados con la herramienta Facepager como se refleja en el gráfico 14.

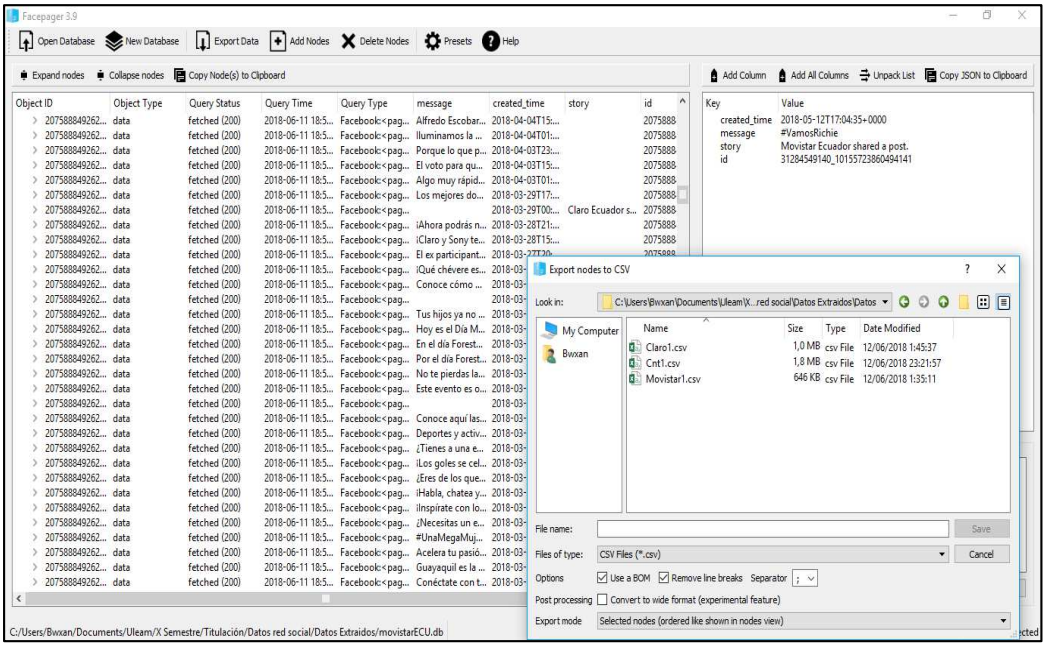


Gráfico 14. Almacenamiento de posts (publicaciones) y comentarios extraídos.

3.2.3 Explotación de los datos

Mediante la herramienta Facepager se pudo obtener los siguientes datos desde enero del 2018 hasta mayo del 2018:

- Publicaciones de las diferentes empresas telefónicas en la red social Facebook (ver tabla 5-7).
- Comentarios de los usuarios seguidores de las páginas oficiales de Movistar, Claro y CNT Ecuador en la red social Facebook (ver tabla 5-7).

Tabulación de la Empresa Movistar

En la tabla 5, se muestran los datos obtenidos de la página oficial de Movistar por medio de la herramienta Facepager, la cual consta con un total de 2762 datos entre publicaciones y comentarios obtenidos en el periodo enero 2018 hasta mayo 2018.

Tabla 5. Datos obtenidos de Facepager empresa Movistar.

Empresa Movistar (enero 2018 – mayo 2018)						
Entradas	Enero	Febrero	Marzo	Abril	Mayo	Total
Publicaciones	4	12	3	7	15	41
Comentarios	305	1478	48	366	524	2721
Total	309	1490	51	373	539	2762

Tabulación de la Empresa Claro

En la tabla 6, se muestran los datos obtenidos de la página oficial de Claro por medio de la herramienta Facepager, la cual consta con un total de 4300 datos entre publicaciones y comentarios obtenidos en el periodo enero 2018 hasta mayo 2018, superando a Movistar con un alto número de interacciones de la empresa con los usuarios seguidores.

Tabla 6. Datos obtenidos de Facepager empresa Claro.

Empresa Claro (enero 2018 – mayo 2018)						
Entradas	Enero	Febrero	Marzo	Abril	Mayo	Total
Publicaciones	93	61	30	47	47	278
Comentarios	669	1353	399	1028	573	4022
Total	762	1414	429	1075	620	4300

- *Tabulación de la Empresa CNT*

En la tabla 7, se muestran los datos obtenidos de la página oficial de CNT por medio de la herramienta Facepager, la cual consta con un total de 8342 datos entre publicaciones y comentarios obtenidos en el periodo enero 2018 hasta mayo 2018, superando a Movistar y Claro con el mayor número de interacciones que posee la empresa con los usuarios seguidores.

Tabla 7. Datos obtenidos de Facepager empresa CNT.

Empresa CNT (enero 2018 – mayo 2018)						
Entradas	Enero	Febrero	Marzo	Abril	Mayo	Total
Publicaciones	75	100	75	109	116	475
Comentarios	1161	1590	1339	2346	1431	7867
Total	1236	1690	1414	2455	1547	8342

3.2.4 Verificar la calidad de los datos

En este apartado se verifica la calidad de los datos obtenidos mediante un análisis e interpretación por cada empresa de telefonía móvil.

- *Análisis e interpretación de la empresa Movistar*

En la tabla 5 y gráfico 15, se refleja el número de datos obtenidos que representan la opinión pública a través de la herramienta Facepager.

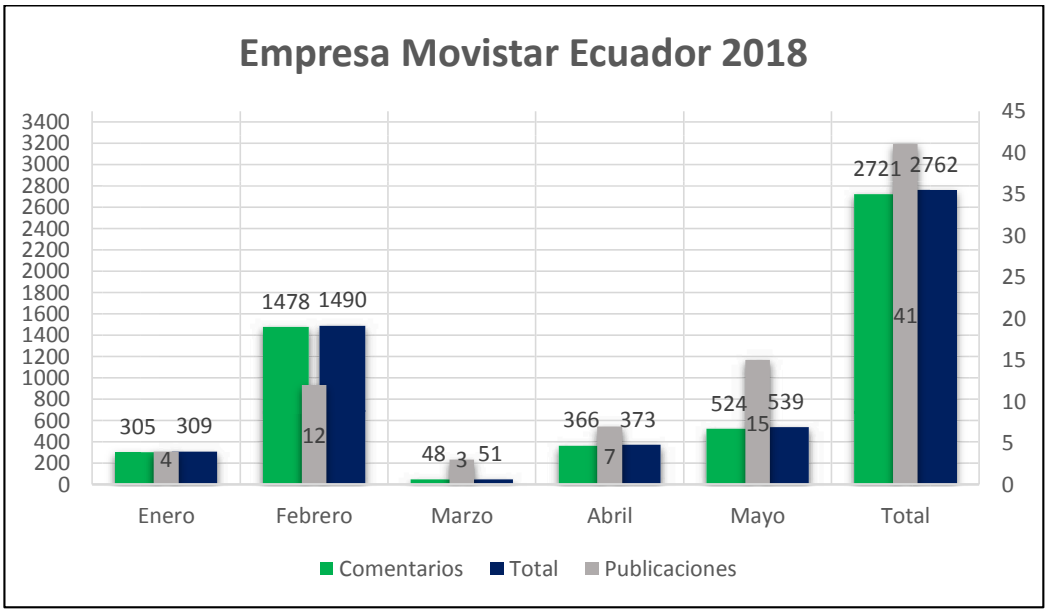


Gráfico 15. Datos generados de Movistar del periodo enero 2018 hasta mayo 2018.

En el cual se destaca lo siguiente:

La empresa movistar cuenta con un total de 41 publicaciones desde enero del 2018 hasta mayo del 2018, en la cual se determina que el mes de marzo tiene el más bajo número de publicaciones realizadas por la empresa con respecto al mes de mayo, destacándose como el mayor en interacción de uso con respecto a publicidad. La interacción de los usuarios hacia la empresa cuenta con un total de 2721 comentarios que a su vez se destaca en el mes de febrero con un total de 1478 comentarios referente al mes de marzo con el menor número de interacción, contando con un total de 2762 datos entre publicaciones por parte de Movistar como de comentarios por parte de los usuarios que interactúan con la empresa.

- *Análisis e interpretación de la empresa Claro*

En la tabla 6 y gráfico 16, se refleja el número de datos obtenidos que representan la opinión pública a través de la herramienta Facepager.

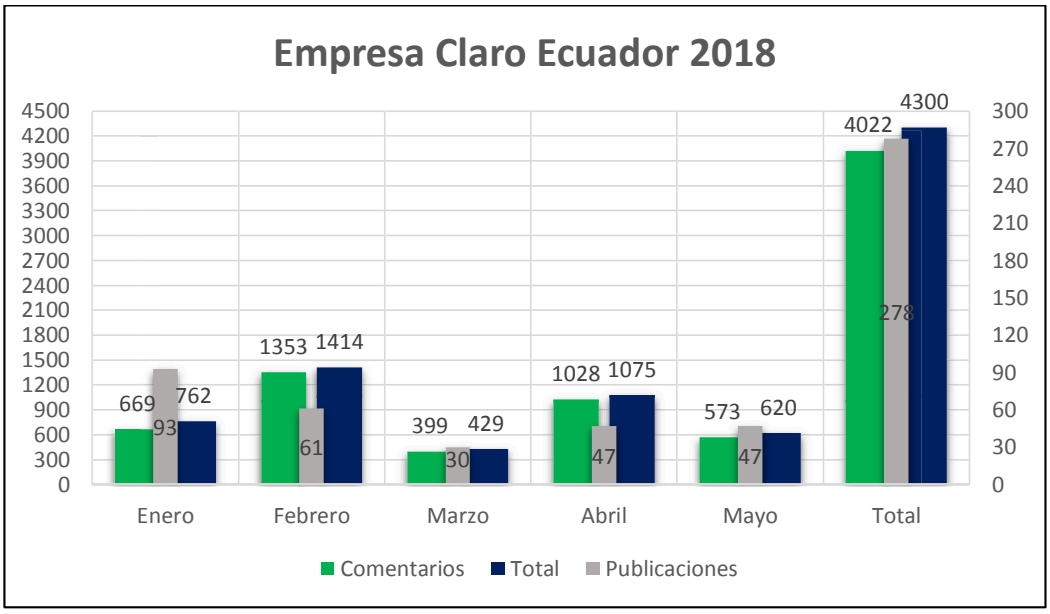


Gráfico 16. Datos generados de Claro del periodo enero 2018 hasta mayo 2018.

En la cual se destaca lo siguiente:

La empresa Claro cuenta con un total de 278 publicaciones desde enero del 2018 hasta mayo del 2018, en la cual se determina que el mes de marzo tiene el más bajo número de publicaciones realizadas por la empresa con respecto al mes de enero, destacándose como el mayor mes en interacción de publicidad. La interacción de los usuarios hacia la empresa cuenta con un total de 4022 comentarios que a su vez se destaca en el mes de febrero con un total de 1353 comentarios referente al mes de marzo con el menor número de comentarios, contando con un total de 4300 datos entre publicaciones por parte de Claro como de comentarios por parte de los usuarios que interactúan con la empresa.

- *Análisis e interpretación de la empresa CNT*

En la tabla 7 y gráfico 17, se refleja el número de datos obtenidos que representan la opinión pública a través de la herramienta Facepager.

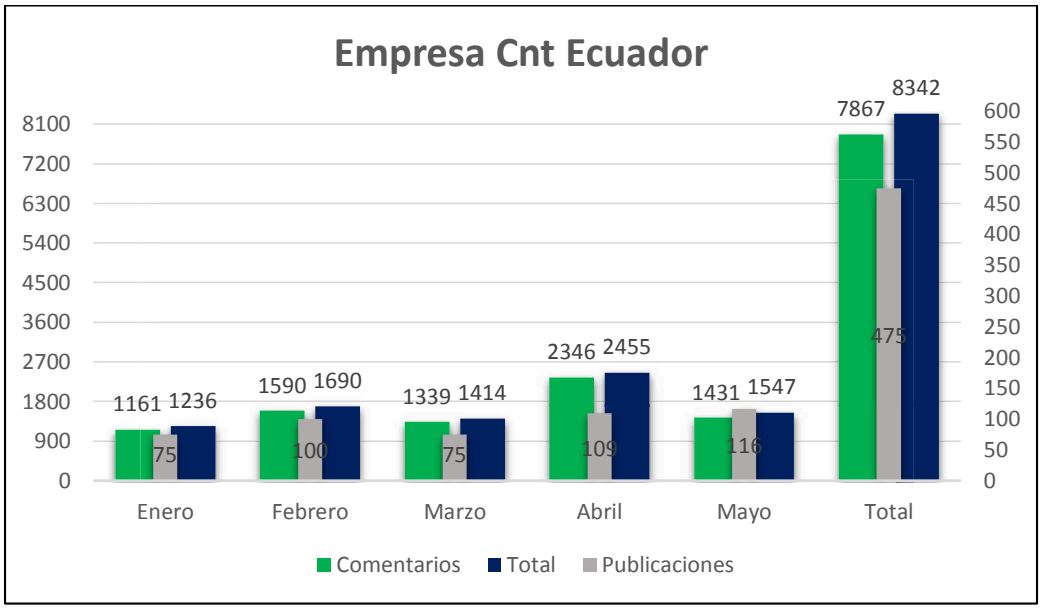


Gráfico 17. Datos generados de CNT del periodo enero 2018 hasta mayo 2018.

En la cual se destaca lo siguiente:

La empresa CNT cuenta con un total de 475 publicaciones desde enero del 2018 hasta mayo del 2018, en la cual se determina que el mes de enero y marzo tienen el más bajo número de publicaciones realizadas por la empresa con respecto al mes de mayo, destacándose como el mayor mes en interacción de publicidad. La interacción de los usuarios hacia la empresa cuenta con un total de 7867 comentarios que a su vez se destaca en el mes de abril con un total de 2346 comentarios referente al mes de enero con el menor número de comentarios, contando con un total de 8342 datos entre publicaciones por parte de CNT como de comentarios por parte de los usuarios que interactúan con la empresa.

3.3 Preparación de los datos

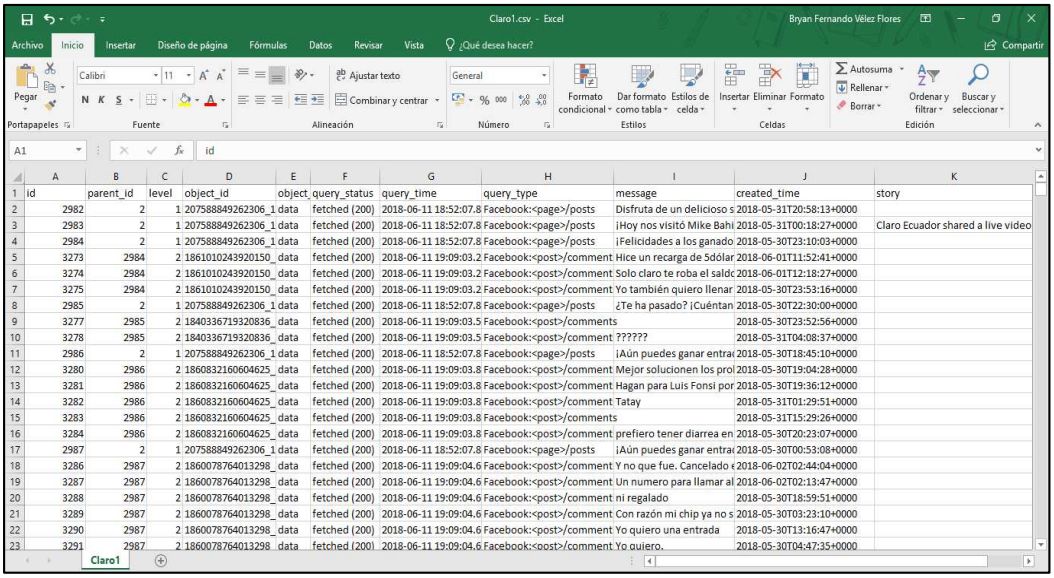
En esta fase de la metodología CRISP-DM, se concentrará en la selección de los datos para el análisis de sentimientos, los cuales serán utilizados en la aplicación de técnicas de minería de datos. Luego se realizará la respectiva limpieza, construcción si es necesario de los datos utilizando la herramienta RStudio, finalmente se integra y se da paso a la fase de modelado en base al estudio realizado sobre el servicio que ofrecen las diferentes empresas telefónicas del país.

3.3.1 Seleccionar los datos

Los registros por usar para el análisis de sentimientos serán los de la base de comentarios de cada empresa telefónica: Movistar, Claro y CNT Ecuador; que se encuentran almacenados en archivos csv por empresa, permitiendo identificar los parámetros de la tabla base como se muestra en la tabla 8, que fue generada de la herramienta Facepager del capítulo anterior y gráfico 18 que contiene los datos extraídos de la herramienta como: posts y comentarios, que están separados por el tipo de parámetros descritos en la tabla 8.

Tabla 8. Descripción de la tabla base

Campo	Tipo	Descripción
id	Numérico	Identificación del campo
id_hijo	Numérico	Identificación segundo nivel
id_objeto	Numérico	Identificación del objeto
nivel	Numérico	Nivel del registro
tipo_objeto	Cadena	Describe si es post o comentario
mensaje	Varchar2	Describe el mensaje del post o comentario
fecha_creación	Date	Fecha de creación de la acción
fecha_descarga	Date	Fecha de descarga de la data



id	parent_id	level	object_id	object_	query_status	query_time	query_type	message	created_time	story
2982	2	1	207588849262306_1	data	fetched (200)	2018-06-11 18:52:07.8	Facebook:<page>/posts	Disfruta de un delicioso s	2018-05-31T20:58:13+0000	
2983	2	1	207588849262306_1	data	fetched (200)	2018-06-11 18:52:07.8	Facebook:<page>/posts	¡Hoy nos visitó Mike Bahi	2018-05-31T00:18:27+0000	Claro Ecuador shared a live video
2984	2	1	207588849262306_1	data	fetched (200)	2018-06-11 18:52:07.8	Facebook:<page>/posts	¡Felicidades a los ganado	2018-05-30T23:10:03+0000	
3273	2984	2	1861010243920150_	data	fetched (200)	2018-06-11 19:09:03.2	Facebook:<post>/comment	Hice un recarga de 5dólar	2018-06-01T11:52:41+0000	
3274	2984	2	1861010243920150_	data	fetched (200)	2018-06-11 19:09:03.2	Facebook:<post>/comment	Solo claro te roba el saldo	2018-06-01T12:18:27+0000	
3275	2984	2	1861010243920150_	data	fetched (200)	2018-06-11 19:09:03.2	Facebook:<post>/comment	Yo también quiero llenar	2018-05-30T23:53:16+0000	
2985	2	1	207588849262306_1	data	fetched (200)	2018-06-11 18:52:07.8	Facebook:<page>/posts	¿Te ha pasado? ¡Cuéntan	2018-05-30T22:30:00+0000	
3277	2985	2	1840336719320836_	data	fetched (200)	2018-06-11 19:09:03.5	Facebook:<post>/comments		2018-05-30T23:52:56+0000	
3278	2985	2	1840336719320836_	data	fetched (200)	2018-06-11 19:09:03.5	Facebook:<post>/comment	??????	2018-05-31T04:08:37+0000	
2986	2	1	207588849262306_1	data	fetched (200)	2018-06-11 18:52:07.8	Facebook:<page>/posts	¡Aún puedes ganar entra	2018-05-30T18:45:10+0000	
3280	2986	2	1860832160604625_	data	fetched (200)	2018-06-11 19:09:03.8	Facebook:<post>/comment	Mejor solucionen los pro	2018-05-30T19:04:28+0000	
3281	2986	2	1860832160604625_	data	fetched (200)	2018-06-11 19:09:03.8	Facebook:<post>/comment	Hagan para Luis Fonsi por	2018-05-30T19:36:12+0000	
3282	2986	2	1860832160604625_	data	fetched (200)	2018-06-11 19:09:03.8	Facebook:<post>/comment	Tatay	2018-05-31T01:29:51+0000	
3283	2986	2	1860832160604625_	data	fetched (200)	2018-06-11 19:09:03.8	Facebook:<post>/comments		2018-05-31T15:29:26+0000	
3284	2986	2	1860832160604625_	data	fetched (200)	2018-06-11 19:09:03.8	Facebook:<post>/comment	prefiero tener diarrea en	2018-05-30T20:23:07+0000	
2987	2	1	207588849262306_1	data	fetched (200)	2018-06-11 18:52:07.8	Facebook:<page>/posts	¡Aún puedes ganar entra	2018-05-30T00:53:08+0000	
3286	2987	2	1860078764013298_	data	fetched (200)	2018-06-11 19:09:04.6	Facebook:<post>/comment	Y no que fue. Cancelado (	2018-06-02T02:44:04+0000	
3287	2987	2	1860078764013298_	data	fetched (200)	2018-06-11 19:09:04.6	Facebook:<post>/comment	Un numero para llamar al	2018-06-02T02:13:47+0000	
3288	2987	2	1860078764013298_	data	fetched (200)	2018-06-11 19:09:04.6	Facebook:<post>/comment	ni regalado	2018-05-30T18:59:51+0000	
3289	2987	2	1860078764013298_	data	fetched (200)	2018-06-11 19:09:04.6	Facebook:<post>/comment	Con razón mi chip ya no s	2018-05-30T03:23:10+0000	
3290	2987	2	1860078764013298_	data	fetched (200)	2018-06-11 19:09:04.6	Facebook:<post>/comment	Yo quiero una entrada	2018-05-30T13:16:47+0000	
3291	2987	2	1860078764013298_	data	fetched (200)	2018-06-11 19:09:04.6	Facebook:<post>/comment	Yo quiero.	2018-05-30T04:47:35+0000	

Gráfico 18. Datos extraídos de la herramienta Facepager de la empresa telefónica Claro.

Cada archivo csv cuenta con datos extraídos de las diferentes empresas telefónicas del país. Como ejemplo se tomó a la empresa Claro, donde se identifican los campos que se obtuvieron de la herramienta Facepager. Cada publicación y comentario realizado en la red social cuenta con un ID personal, donde se diferencia una publicación de un comentario por el tipo de campo, en este caso el nivel (level) de nodo, el cual permite identificar el nivel del registro, siendo 1 un post (publicación) y 2 un comentario, que a su vez el campo tipo de objeto (query_type) determina lo mismo, pero de una manera descriptiva como se detalla la tabla 8 y en las columnas del gráfico 18.

3.3.2 Limpiar los datos

Para realizar el análisis de sentimientos se requiere tener los datos de manera estructurada y limpia, por tanto, se procede a eliminar los registros de las diferentes empresas de telefonía móvil que contengan caracteres especiales, campos vacíos o comentarios sin sentido alguno, para ello se utilizará la herramienta RStudio.

- *Crear new Project*

En RStudio se crea un nuevo proyecto en este caso uno que englobe a las empresas de telefonía móvil: Movistar, Claro y CNT Ecuador como se muestra en el gráfico 19.

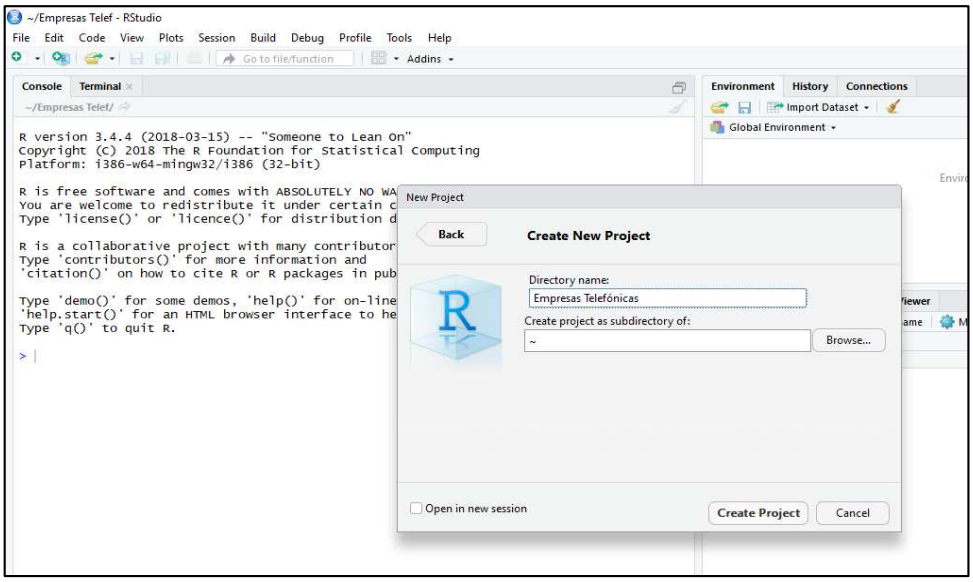


Gráfico 19. Creación de nuevo proyecto en RStudio por cada empresa de telefonía móvil.

- *Importar librería*

Como todo lenguaje, RStudio tiene paquetes de librerías las mismas que permite realizar análisis de datos de acuerdo con nuestro interés en este caso como se muestra en el gráfico 20, las librerías a utilizar para la elaboración del análisis de sentimientos.

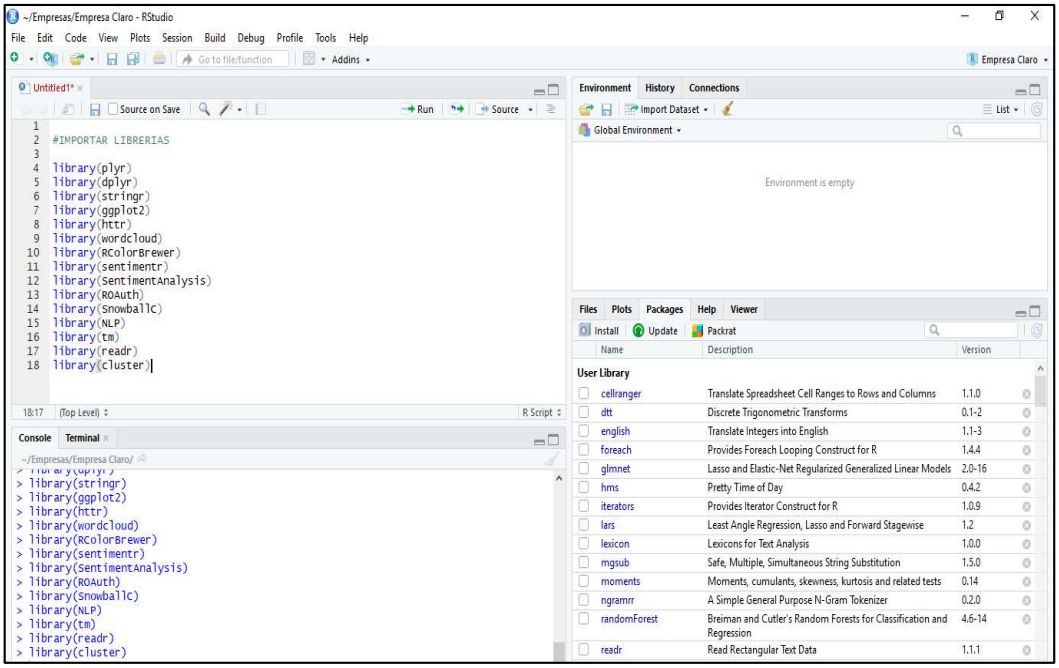


Gráfico 20. Importación de librerías en RStudio.

Importar datos

Para importar datos en RStudio se define un nombre a la variable de importación para que la herramienta pueda importar e interpretar los datos, luego el comando *read* y el tipo de archivo en este caso csv, seguido del nombre del archivo que contiene los posts y comentarios, y la función *sep* que permite separar datos como se muestra en el gráfico 21.

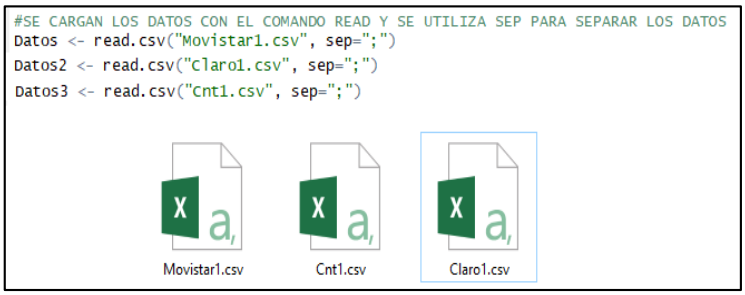


Gráfico 21. Importar datos en RStudio de las empresas de telefonía móvil del país.

- *Preprocesamiento*

Para realizar el preprocesamiento que viene siendo la limpieza de los datos, se crea un corpus el cual contiene los datos de la variable que se va a utilizar, en este caso de la variable *message* que contiene los posts y comentarios de la red social Facebook de las diferentes empresas telefónicas como se muestra en el gráfico 22.

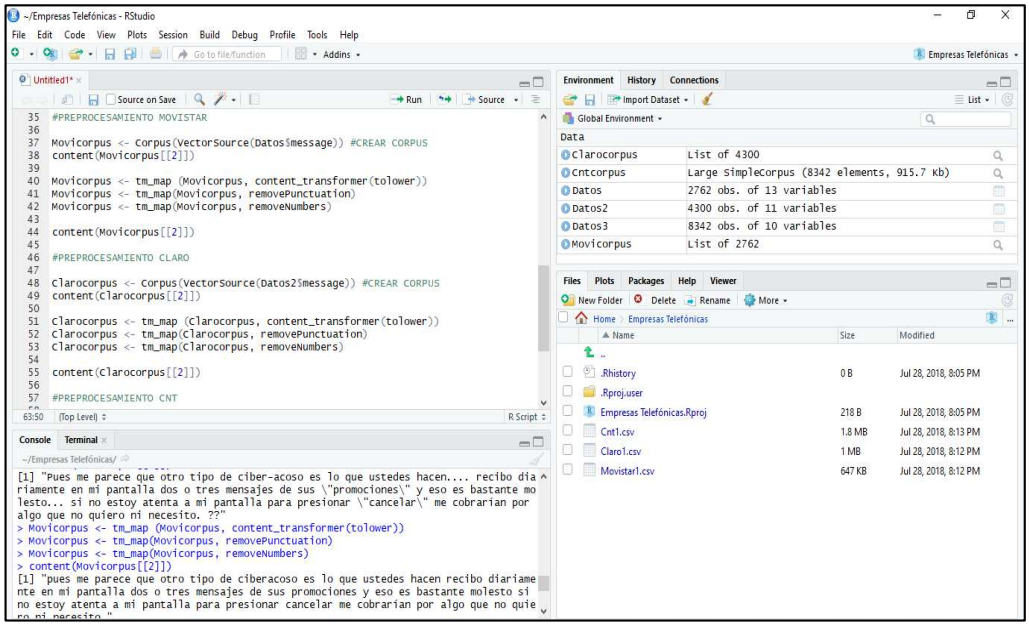


Gráfico 22. Preprocesamiento de los datos en RStudio.

Se utiliza *tm_map* para aplicar funciones de transformación como se muestra en el gráfico 22. La función *content_transformer* transforma las palabras o letras que estén en mayúsculas a minúsculas, *removePunctuation* remueve los signos y caracteres de puntuación del texto y la función *removeRumbers* permite remover los números que estén incluidos en los textos. Al final de cada función o de cada preprocesamiento utilizamos la función *content* seguido del nombre de nuestro corpus para comprobar si se ha realizado o no correctamente la limpieza de acuerdo con cada función anteriormente mencionada.

Dentro del preprocesamiento se utiliza el comando *stopword*, el cual permite suprimir las palabras vacías, refiriéndose a todas aquellas palabras que carecen de un significado por si solas como las preposiciones, conjunciones, pronombres, etc. Utilizando el corpus de cada empresa telefónica como variable por medio de la función *tm_map,* que es una función de transformación, toma un documento de texto (corpus) como entrada y devuelve otro de las mismas características pero con los parámetros establecidos, en este caso se define utilizando la función *removewords* seguido de la concatenación C, permitiendo unir el removedor de palabras con la función *stopword,* seguido del idioma el cual suprimirá las palabras vacías (palabras que carecen de significado), deteniendo todas las palabras en ese idioma, en este caso detendrá y eliminará las palabras vacías que estén en inglés y en español de cada empresa de telefonía móvil como se muestra en el gráfico 23, tomando como prueba a la empresa CNT Ecuador, el cual muestra un mensaje que está contenido en el nodo 2 por medio de la función *content*. El mensaje muestra algunas palabras vacías (palabras que carecen de significado) y palabras que tienen significado propio, los mismos que luego de aplicar la función *removewords* concatenado con la función *stopword* los elimina, dejando solo las palabras que contengan significado y reduciendo el mensaje con un margen de palabras claves de su estado original.

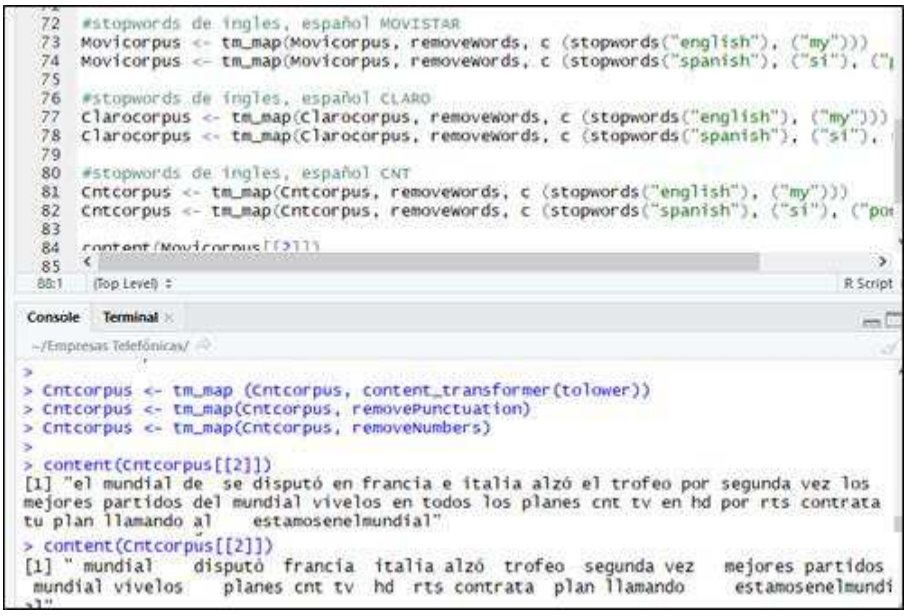


Gráfico 23. Removedor de palabras vacías y pronombres en idiomas inglés y español.

Luego de transformar el texto en minúscula, remover los signos de puntuación, remover números y de eliminar las palabras vacías de los posts y comentarios emitidos por las empresas de telefonía móvil y de los usuarios seguidores de la red social Facebook, se procede a remover las URL de cada texto, obviando por ser direcciones web que no sirven para el caso de estudio. En el gráfico 24 se muestra como la función *removeURL* permite remover los enlaces que se emiten en los textos de las empresas de telefonía móvil de la red social Facebook, luego se utiliza la función *content* para comprobar si ha funcionado el removedor de *URLs* aplicado para cada empresa telefónica.

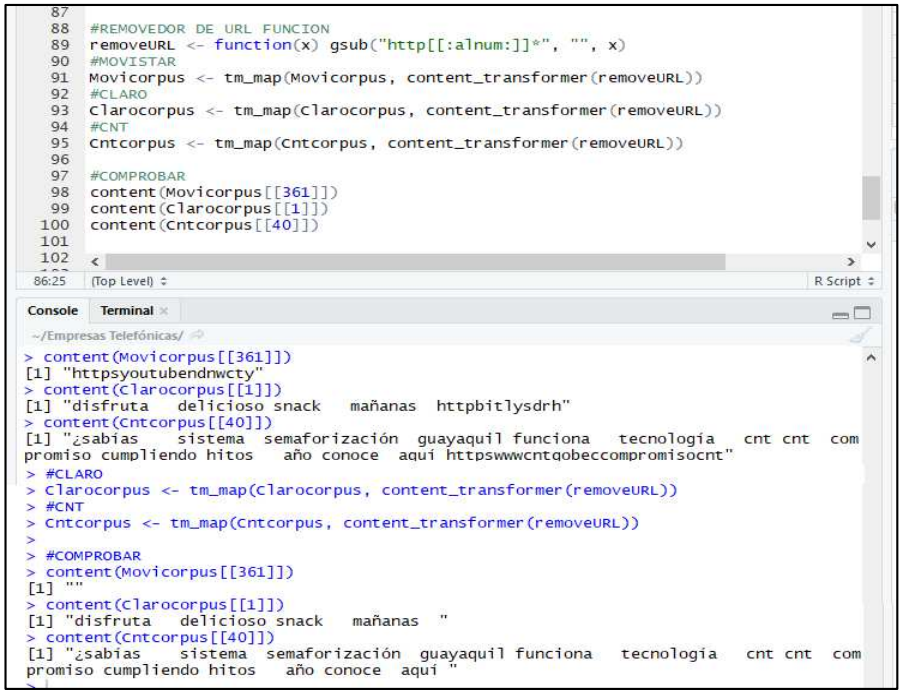


Gráfico 24. Removedor de enlaces y direcciones URLs.

Dentro del contexto del análisis de sentimientos se encuentra el código **ASCII**, siendo un código estándar americano para el intercambio de información, que define los caracteres que se utilizan en el ordenador. Como existen múltiples plataformas hardware en el mercado e infinidad de sistemas operativos se han ajustado los códigos utilizados por el ordenador atendiendo a estas razones y al idioma de los usuarios. En la red social Facebook y específicamente en el formato que emiten los posts y comentarios de dicha red social, cuenta con este código que a su vez es difícil de entender o de hacer entender al ordenador, para esto se utiliza una función que invierte caracteres de código **ASCII** a **UTF-8** para que la herramienta RStudio pueda leer los caracteres que estén en dicho formato y sea entendible para el estudio que se realiza. En el gráfico 25, se muestra la función que invierte la traducción de cada corpus de las empresas de telefonía móvil. Al momento de ejecutar esta función lo que hace con el texto es invertir los caracteres que no se podían leer, adaptándose al lenguaje de la herramienta. Seguido de esto

se vuelve a ejecutar la función *removePunctuation* para que RStudio pueda eliminar los caracteres que no eran entendibles para el ordenador.

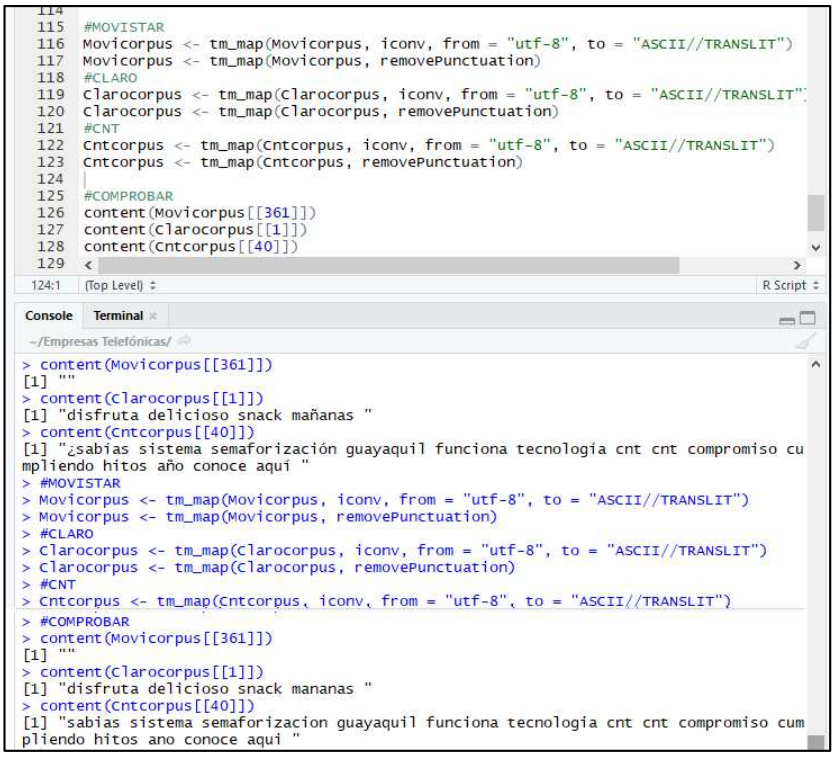


Gráfico 25. Función para la traducción de código ASCII.

El gráfico 26, muestra un resumen sobre la fase de preprocesamiento, el cual se divide por funciones, separando la transformación de datos de los caracteres removidos. Siguiendo el proceso de los scripts para la limpieza del texto, y por último la transformación del código **ASCII** a **UTF-8**, luego repite el proceso de remover signo de puntuación y culmina con la salida del texto limpio de cada empresa de telefonía móvil para dar paso a la siguiente fase de clasificación.

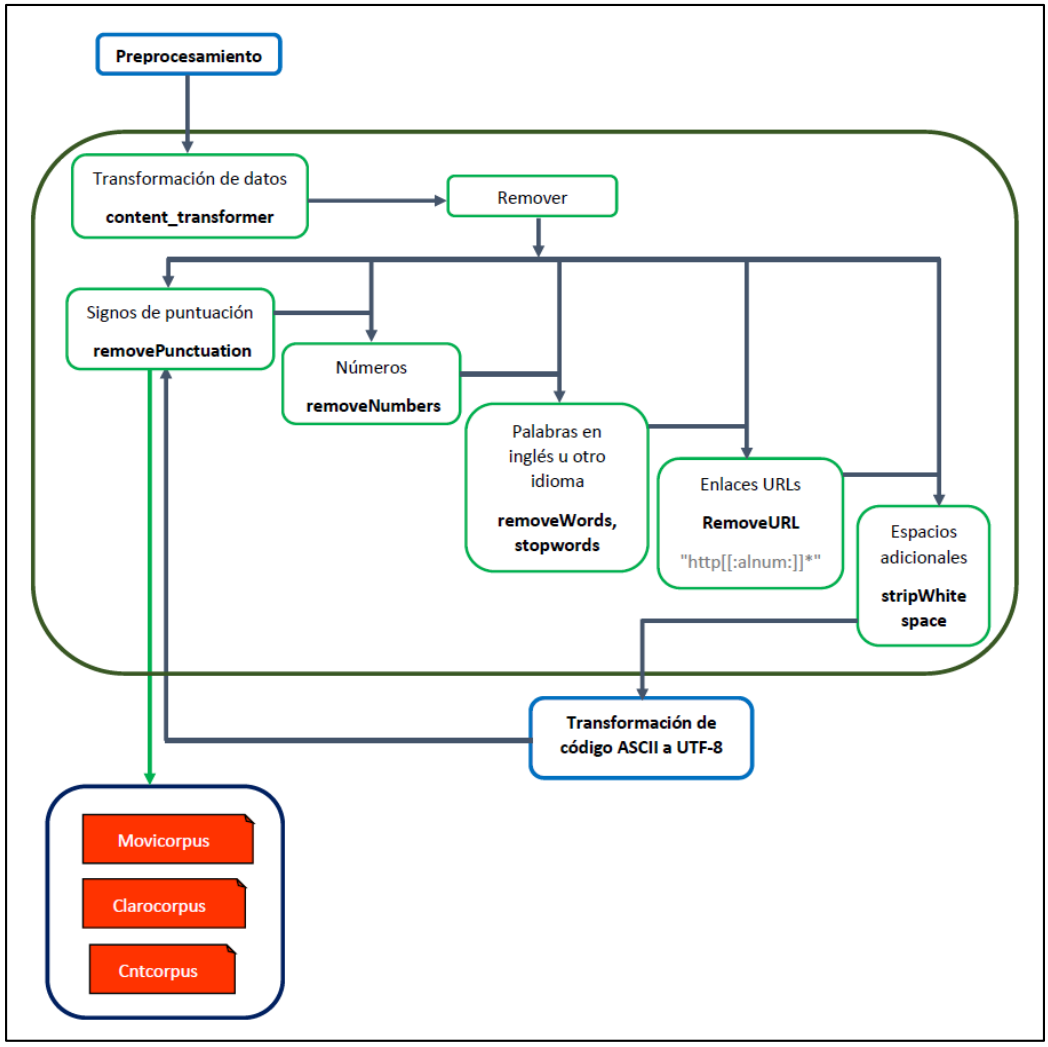


Gráfico 26. Resumen de las funciones del preprocesamiento de texto.

3.3.3 Estructurar los datos

Para la estructuración de datos se aplicará el uso de matrices de términos, siendo la base para la realización de tareas de data mining. La matriz de términos permite la visualización de histogramas de frecuencias, ranking de términos más frecuentes en el caso del servicio de telefonía móvil, los términos que más se utilizan en base al servicio que ofrecen y en respuesta que emiten los usuarios seguidores de: Movistar, Claro y CNT Ecuador.

- *Matriz de documento de términos*

Se crean matrices de documentos de términos para que el computador pueda interpretar y determinar que palabras son las más frecuentes ya que solo entiende números y en este caso interpretará las palabras de cada publicación y comentario de la red social Facebook sobre el servicio de telefonía móvil.

Para ello se crea una matriz de frecuencia por cada empresa de telefonía móvil haciendo referencia a: Movistar, Claro y CNT Ecuador, aplicando con la función *DocumentTermMatrix* como se muestra en el gráfico 27, luego se ejecuta las frecuencias de cada empresa para inspeccionar los elementos que brinda información de cada matriz de documento; el número que tienen las matrices, numero de términos, el largo máximo de columnas, entre otras características.

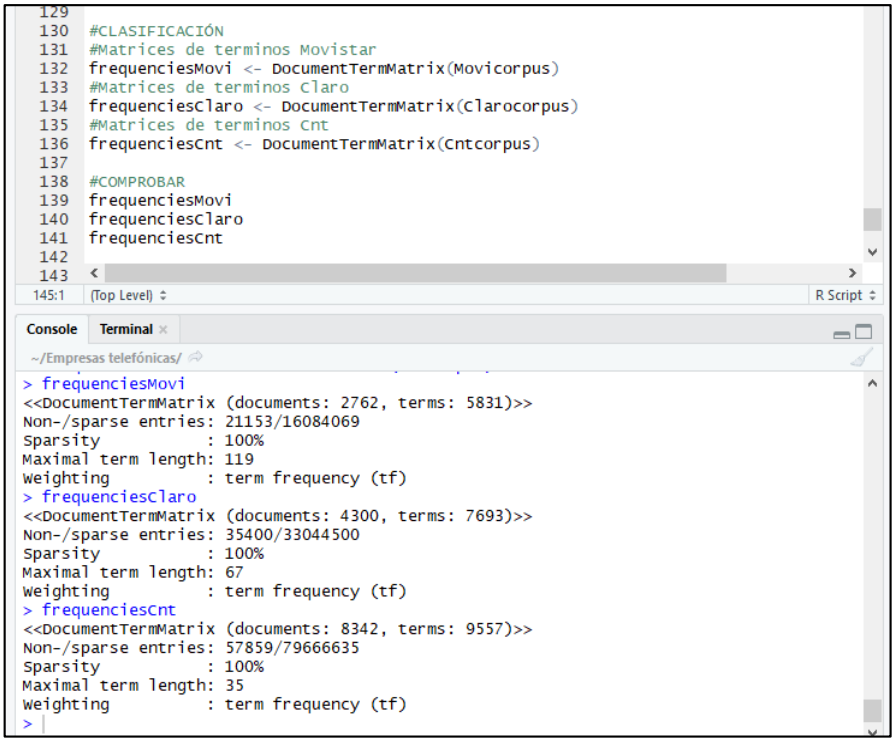


Gráfico 27. Creación de matrices de términos de documentos.

En las siguientes tablas se mostrará un rango de 6 filas que son los documentos que contienen los posts y comentarios, y 6 columnas que representan

la frecuencia de algunas palabras utilizadas en los comentarios de cada empresa de telefonía móvil, esta frecuencia indica el número de veces que se repite la palabra en cada documento, utilizando la función *inspect(frequencies[90:95, 335:340])* en el que se define un rango para las matrices de: Movistar, Claro y CNT Ecuador.

Tabla 9. Matriz de términos frecuentes Movistar

Docs	quemando	rapido	siguen	sms	trabajo	ven
90	0	0	0	0	0	0
91	0	0	0	0	0	0
92	0	0	0	0	0	0
93	0	0	0	0	0	0
94	0	0	0	0	0	0
95	0	0	1	0	0	0

La tabla 9, muestra una previa de lo que es la matriz de la empresa de telefonía móvil Movistar Ecuador, en el cual muestra un rango determinado de las palabras más frecuentes y el número de veces que se repite en un documento definido previamente en el rango anteriormente mencionado.

Tabla 10. Matriz de términos frecuentes Claro

Docs	campeonato	cuando	nacional	satelital	transmision	vivo
90	0	0	0	0	0	0
91	0	0	0	0	0	0
92	1	0	1	1	0	0
93	0	0	0	0	0	0
94	1	0	1	0	0	0
95	1	0	0	0	0	0

La tabla 10, muestra una previa de lo que es la matriz de la empresa de telefonía móvil Claro Ecuador, en el cual muestra un rango determinado de las

palabras más frecuentes y el número de veces que se repite en un documento definido previamente en el rango anteriormente mencionado. En comparación con la matriz de la empresa Movistar Ecuador en el rango determinado Claro muestra dominio sobre palabras y el número de veces que se repiten.

Tabla 11. Matriz de términos frecuentes CNT

Docs	denuncia	dicha	inclui	institucion	misiva	tarifa
90	0	0	0	0	0	0
91	0	0	0	0	0	0
92	0	0	0	0	0	0
93	0	0	0	0	0	0
94	0	0	0	0	0	0
95	0	0	0	0	0	0

Por otro lado, en la tabla 11 de la matriz de la empresa CNT Ecuador, no muestra ninguna frecuencia en comparación de las matrices de las empresas anteriormente mencionadas. Esto se debe a que el rango definido solo es una inspección de lo realizado para comprobar cómo funciona una matriz de frecuencia y dependiendo del rango puede mostrar más o menos el nivel de frecuencia dependiendo las veces que se repitan las palabras en cada documento de post y comentario.

Se considera una matriz dispersa a una matriz que tenga más de un 60% de ceros, en este caso se procede a reducir las palabras que se repiten muy poco o son poco frecuentes, utilizando la función *removeSparseTerms* y tomando como primer argumento las matrices y como segundo argumento la dispersión como se muestra en el gráfico 28, teniendo en cuenta que la dispersión no puede tomar valores de 0 o 1.0 solo valores intermedios. En este caso se define un 0.999 como

argumento, permitiendo eliminar solo los términos que son muy poco mencionados.

```
171  #Reducir las palabras que se repiten muy poco o son poco frecuentes
172  sparseMovi <- removeSparseTerms(frequenciesMovi, 0.999)
173  sparseClaro <- removeSparseTerms(frequenciesClaro, 0.999)
174  sparseCnt <- removeSparseTerms(frequenciesCnt, 0.999)
175
176  #Revisar los sparse de cada empresa
177  sparseMovi
178  sparseClaro
179  sparseCnt
180
181
182
168:52  FORMA A

Console  Terminal

~/Empresas Telefónicas/

> sparseMovi
<<DocumentTermMatrix (documents: 2762, terms: 1416)>>
Non-/sparse entries: 15960/3895032
Sparsity           : 100%
Maximal term length: 17
Weighting          : term frequency (tf)
> sparseClaro
<<DocumentTermMatrix (documents: 4300, terms: 1314)>>
Non-/sparse entries: 25960/5624240
Sparsity           : 100%
Maximal term length: 23
Weighting          : term frequency (tf)
> sparseCnt
<<DocumentTermMatrix (documents: 8342, terms: 980)>>
Non-/sparse entries: 42193/8132967
Sparsity           : 99%
Maximal term length: 19
Weighting          : term frequency (tf)
>
```

Gráfico 28. Reducción de términos pocos frecuentes.

Luego se procede a crear un nuevo data frame con los documentos y términos más frecuentes. Utilizando la función *colSums* seguido del data frame como argumento para sumar el número de veces que se repite cada termino en los documentos, luego se ordena de manera decreciente permitiendo identificar la frecuencia de las diez primeras palabras de cada empresa de telefonía móvil como se muestra en el gráfico 29.

```
> frequenciesMovi[1:10]
    plan movistar servicio     dias      mas    puedo   quiero     solo  gracias     hola
     410      403      291      129      123      121      119      117      115      106
> frequenciesClaro[1:10]
servicio    claro internet     plan     solo   pesimo    senal  ecuador     peor    mejor
     705      689      308      307      296      198      192      191      172      168
> frequenciesCnt[1:10]
     cnt internet servicio     plan   quiero    senal   pesimo     dias  gracias     solo
    1748     1183     1170      728      389      384      377      354      339      337
>
```

Gráfico 29. Diez palabras más frecuentes por empresa.

Como se muestra en el gráfico 29, se tienen los términos y frecuencias de manera agrupada, para ello se procede a separar las palabras de la frecuencia por cada empresa convirtiendo en una tabla de dos columnas como un nuevo *data.frame* el cual contiene *(word=names(),freq=())*, agregando la variable que contiene las matrices con mayor frecuencia, esto permitirá separar las palabras de las frecuencias en dos columnas.

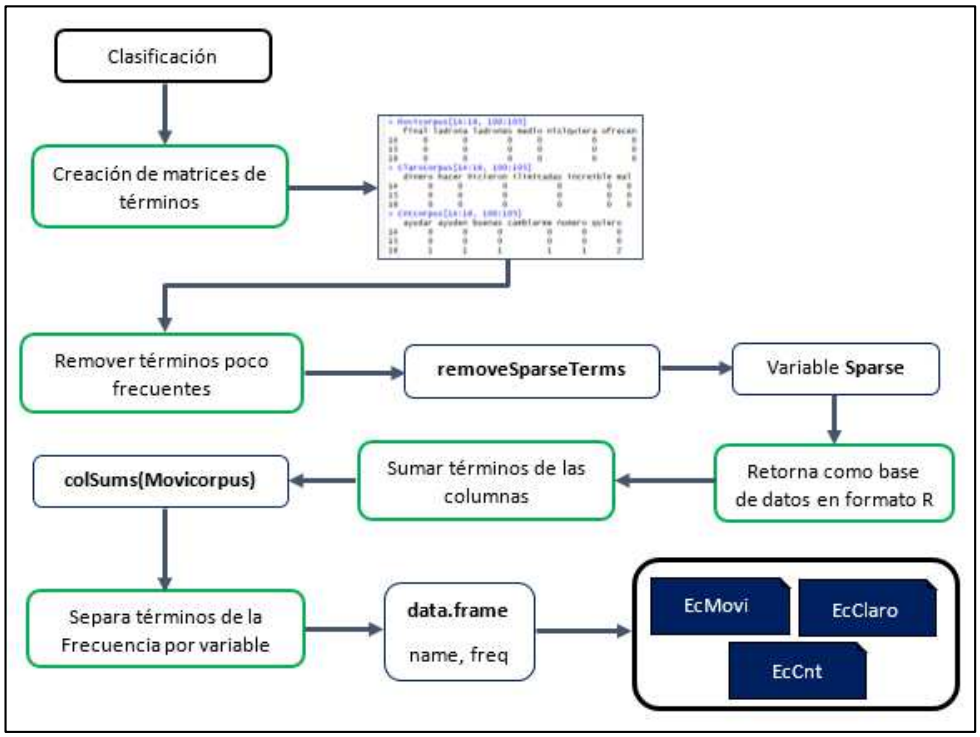


Gráfico 30. Resumen de las funciones de la clasificación de términos.

El gráfico 30, muestra un resumen sobre la fase de clasificación dividido por funciones, siguiendo una secuencia ordenada de los subprocesos, el cual empieza creando la matriz de documentos y eliminando los términos poco frecuentes, luego suma las veces que se repiten los términos en cada documento mostrando como salida los términos más importantes de cada empresa en un nuevo data frame.

En la tabla 12, se muestra cómo queda cada empresa y sus variables: word (palabras) y freq (frecuencia), separadas en diferentes columnas. Cabe recalcar que cada empresa tiene un número diferente de entradas y a su vez con términos similares entre sí, pero de frecuencias diferentes.

Tabla 12. Referencia de las tablas de Movistar, Claro y CNT Ecuador.

EcMovi		**EcClaro**		**EcCnt**	
Word	**Freq**	**Word**	**Freq**	**Word**	**Freq**
plan	410	servicio	705	cnt	1748
movistar	403	claro	689	internet	1183
servicio	291	internet	308	servicio	1170
dias	129	plan	307	plan	728
puedo	121	solo	296	quiero	389
quiero	119	senal	192	señal	384
solo	117	ecuador	191	dias	354
gracias	115	peor	172	gracias	339
hola	106	mejor	168	buenas	337
numero	104	dia	163	solo	337
buenas	103	cliente	155	hola	326
equipo	99	quiero	147	puedo	322
celular	97	ahora	142	barcelona	313
cliente	97	días	142	favor	308
favor	94	megas	142	ver	285

3.3.4 Integrar datos

Para la integración de datos se procede a importar el diccionario de datos el cual servirá para identificar las palabras cuyo sentimiento sea: positivo, negativo o neutro y que corresponda de manera correcta la integración del sentimiento a las palabras de cada empresa de telefonía móvil para evaluar la emoción del texto.

El paquete *tidytext* contiene varios léxicos de sentimiento como un conjunto de datos divididos por sentimiento.

Los tres léxicos de propósito general son:

- *Afinn* de Finn Arup Nielsen
- *Bing* de Bing Liu y sus colaboradores
- *Nrc* de Saif Mohammad y Peter Turney

Los tres de estos léxicos se basan en *unigrams*, es decir, palabras sueltas. Estos léxicos contienen muchas palabras en inglés y las palabras tienen puntajes asignados para sentimientos positivos, negativos y posiblemente también emociones como alegría, ira, tristeza, etc.

En este caso se utiliza el léxico *Bing* el cual clasifica las palabras de forma binaria en categorías positivas y negativas. Las palabras neutras no son consideradas en el léxico por lo cual se deben agregar y se asignar de manera manual. El léxico *Bing* es un diccionario cuyas palabras y sentimiento están en ingles por lo que se procede a exportar para su correcta traducción y luego se vuelve a importar, tal y como se refleja en el gráfico 31.

```
###################DICCIONARIO DE DATOS#####################################

######IMPORTAR DICCIONARIO DE R LEXICON BING PARA POSITIVO Y NEGATIVO
#NOMBRE DE VARIABLE#####
sentimental <- get_sentiments(lexicon = "bing")
head(sentimental)
##COMO EL DICCIONARIO ESTA EN INGLES SE PROCEDE A EXPORTAR PARA
##TRADUCIR A ESPAÑOL##
write.csv(sentimientos, file="sentimiento.csv")
##DESPUES SE VUELVE A IMPORTAR EL DICCIONARIO YA TRADUCCIDO
##Y CON PALABRAS AGREGADAS MANUALMENTE##
espanish <- read.csv("diccionarioespañol.csv", sep=";")
#############################################################################
```

Gráfico 31. Diccionario de datos.

Luego de importar el diccionario de datos con el nombre de *diccionarioespañol.csv* se procede a aplicar la función *merge*, el cual me permite hacer un cruce de palabras entre los datos originales y el diccionario de datos. Lo que se realiza tal y como se refleja en el gráfico 32, es crear una nueva tabla

juntando filas de otras columnas, pero de las palabras que se encuentren en el diccionario de datos y las que están en los datos originales de las empresas de telefonía móvil, antes de ello se procede a agregar palabras que no se encuentren en el diccionario de datos de manera manual y para la cual sean reconocidas al momento de aplicar la función *merge*.

```
#La función merge()
#consigue que se muestren todos los datos de ambos dataframes,
#o solo aquellos que son comunes a ambos
EcMovi = merge(EcMovi, espanish)
EcClaro = merge(EcClaro, espanish)
EcCnt = merge(EcCnt, espanish)

#CAMBIAR NOMBRE DE VARIABLES
#############MOVISTAR#############
colnames (EcMovi) [colnames (EcMovi) == "sentiment"] <- "sentimentMovi"
colnames (EcMovi) [colnames (EcMovi) == "freq"] <- "freqMovi"
#############CLARO#############
colnames (EcClaro) [colnames (EcClaro) == "sentiment"] <- "sentimentClaro"
colnames (EcClaro) [colnames (EcClaro) == "freq"] <- "freqClaro"
#############CNT#############
colnames (EcCnt) [colnames (EcCnt) == "sentiment"] <- "sentimentCnt"
colnames (EcCnt) [colnames (EcCnt) == "freq"] <- "freqCnt"
```

Gráfico 32. Cruce de tablas con la función merge y renombre de variables.

Seguido de ello cambiamos el nombre de las columnas de frecuencia y sentimiento de cada empresa de telefonía móvil, esto de modo que al momento de realizar un nuevo cruce con la función *merge* no se vean afectado estas columnas, por último, se realiza un reemplazo de caracteres a los espacios en blanco agregando la palabra 'neutro' a los espacios en blanco de la columna de sentimientos como se muestra en la tabla 13.

Tabla 13. Referencia de las tablas con la columna sentimiento por empresa.

EcMovi		EcClaro		EcCnt	
Word	**Sentimiento Movistar**	**Word**	**Sentimiento Claro**	**Word**	**Sentimiento Cnt**
plan	Positivo	servicio	Positivo	cnt	Neutro
movistar	Neutro	claro	Positivo	internet	Positivo
servicio	Positivo	internet	Positivo	servicio	Positivo
dias	Neutro	plan	Positivo	plan	Positivo
puedo	Positivo	solo	Neutro	quiero	Positivo
quiero	Positivo	senal	Neutro	senal	Neutro
solo	Neutro	ecuador	Neutro	dias	Neutro
gracias	Positivo	peor	Negativo	gracias	Positivo
hola	Positivo	mejor	Positivo	buenas	Positivo
numero	Neutro	dia	Neutro	solo	Neutro
buenas	Positivo	cliente	Neutro	hola	Positivo
equipo	Neutro	quiero	Positivo	puedo	Positivo
celular	Neutro	ahora	Positivo	barcelona	Neutro
cliente	Neutro	dias	Neutro	favor	Positivo
favor	Positivo	megas	Positivo	ver	Positivo

3.4 Modelado

En esta fase de la metodología CRISP-DM, se evaluará las técnicas de minería de datos más apropiadas para el procesamiento de texto en base al estudio realizado sobre el servicio que ofrecen las diferentes empresas de telefonía móvil: Movistar, Claro y CNT Ecuador, se establecerá si el sentimiento es positivo, negativo o neutro en base al diccionario de datos, luego proceder a generar los modelos y evaluar los resultados para determinar qué empresa de telefonía móvil brinda un mejor servicio en base a los criterios de las experiencias de los usuarios seguidores de la red social Facebook.

3.4.1 Selección de la técnica de modelado

Para la selección de las técnicas que se van a utilizar se hizo acorde al tipo de datos que se obtuvieron con la fase anterior, para cumplir con el objetivo principal que es determinar qué empresa de telefonía móvil de las anteriormente mencionadas tiene mejor servicio en base a las experiencias de los usuarios

seguidores a través de red social Facebook. Entre las técnicas seleccionadas se tiene las siguientes: gráficos de barras, nubes de palabras y pirámides.

3.4.2 Construcción del modelo

- *Gráficos de Barras*

Un gráfico de barras es una forma de resumir un conjunto de datos por categorías. Muestra los datos usando varias barras de la misma anchura, cada una de las cuales representa una categoría concreta. La altura de cada barra es proporcional a una agregación específica y de este modo obtener una primera vista general de los posts y comentarios extraídos de la red social Facebook.

En el gráfico 33, se explica la función para realizar los gráficos de barras con respecto a la frecuencia de términos. Para ello se utiliza un paquete de visualización llamada *ggplot*, que permite crear gráficos que representan datos numéricos y categóricos tanto univariados como multivariantes de una manera directa. El agrupamiento se puede representar por color, símbolo, tamaño y transparencia, como parámetro principal está el nombre de la tabla de datos en este caso el de las empresas de telefonía móvil, combinado con la función *aes* que permite construir mapeos estéticos conjunto con *ggplot* y dentro de la función *aes* se agrega parámetros x que será la variable sentimiento y *fill* (llenar) que será la lista de sentimientos existentes. Se utiliza el símbolo '+' para concatenar una función con otra, luego se utiliza *geom_histogram* que muestra el recuento con barras y como parámetro el *fill* de la columna sentimiento, y *stat 'count'* como conteo del nivel de las veces que se repite cada sentimiento, luego simplemente se agrega título y nombre a los ejes tanto para *x* como para *y*, seguido de una

breve descripción por cada gráfico. Cabe recalcar que la función es la misma para cada empresa, solo se modificaron las variables.

```
#######################GRÁFICO DE BARRAS############################

#GRÁFICO DE BARRAS MOVISTAR
ggplot(EcMovi, aes(x = sentimentMovi, fill = Etiquetas)) +
  geom_histogram(aes(fill = sentimentMovi),stat = "count") +
  xlab("Sentimiento") + ylab("Cantidad") +
  labs(title = "Gráfico de barras Movistar",
       subtitle = "Cantidad de términos divididos por sentimiento",
       caption = "Datos obtenidos de Facebook")

table(EcMovi$sentimentMovi)
########################################
#GRÁFICO DE BARRAS CLARO
ggplot(EcClaro, aes(x = sentimentClaro, fill = Etiquetas)) +
  geom_histogram(aes(fill = sentimentClaro),stat = "count") +
  xlab("Sentimiento") + ylab("Cantidad") +
  labs(title = "Gráfico de barras Claro",
       subtitle = "Cantidad de términos divididos por sentimiento",
       caption = "Datos obtenidos de Facebook")

table(EcClaro$sentimentClaro)
########################################
#GRÁFICO DE BARRAS CNT
ggplot(EcCnt, aes(x = sentimentCnt, fill = Etiquetas)) +
  geom_histogram(aes(fill = sentimentCnt),stat = "count") +
  xlab("Sentimiento") + ylab("Cantidad") +
  labs(title = "Gráfico de barras CNT Ecuador",
       subtitle = "Cantidad de términos divididos por sentimiento",
       caption = "Datos obtenidos de Facebook")

table(EcCnt$sentimentCnt)
```

Gráfico 33. Función de frecuencia general por empresa.

En el gráfico 34, se observan 3 barras de diferentes colores etiquetados, el eje *x* representa el sentimiento: positivo, negativo y neutro, mientras que el eje *y* representa la cantidad de términos por el sentimiento. De forma generalizada las tres barras representan a todos los datos obtenidos por parte de los usuarios seguidores en la red social Facebook sobre el servicio de telefonía móvil de la empresa Movistar, a simple se determina que la empresa consta con 319 términos positivos, 134 términos negativos y 145 términos neutros, con un total de 598 términos en forma generalizada.

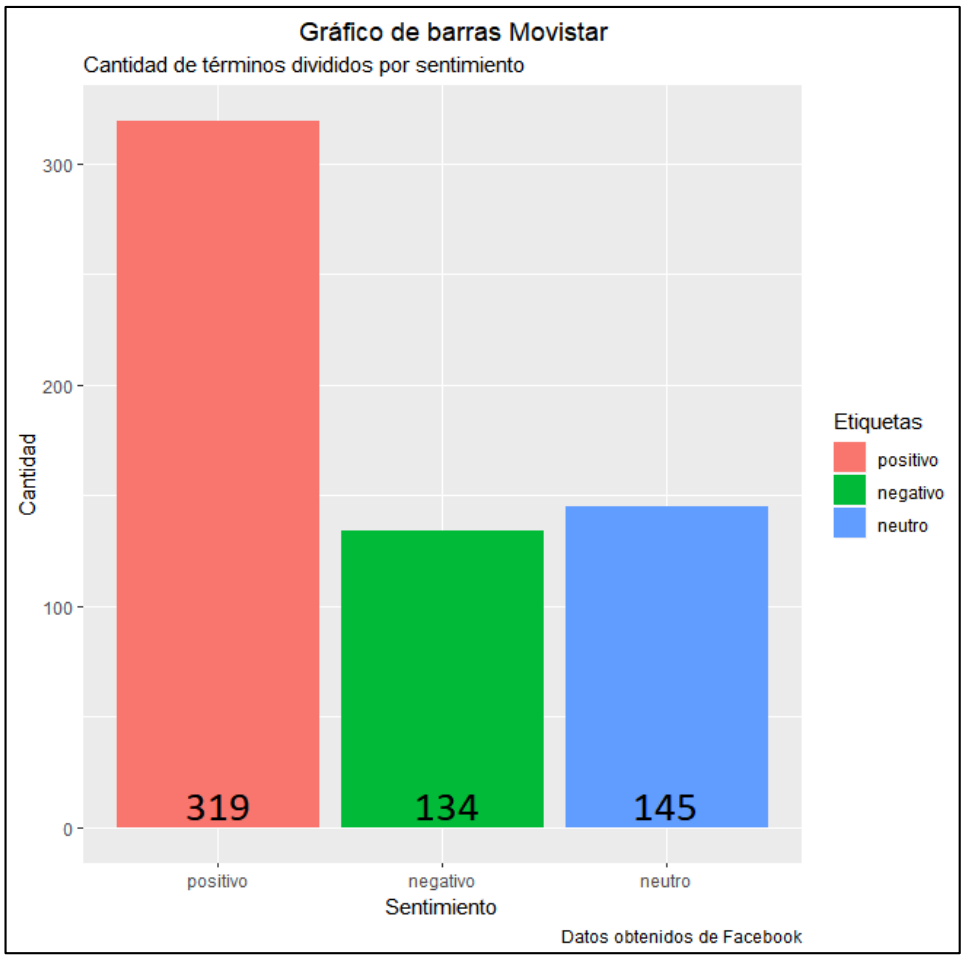


Gráfico 34. Gráfico de barras Movistar.

En el gráfico de barras de la empresa Claro como se muestra en el gráfico 35, a diferencia del gráfico 34 de la empresa Movistar, se puede observar a simple vista una diferencia en la poca participacion de los usuarios seguidores de movistar con respecto a los de Claro en la red social facebook, con una diferencia de 152 términos, incluyendo en su mayoria a términos positivos con una diferencia de 76 términos, de igual manera cuenta un total de 750 términos, entre ellos 395 positivos, 137 negativos y 218 neutros.

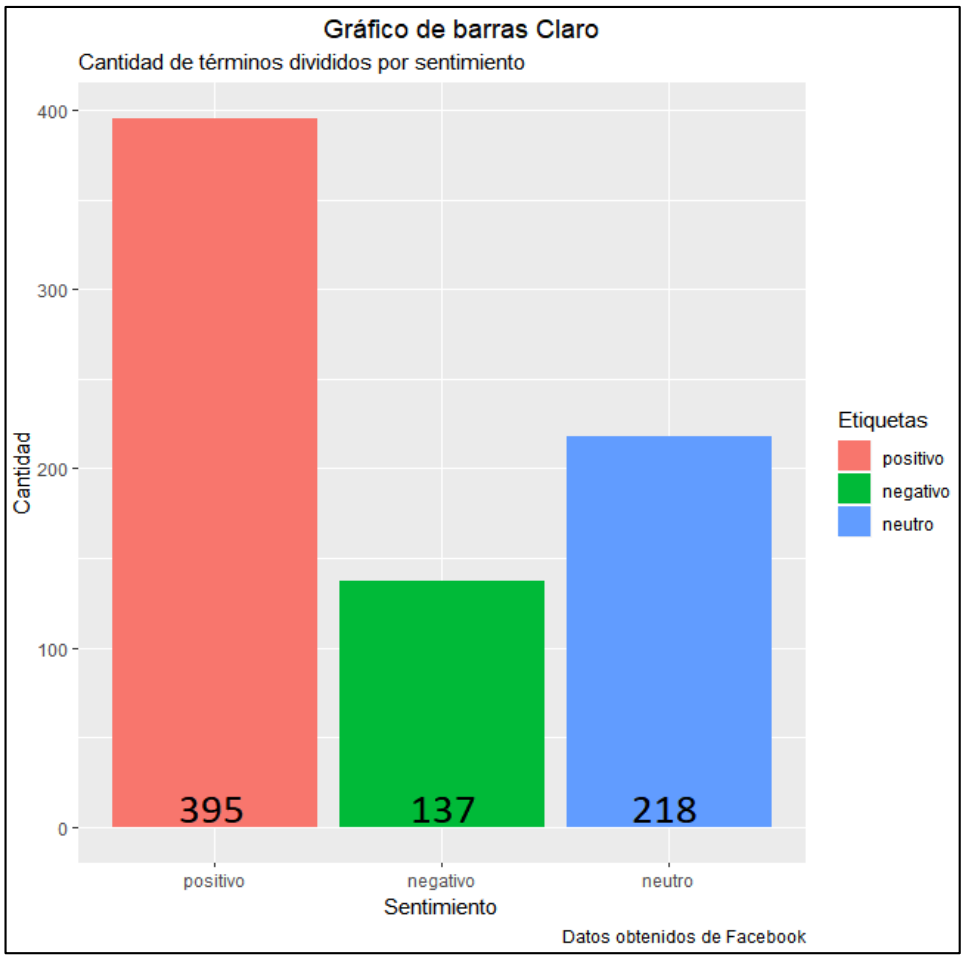


Gráfico 35. Gráfico de barras Claro.

El gráfico de barras de CNT a pesar de ser una empresa pública el nivel de sentimiento como críticas en términos positivos es el más bajo como se muestra en el gráfico 36, con un total de 296 términos positivos, comparado con las otras operadoras de telefonía móvil, pero a diferencia de ello a simple vista se nota que cuenta con el nivel más bajo referente a críticas negativas de 87 términos, llegando a una diferencia de 50 términos por las otras empresas, lo que le da una ventaja en criterio general.

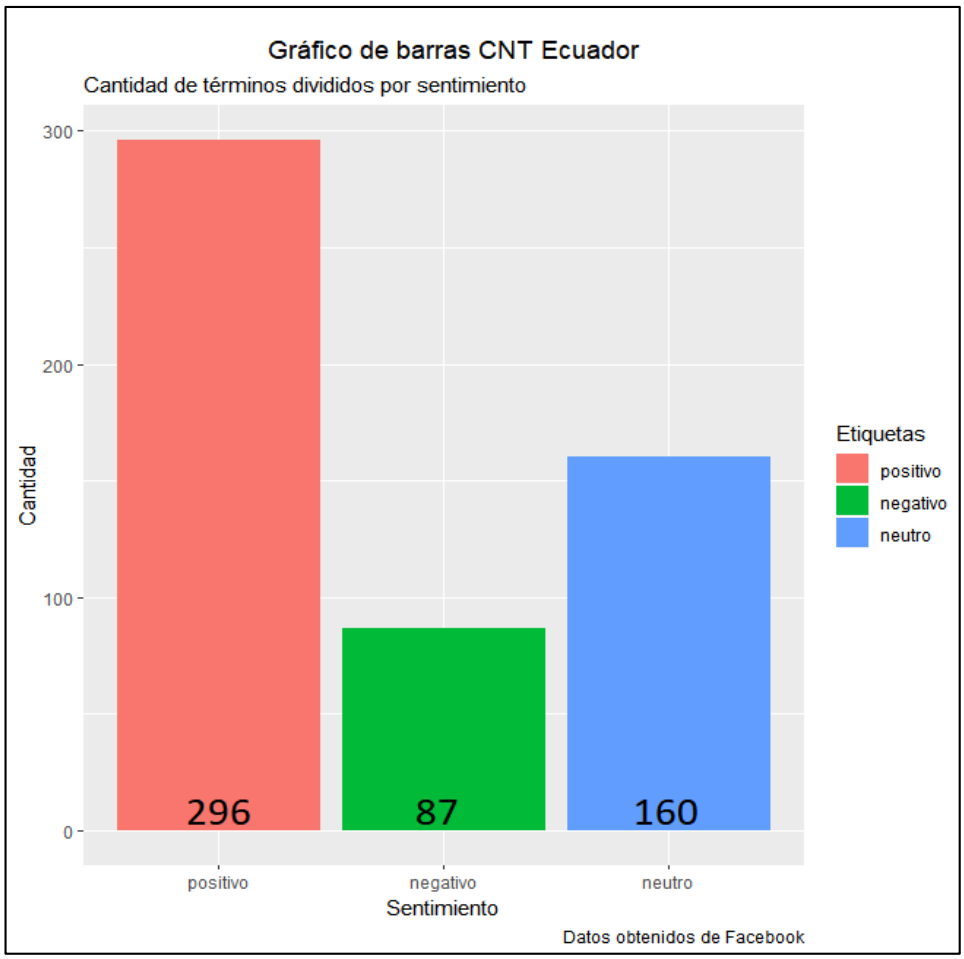


Gráfico 36. Histograma de frecuencia CNT Ecuador.

- *Gráfico de barras de los 10 términos más frecuentes*

Como se mencionó anteriormente los gráficos de barras son usados para comparar dos o más valores y pueden orientarse horizontal o verticalmente. Usamos el diagrama de barras cuando pretendemos hacer una gráfica diferencial. Las barras deben ser estrechas para representar que los valores que toma la variable son discretos, cuyas observaciones se agrupan en categorías. En este caso se utilizarán los primeros diez términos con mayor frecuencia y categorizar por sentimiento, para poder determinar qué es lo más relevante dentro del contexto del servicio de telefonía móvil según criterios de los usuarios seguidores.

Antes de crear los gráficos de los términos más frecuentes por los usuarios seguidores de las empresas de telefonía móvil se ordena de manera decreciente. Lo relevante del gráfico 37, es que toma los datos para construir las gráficas solicitando los renglones del objeto en este caso de la tabla EcMovi, luego utilizamos *ggplot* y se definen los parámetros, en este caso el eje *x* serán los términos (palabras) mientras que el eje *y* serán las frecuencias. Seguido de esto se utiliza *geom_bar* para generar las tablas y como parámetros los colores para las barras, luego de esto se usa *geom_text* el cual ayudará a definir el color del texto dentro de las barras, en este caso a las etiquetas de sentimiento asignando un color blanco con un ajuste de 1.3. Se usa *coord_flip* para voltear el orden de las coordenadas siendo ahora el eje *x* la frecuencia y el eje *y* los diez términos más frecuentes, para una mejor visualización y luego simplemente se agrega título y nombre de los ejes.

```
##############GRÁFICO DE BARRAS DE LAS 10 PALABRAS MÁS FRECUENTES############

#####################################################################
#Ordena de manera decreciente
EcMovi <- EcMovi[order(EcMovi[, 2],decreasing = T),]
#####DIEZ TÉRMINOS MAS FRECUENTES DE MOVISTAR#####
EcMovi[1:10, ] %>%
  ggplot(aes(word, freqMovi)) +
  geom_bar(stat = "identity", color = "black", fill = "dodgerblue4") +
  geom_text(color = "floralwhite",aes(hjust = 1.3, label = sentimentMovi)) +
  coord_flip() +
  labs(title = "Diez términos más frecuentes de Movistar",
       x = "Términos", y = "Frecuencia de uso")
```

Gráfico 37. Función de las palabras más frecuencia.

En el caso de la empresa Movistar como se muestra en el gráfico 38, sobre los diez términos más frecuentes, existe una mezcla de sentimientos con términos etiquetados, en el cual por medio del gráfico de barras, se determina que la palabra 'plan' es la más utilizada por los usuarios seguidores de la empresa Movistar, siendo una palabra con sentimiento positivo y con una frecuencia mayor de 400,

se puede determinar que los usuarios tienen buen criterio sobre el plan o planes que ofrece la empresa como tal. Por otro lado, tenemos la palabra 'numero' con el nivel más bajo de frecuencia, pero dentro de los diez términos más frecuentes, determinando que a pesar de ser una palabra neutra se puede asociar con la palabra 'plan' y determinar que los usuarios realizan muchas menciones en comentarios sobre los planes que adquieren de sus números telefónicos en la red social Facebook.

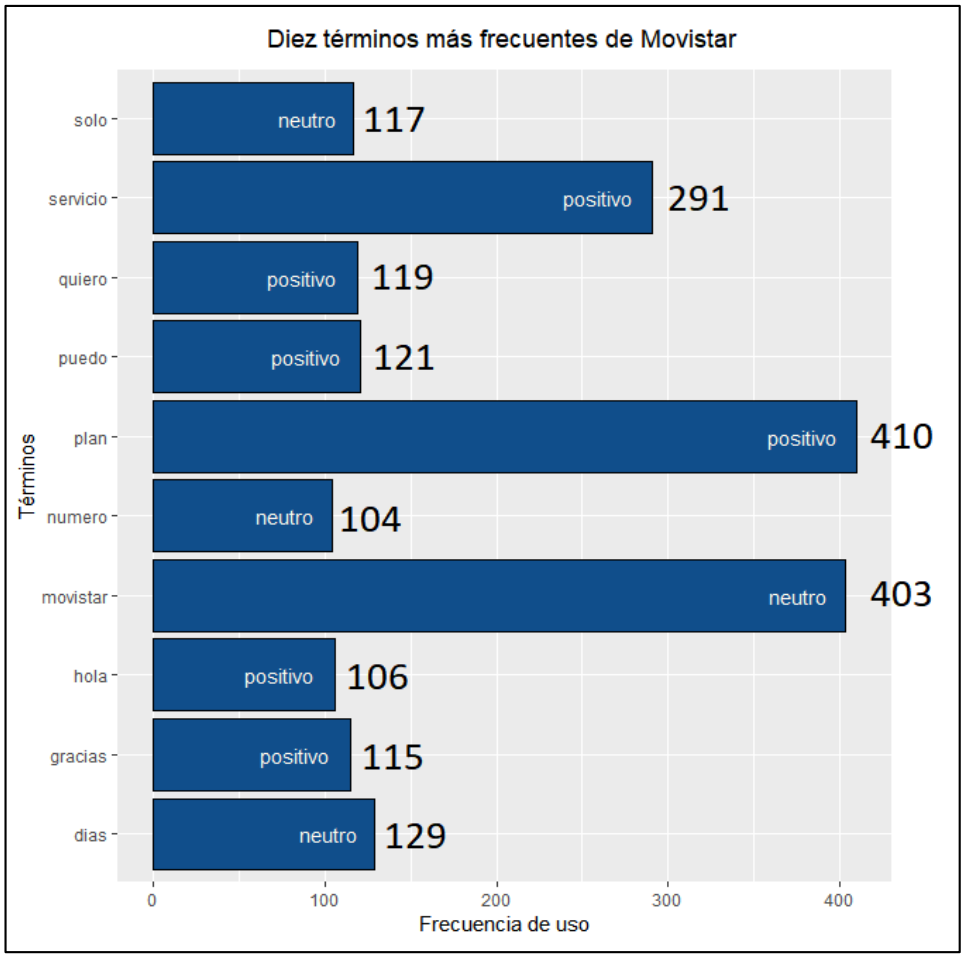


Gráfico 38. Gráfico de barras de los diez términos más frecuentes de Movistar.

En el caso de la empresa Claro, como se muestra en el gráfico 39, la palabra más frecuente es 'servicio' determinando que los usuarios interactúan en la red social de la empresa Claro, realizando una mayor mención sobre el servicio que ofrece la empresa, a pesar de ser tomado como un término de sentimiento positivo

existe posibilidad que no lo sea del todo, dependerá de los demás términos y de los criterios negativos para poder definir si en verdad el servicio es considerado bueno o malo. Luego se tiene la palabra 'dia' con un valor de 163 como frecuencia, siendo la más baja pero dentro del rango de los diez términos más frecuentes y de sentimiento neutro. En comparación de la empresa Movistar, no se puede realizar una analogía de términos con la palabra 'dia', pero sí una relación de manera decreciente en base al nivel de frecuencia con las palabras 'claro' e 'internet' determinando que los usuarios realizan más comentarios sobre el servicio de internet de la empresa.

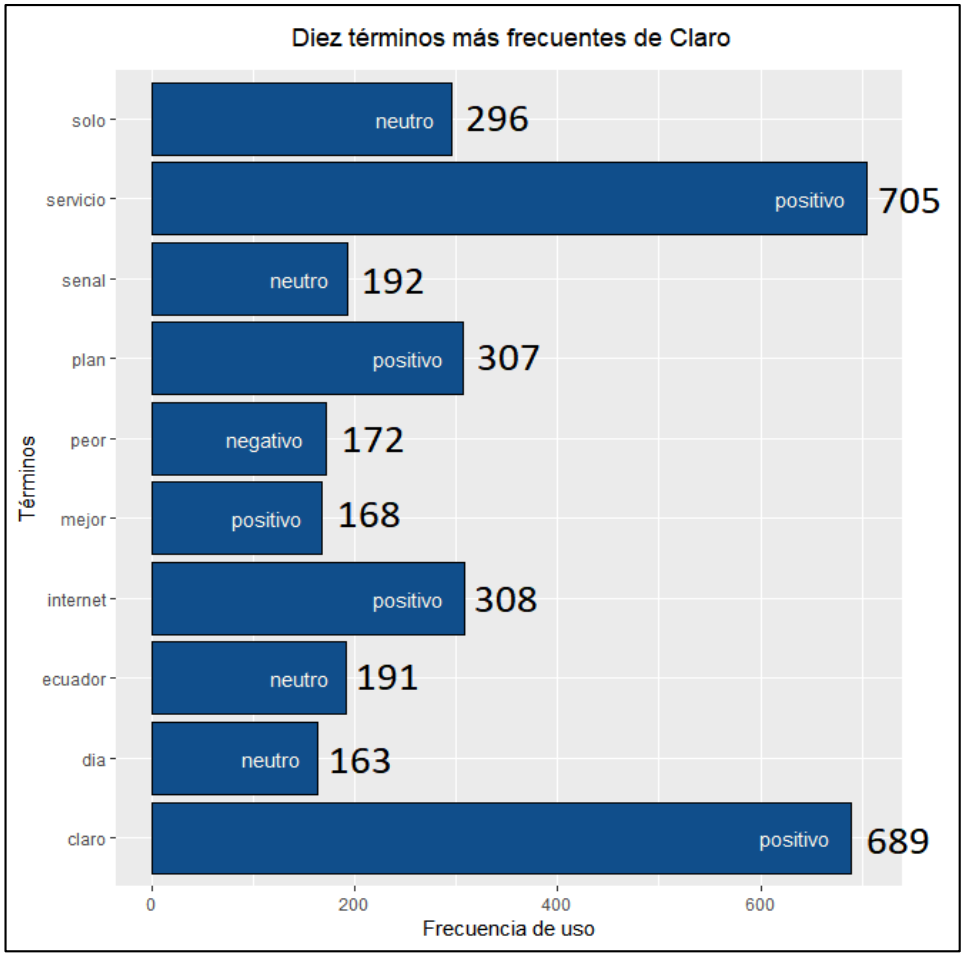


Gráfico 39. Gráfico de barras de los diez términos más frecuentes de Claro.

Por otro lado en el gráfico de frecuencias de la empresa CNT Ecuador, como se refleja en el gráfico 40, se observa claramente que la palabra 'cnt' tiene

dominio por sobre el resto de palabras con una frecuencia de 1748 catalogada como una palabra neutra pero recalcando que CNT es una empresa pública no muy bien vista por los ciudadanos pero es la que mayor interacción posee, por otro lado está la palabra 'internet' y 'servicio' sobrepasando una frecuencia de 1000 catalogando como un sentimiento positivo dentro del contexto generalizado y determinando que los usuarios en la red social Facebook de la empresa CNT, tienen un dominio respecto a comentarios sobre el servicio de internet.

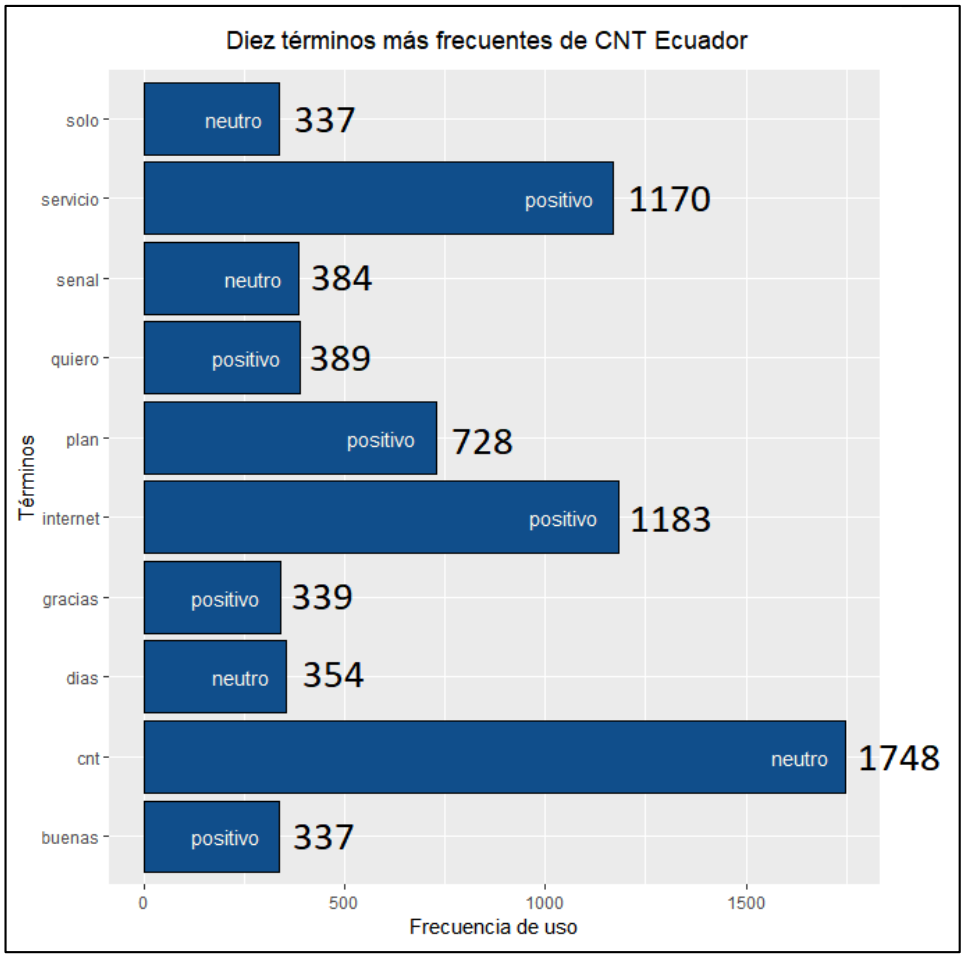


Gráfico 40. Gráfico de los diez términos más frecuentes de CNT Ecuador.

- *Gráfico de barras de los 10 términos más frecuentes como proporción*

La misma información expresada como proporción de uso de cada palabra. Para esta gráfica se usa la función *mutate* de *dplyr* para obtener el porcentaje de uso de cada palabra antes de graficar como se muestra en el gráfico 41.

```
#####PORCENTAJE DE TÉRMINOS DE MOVISTAR#####
EcMovi %>%
  mutate(perc = (freqMovi/sum(freqMovi))*100) %>%
  .[1:10, ] %>% #se usa el punto para evitar que tome la suma solo de los 10 términos
  ggplot(aes(word, perc)) +
  geom_bar(stat = "identity", color = "black", fill = "darkcyan") +
  geom_text(color = "floralwhite",aes(hjust = 1.3, label = round(perc, 2))) +
  coord_flip() +
  labs(title = "Diez términos más frecuentes de Movistar",
       x = "Términos", y = "Porcentaje de uso")

sum(EcMovi$freqMovi)
```

Gráfico 41. Función de frecuencia expresada como proporción.

La única diferencia de la función anterior es que se agrega *mutate*, siendo una función que permite agregar nuevas variables y conservar las existentes. Dentro de la función *mutate* usamos *perc*, el cual se puede usar para extraer percentiles de la distribución de muestreo de una estadística. Percentil es una medida de posición usada en estadística que indica, una vez ordenado los datos de menor a mayor, el valor de la variable por debajo del cual se encuentra un porcentaje dado de observaciones en un grupo de observaciones.

En el gráfico 41, dentro de *perc* se toma como parámetro una función matemática, tomando valores de frecuencia divididos para la suma total de las frecuencias de todas las palabras por cien, para determinar la proporción de los primeros diez términos con mayor frecuencia, tomando el eje *x* como el porcentaje de uso y el eje *y* como los términos más frecuentes.

A partir de las gráficas anteriores se pudo observar que, aunque 'plan' es la palabra más usada, representa el 3.35% del total de uso de términos por parte de la empresa Movistar como se muestra en el gráfico 42.

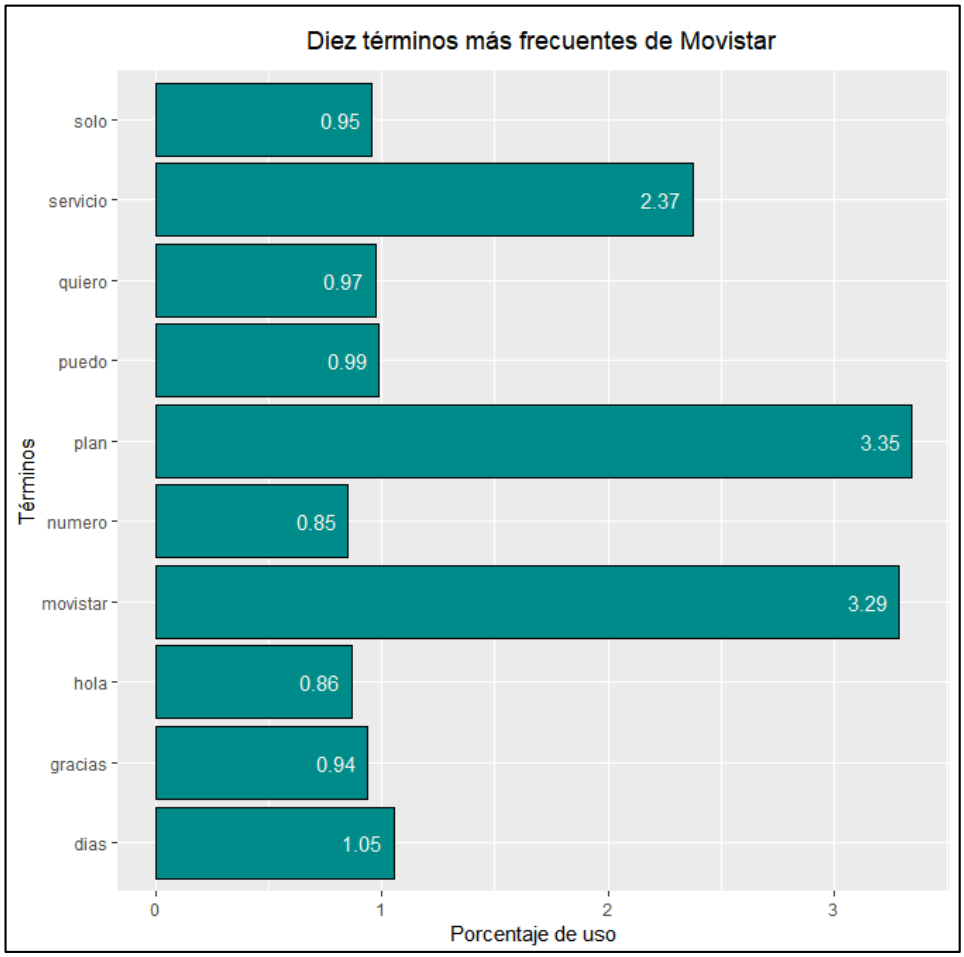


Gráfico 42. Gráfico de barras de los diez términos más frecuentes como proporción de Movistar.

Por parte de la empresa Claro la palabra 'servicio', es la más utilizada por los usuarios, representa el 2.97% del total de uso de términos por parte de la empresa Claro como se muestra en el gráfico 43.

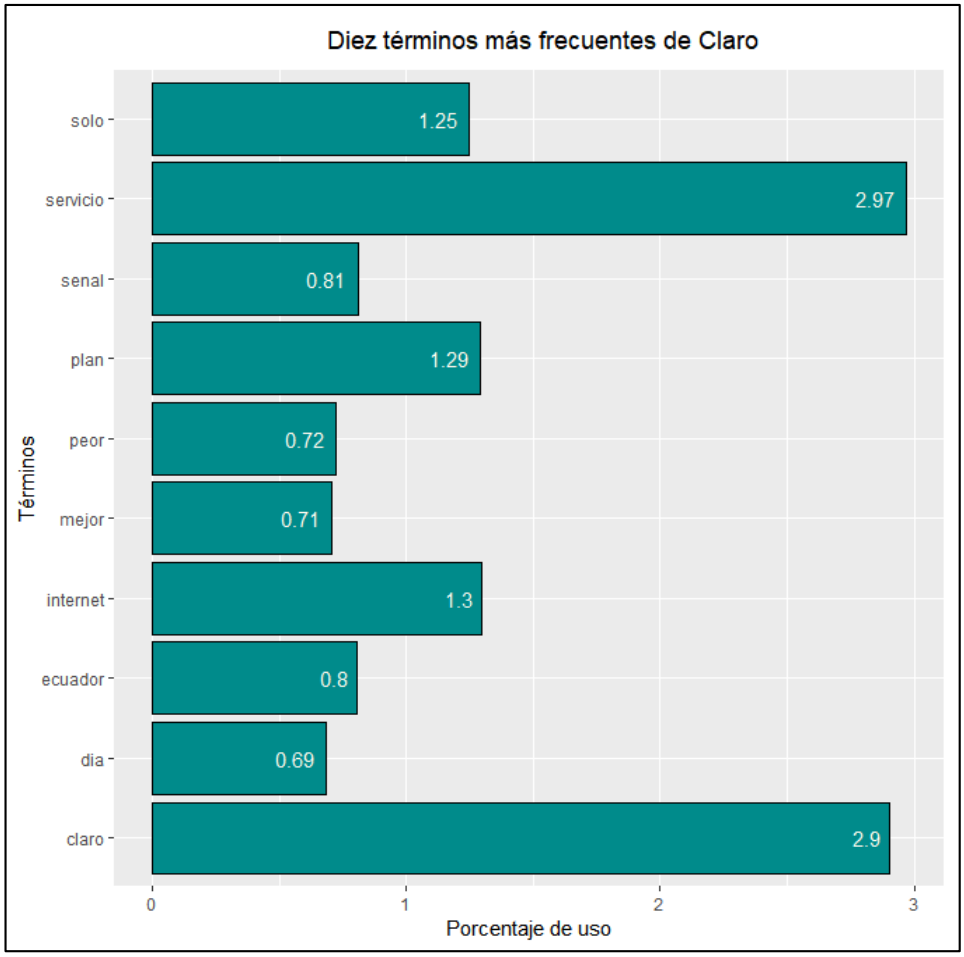


Gráfico 43. Gráfico de barras de los diez términos más frecuentes como proporción de Claro.

En CNT Ecuador la palabra 'cnt', es la más utilizada por parte de los usuarios seguidores, pero a su vez representa solo el 5.06% del total de uso de términos como se muestra en el gráfico 44.

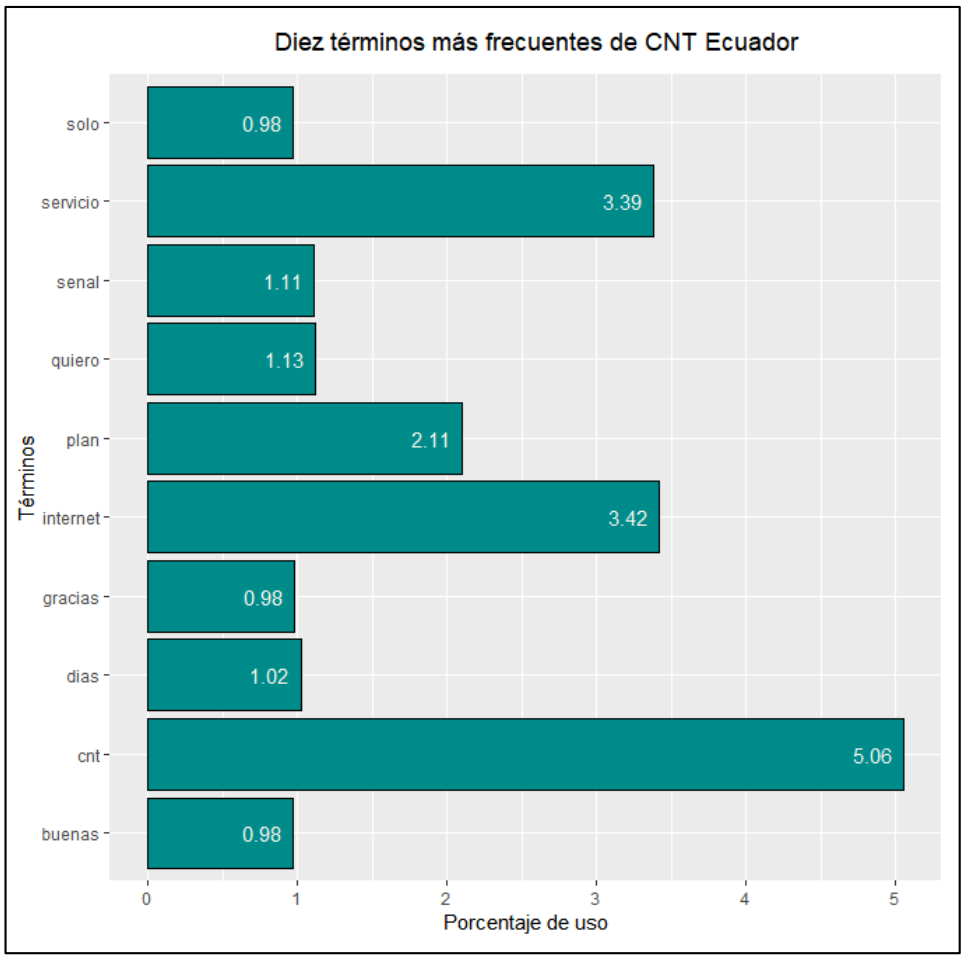


Gráfico 44. Gráfico de barras de los diez términos más frecuentes como proporción de CNT Ecuador.

- *Nubes de palabras*

Las nubes de palabras o wordcloud es un método de visualización que muestra la frecuencia con la que aparece una palabra en un cuerpo de texto dado, siendo esta una forma más atractiva que el diagrama de barras. En estas nubes, el tamaño de la fuente, el color, y la ubicación de la palabra es significativo y facilita la interpretación, es decir que, a mayor tamaño y ubicación central del término, mayor es su frecuencia de uso y las palabras de un mismo color se caracterizan por tener frecuencias similares.

Se ha aplicado el método de nube de palabras a los datos procesados en la fase anterior de las empresas telefónicas Movistar, Claro y CNT Ecuador para

verificar las palabras que tienen mayor frecuencia con un máximo de 100 palabras, para así poder sacar una conclusión del sentimiento de los comentarios emitidos por los usuarios en Facebook.

```
##################NUBE DE PALABRAS#################################
#NUBE DE MOVISTAR
wordcloud(EcMovi$word, max.words = 100,EcMovi$freqMovi,
          random.order = F, colors = brewer.pal(name = "Dark2", n = 8))

#NUBE DE CLARO
wordcloud(EcClaro$word, max.words = 100,EcClaro$freqClaro,
          random.order = F, colors = brewer.pal(name = "PuOr", n = 10))

#NUBE DE CNT
wordcloud(EcCnt$word, max.words = 100,EcCnt$freqCnt,
          random.order = F, colors = brewer.pal(name = "BrBG", n = 11))
```

Gráfico 45. Función de wordcloud por empresa.

La función de *wordcloud* permite crear la nube de palabras como muestra en el gráfico 45, donde *EcMovi$word* es la tabla donde se tiene almacenado los datos procesado en la fase anterior de la empresa telefónica Movistar, se usa el signo de dólar para llamar la variable *Word,* con máximo de 100 palabras y para ello se utiliza *max.word=100, EcMovi$FreqMovi* sirve para llamar la variable *FreqMovi* usando las palabras más frecuente, *random.order=F* sirve para trazar palabras en orden aleatorio, pero como es falso se trazaran en frecuencia decreciente y por último el argumento *colors = brewer.pal (name = "Dark2", n=8)* es para definir el color de las palabras que aparecerán en la nube de palabras de menor a mayor frecuencia, usando el paquete *brewer.pal* para cumplir con esta función.

Básicamente esta es la forma más sencilla que se ha implementado para crear las nubes de palabras, la función que se explicó se usará para las otras dos empresas telefónicas Claro y CNT Ecuador, puesto que solo se cambiaran las variables donde se encuentra procesados los datos.

Gráfico 46. Nube de palabras de la empresa telefónica Movistar.

En base a los comentarios extraídos de Facebook de la empresa Movistar en el gráfico 46, se puede observar que las palabras 'movistar' y 'plan' tienen una mayor frecuencia y relevancia, seguido de la palabra 'servicio', dándole a esto un significado positivo, pero a su vez también se pueden observar palabras que son poco frecuentes por el bajo volumen de intensidad como 'saldo', 'llamadas' y 'cambiar'.

Gráfico 47. Nube de palabras de la empresa telefónica Claro.

En el gráfico 47, las palabras 'servicio' y 'claro' son las de mayor volumen, destacando el servicio de telefonía móvil en la empresa, a su vez le siguen palabras como 'internet', 'plan', 'peor' y 'mejor' de tal manera que se puede determinar que los planes de internet están en debate porque la mitad de los usuarios aceptan que el servicio es bueno y la otra mitad no lo cree.

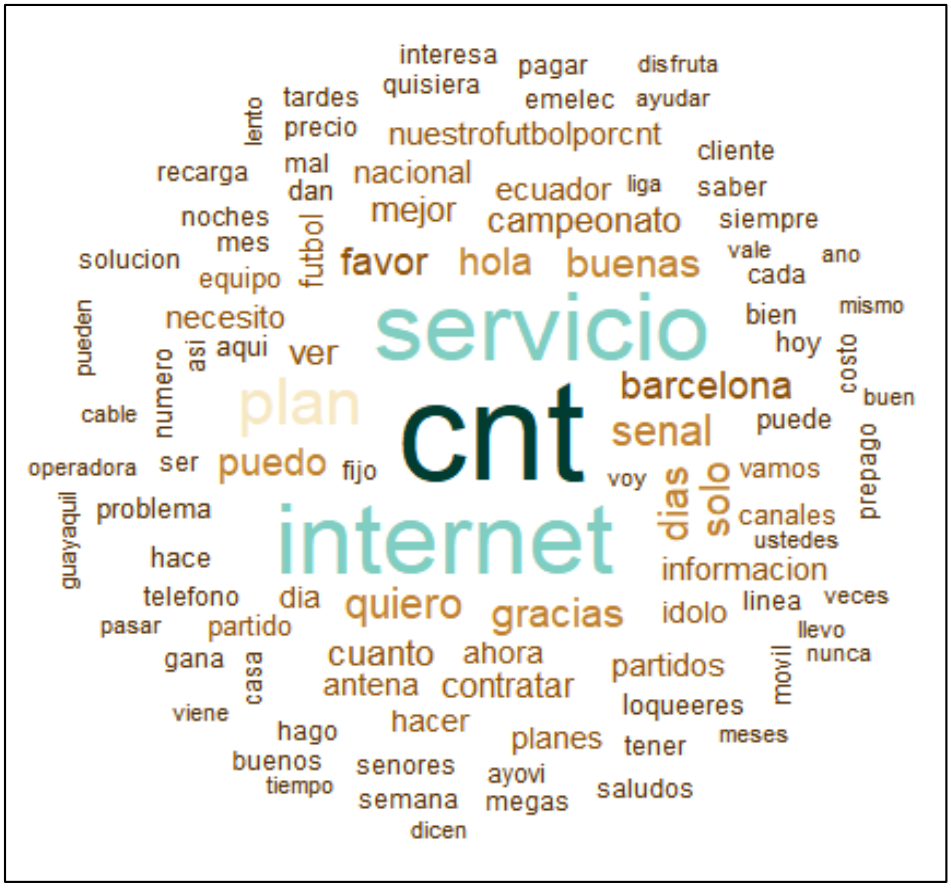


Gráfico 48. Nube de palabras de la empresa telefónica CNT Ecuador.

De la misma manera en el gráfico 48, los comentarios de los seguidores de la empresa CNT al igual que Claro y Movistar guarda cierta relación en sus palabras más frecuentes, se puede observar que las palabras con mayor frecuencia son 'cnt', 'servicio' e 'internet', pero también las de menor frecuencia como 'quiero', 'senal', 'guayaquil', determinando que a pesar de ser una empresa pública y con mayor número de interacciones de usuarios, la mayor parte de los internautas de la ciudad de Guayaquil comentan mucho sobre la señal que ofrece la empresa.

Dada la cantidad de comentarios, opiniones, emociones, que se manifiesta a través de la red social Facebook resultó útil la extracción de este tipo de información para realizar un análisis de las tres empresas telefónicas, con el fin de monitorear las opiniones de los seguidores de cada empresa por el tamaño de

frecuencia y sacar una conclusión de cuál de ellas tiene un mejor estatus a nivel nacional con respecto a los servicios que ofrecen. Por ello se crearon nubes de palabras por cada empresa evaluando sus palabras en forma generalizada. Ahora se elaborará dos nubes de palabras de las tres empresas antes mencionada para hacer un análisis de las palabras positivas y negativas de las 3 empresas juntas, es decir una nube de palabras positiva y la otra negativa.

```
#CREAR BASE DE DATOS DE PALABRAS POSITIVAS
Movipositivo=subset(EcMovi,sentimentMovi=="positivo")
Claropositivo=subset(EcClaro,sentimentClaro=="positivo")
Cntpositivo=subset(EcCnt,sentimentCnt=="positivo")

#AGRUPAMIENTO DE PALABRAS POSITIVAS QUE SE REPITAN ENTRE LAS 3 EMPRESAS
PositivosCC =Reduce(merge, list(Movipositivo, Claropositivo, Cntpositivo))

#DIFERENCIA DE FRECUENCIA DE SENTIMIENTO POSITIVO
difference <- abs(PositivosCC[, 2] - PositivosCC[,4] - PositivosCC[,6])
#AGREGAR COLUMNA DIFERENCIA A LA TABLA POSITIVOS
PositivosCC <- cbind(PositivosCC, difference)
#ORDENAR DE MANERA DECRECIENTE
PositivosCC <- PositivosCC[order(PositivosCC[, 8],decreasing = T), ]
```

Gráfico 49. Función para la creación de base de datos de palabras positivas de las empresas telefónicas Movistar, Claro y CNT Ecuador.

El gráfico 49, muestra cómo se crea la base de datos de las palabras positivas, con la función *subset* que permite tomar todas las palabras que tienen un sentimiento positivo, de los datos originales que se encuentran en tablas de las empresas telefónicas Movistar, Claro y CNT Ecuador, luego la variable *sentimientMovi* debe ser igual al sentimiento que se desea extraer en una nueva base de datos y se usa el doble igual para determinar una igualdad para que el sentimiento sea positivo, el mismo procedimiento se realiza para las otras dos empresas. Una vez creada la nueva base de datos se realiza un agrupamiento de palabras positivas que se repitan en las tres empresas utilizando la función *reduce* y *merge* que sirve para juntar las filas que tienen el mismo valor en los campos de cruce y toma los nombres de las tablas como parámetros.

Continuando con el código, la función *difference* determina el orden de la diferencia de frecuencia de las palabras agrupadas y *abs* se usa para encontrar un valor positivo absoluto en la tabla *PositivosCC* y en corchete se colocan los valores de la columna dos que son los valores de frecuencia y realiza una resta tomando la misma variable, pero referenciando a la columna cuatro de la segunda frecuencia, después de la misma manera se hace una resta con la columna seis de la tercera frecuencia, luego con la función *cbind* se agrega la columna *difference* a la variable *PositivosCC,* se ordena de manera decreciente con la siguiente función que es true (verdadero) *decreasing=T.*

De la misma manera se sigue los mismos pasos para crear la nube de palabras negativas con la única diferencia que se cambian las variables.

```
#WORDCLOUD DE GRUPOS POSITIVOS Y NEGATIVOS#

#NUBE DE PALABRAS POSITIVAS AGRUPADAS DE LAS EMPRESAS MOVISTAR-CLARO-CNT
wordcloud(PositivosCC$word,PositivosCC$difference,
          random.order = F, colors = brewer.pal(name = "Dark2", n = 8))

#NUBE DE PALABRAS NEGATIVAS AGRUPADAS DE LAS EMPRESAS MOVISTAR-CLARO-CNT
wordcloud(NegativosCC$word,NegativosCC$difference,
          random.order = F, colors = brewer.pal(name = "RdGy", n = 11))
```

Gráfico 50. Función de wordcloud positivos y negativos.

El gráfico 50, muestra el código de cómo crear las nubes de palabras positivas y negativas básicamente en la misma función que se usó en el gráfico 45, la única diferencia es el cambio de variables: *PositivoCC, NegativoCC* y tomando la columna con los valores de *difference.*

Las siguientes tablas hacen referencias a palabras separadas y catalogadas como positivas y negativas de las empresas telefónicas Movistar, Claro y CNT Ecuador:

Tabla 14. Referencia de las palabras positivas de las empresas telefónicas Movistar, Claro y CNT Ecuador.

Word	freqMovi	freqClaro	freqCnt	difference
servicio	291	705	1170	1584
internet	66	308	1183	1425
claro	45	689	53	697
plan	410	307	728	625
quiero	119	147	389	417
mejor	45	168	262	385
favor	94	137	308	351
gracias	115	113	339	337
buenas	103	81	337	315
hola	106	89	326	309

Tabla 15. Referencia de las palabras negativas de las empresas telefónicas Movistar, Claro y CNT Ecuador.

Word	freqMovi	freqClaro	freqCnt	difference
peor	47	172	78	203
mal	43	81	122	160
problema	89	103	144	158
esperando	26	94	86	154
lento	5	41	105	141
nadie	37	81	72	116
pesima	20	74	62	116
problemas	30	60	77	107
nunca	50	69	87	106
quisiera	45	28	123	106

En la tabla 14, se muestra las palabras positivas y en la tabla 15, las palabras negativas que se agrupan con mayor frecuencia de las empresas telefónicas Movistar, Claro y CNT Ecuador, la frecuencia de cada una y el orden de diferencia que existe entre las tres empresas.

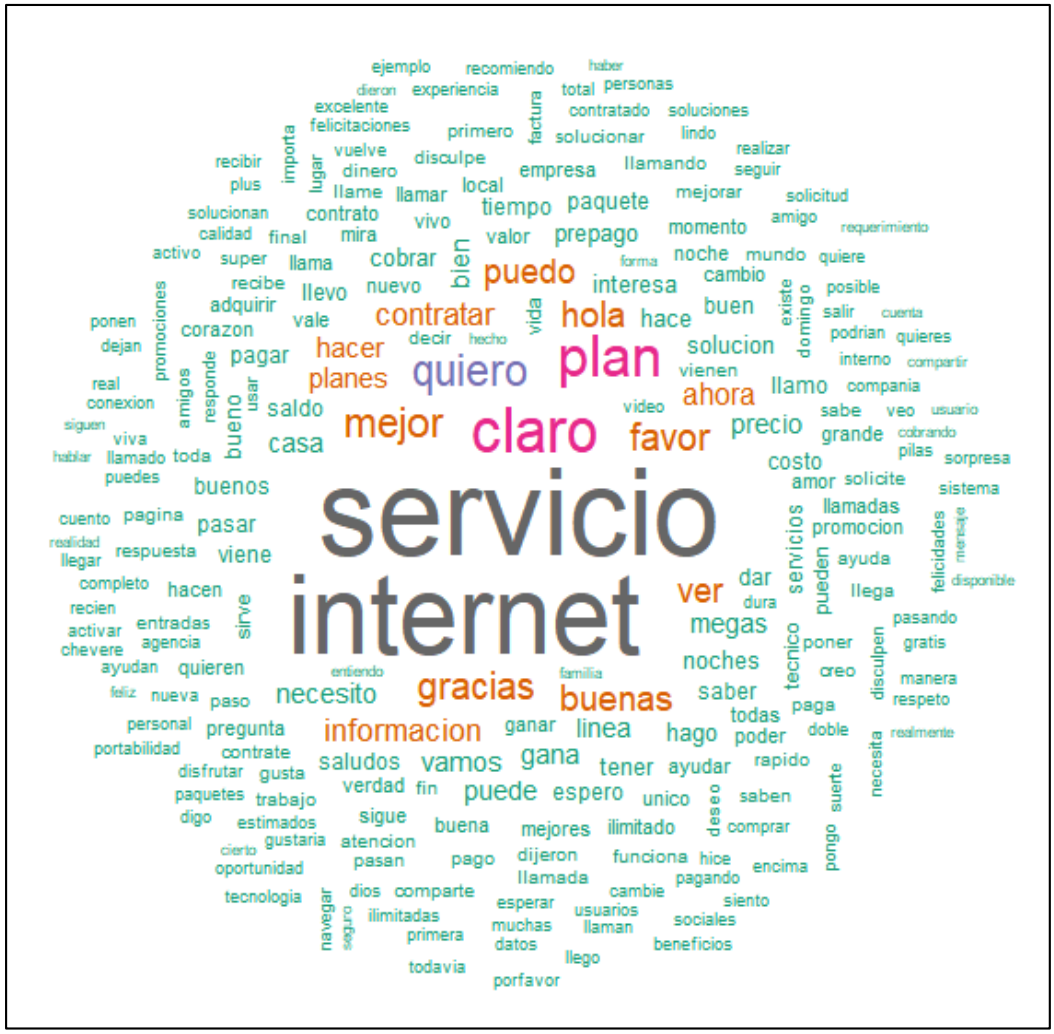


Gráfico 51. Nube de palabras positivas de las empresas telefónicas Movistar, Claro y CNT Ecuador.

En el gráfico 51, se puede observar una agrupación de las palabras positivas que existen en los comentarios extraídos de la red social Facebook de las empresas telefónicas Movistar, Claro y CNT Ecuador, dando como resultado las palabras con mayor frecuencia 'servicio' e 'internet'. Lo curioso de la agrupación de palabras positivas relacionadas entre empresas es que dentro de este conjunto de términos se encuentra la palabra 'claro', es decir que dentro del contexto de las empresas agrupadas los usuarios seguidores hablan de Claro de una manera frecuente, realizando comparaciones y diferencias sobre los servicios que ofrece.

Gráfico 52. Nube de palabras negativas de las empresas telefónicas Movistar, Claro y CNT Ecuador.

En el gráfico 52, se muestra una agrupación de palabras negativas acerca de las empresas telefónicas, pero esta tiene un menor volumen que el gráfico 51, es decir que tiene pocos comentarios negativos referente a los productos y servicios que ofrecen a través de la red social Facebook. Recalcando como puntos negativos las palabras 'problema', 'peor', 'lento' y 'pesimo', es decir que un grupo determinado de personas de las tres empresas de telefonía móvil, realizan referencias de comentarios sobre los problemas de servicios que ofrecen más que todo de internet, siendo este lento y pésimo.

- *Pirámides*

Las pirámides son gráficos de barras horizontales cuya longitud es proporcional a la frecuencia de palabras comunes que existen entre empresas telefónicas, dicha información que muestra las pirámides se toma en base a la resta de la frecuencia mayor con la menor y con este resultado se forma la pirámide de manera decreciente con las palabras comunes con mayor frecuencia.

Estos tipos gráficos son ideales para obtener una idea general de las características comunes que existen entre las empresas telefónicas de Movistar, Claro y CNT Ecuador.

```
###############PIRAMIDE DE TÉRMINOS POSITIVOS MOVISTAR - CLARO ###################

Piramide1 = merge(Movipositivo,Claropositivo)#AGRUPAR DATOS DE MOVISTAR Y CLARO
head(Piramide1)

#DIFERENCIA DE FRECUENCIA
difference <- abs(Piramide1[, 2] - Piramide1[,4])
#AGREGAR COLUMNA DIFERENCIA A LA TABLA AGRUPADA DE MOVISTAR Y CLARO
Piramide1 <- cbind(Piramide1, difference)
#ORDENAR DE MANERA DECRECIENTE
Piramide1 <- Piramide1[order(Piramide1[, 6],decreasing = T), ]

#DATA FRAME DE LOS PRIMEROS 10 TÉRMINOS
Piramide1 <- data.frame(x = Piramide1[1:10, 2],
                        y = Piramide1[1:10, 4],
                        labels = Piramide1[1:10,1])

#PIRAMIDE MOVISTAR - CLARO
pyramid.plot(Piramide1$x, Piramide1$y,labels = Piramide1$labels,
             main = "Términos positivos en común",gap = 100,
             laxlab = NULL,raxlab = NULL, unit = NULL,
             top.labels = c("Movistar","Términos","Claro"))
############################################################
```

Gráfico 53. Función de pirámide positiva.

Para realizar una pirámide positiva se hace uso del siguiente código como se muestra en el gráfico 53. Se crea una variable llamada *Piramide1* donde agrupa los datos positivos de las empresas telefónicas Movistar y Claro, con la siguiente función *difference <- abs (Piramide1[,2] - Piramide1[,4])* el cual permitirá realizar una diferencia de frecuencia, *cbind* permite agregar una columna de la diferencia de las tablas de datos agrupados, *Piramide1[order(Piramide1[,6],*

decreasing =T),] ordena la pirámide de manera decreciente y con el *data.frame* se realiza la lista de las palabras más frecuentes y la relación en común de ambas empresas. Luego se usa la función *pyramid.plot* para crear la pirámide en barras horizontales, tomando de la variable *Piramide1* donde se tiene almacenado los datos de las empresas Claro y Movistar, se asigna una etiqueta a cada valor de *x* y *y*, con la función *main* se agrega el título principal 'Términos positivos en común', los argumentos *laxlab, raxlab* y *unit* se declaran nulos, no se asigna ninguna etiqueta adicional al eje izquierdo, derecho y unidad de trama y para finalizar con *top.label* se le da un nombre a las categorías representadas en el lado izquierdo 'Movistar' y en el lado derecho de la trama 'CNT Ecuador' y un encabezado para la etiqueta en el centro 'Términos'. Es así como se realiza el gráfico de pirámide, este mismo código se usa para hacer la comparación entre las tres empresas por términos tanto positivos como negativos.

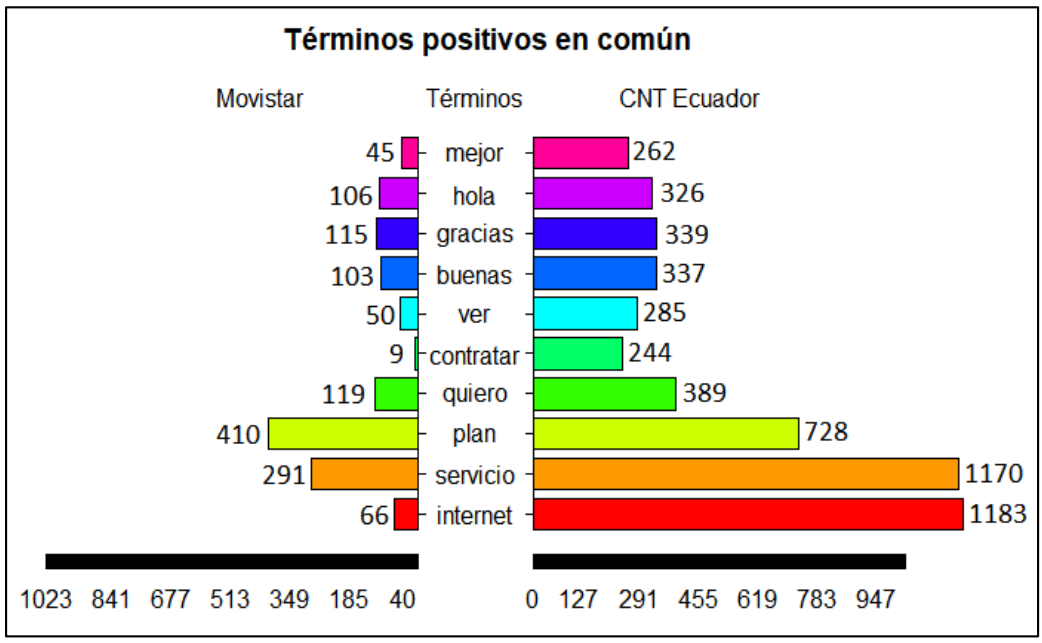


Gráfico 54. Términos positivos en común entre empresas Movistar y CNT Ecuador.

En el gráfico 54, se observa un gráfico doble de frecuencias que se disponen de forma horizontal, sobre la línea de las abscisas que indican dos

empresas telefónicas, de lado izquierdo Movistar y CNT Ecuador de lado derecho. En el eje de las ordenadas se disponen los términos positivos en común que existen entre las empresas antes mencionadas, el volumen de las barras depende de la frecuencia que tenga cada palabra. En el gráfico de pirámides podemos destacar que las palabras 'internet', 'servicio' y 'plan' son más frecuentes en CNT que en Movistar es decir que entre las dos empresas las personas prefieren contratar más el servicio de planes de internet que ofrece CNT en referencia a los términos positivos.

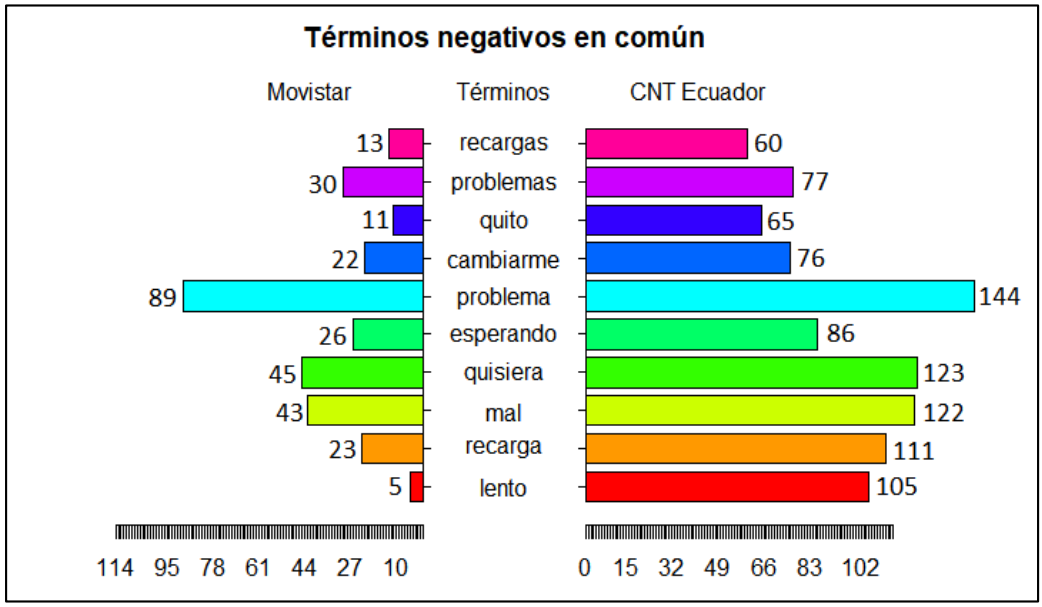


Gráfico 55. Términos negativos en común entre empresas Movistar y CNT Ecuador.

En el gráfico 55, se realiza una comparación de términos negativos más comunes entre las empresas Movistar y CNT Ecuador, donde destacamos las palabras 'problema', 'lento', 'cambiarme', determinando que el uso de palabras negativas es más frecuente en CNT que en Movistar, sobre todo con respecto a problemas de servicio lento y que la mayoría de los usuarios desean cambiarse de operadora.

En comparación de términos positivos y negativos entre ambas empresas las personas desean contratar más los servicios de CNT por sus promociones y planes, pero los que ya poseen esos servicios tienen malas experiencias por lo que sugieren cambiarse de operadora móvil.

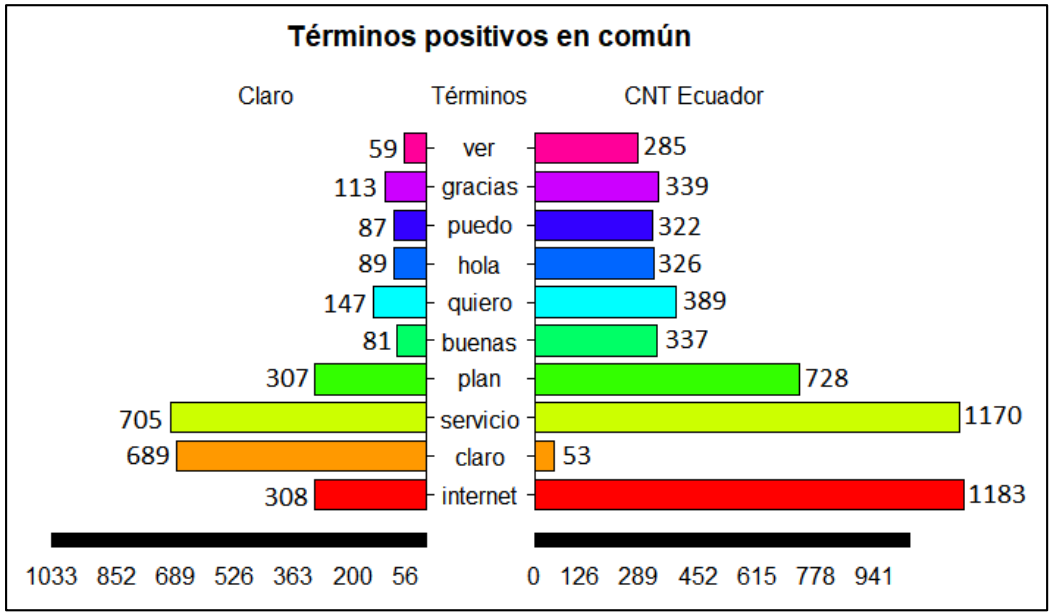


Gráfico 56. Términos positivos en común entre empresas Claro y CNT Ecuador.

De la misma manera en el gráfico 56, hace una comparación de los términos positivos en común entre las empresas telefónicas Claro y CNT Ecuador, donde se observa que CNT tiene una mayor frecuencia en sus palabras que en Claro, destacando las palabras 'internet', 'servicio' y 'plan', a pesar de ser una de las empresas más criticadas, CNT cuenta con un mayor número de interacciones de manera positiva en palabras frecuentes con respecto de Claro que a pesar de ser también una de las empresas con una buena reputación no supera el nivel de frecuencia de CNT.

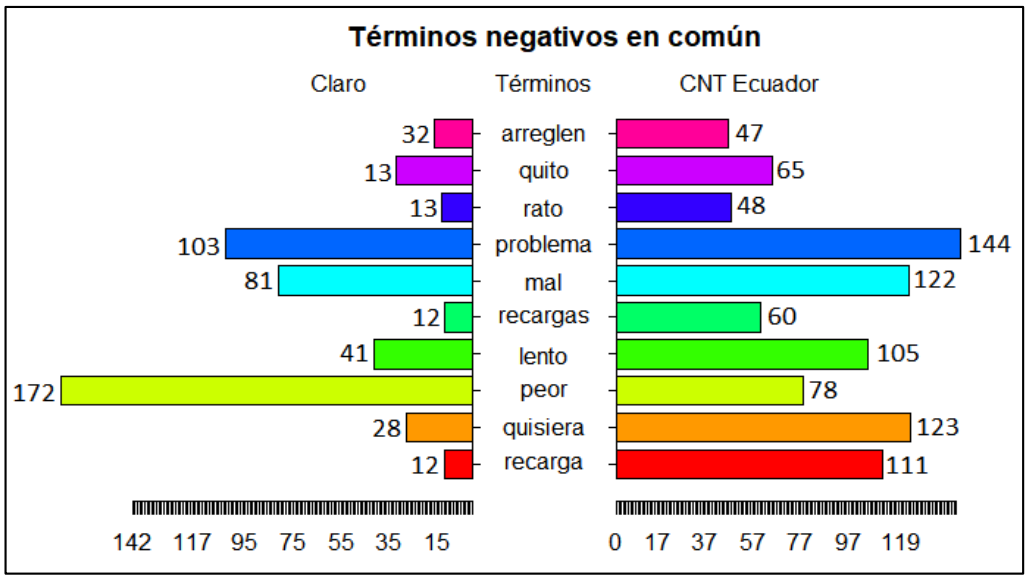


Gráfico 57. Términos negativos en común entre empresas Claro y CNT Ecuador.

En el gráfico 57, se observan los términos negativos que se relacionan entre Claro y CNT, destacando las palabras 'problema', 'peor', 'mal' y 'quisiera'. En el cual determinamos que en CNT las personas tienden a realizar muchos comentarios negativos con respecto a problemas que presentan con el servicio que ofrece la telefonía móvil, a pesar de que en ambas empresas el nivel de frecuencia de la palabra 'problema' no presenta mucha variación, los usuarios prefieren cambiarse de operadora más en CNT que en Claro.

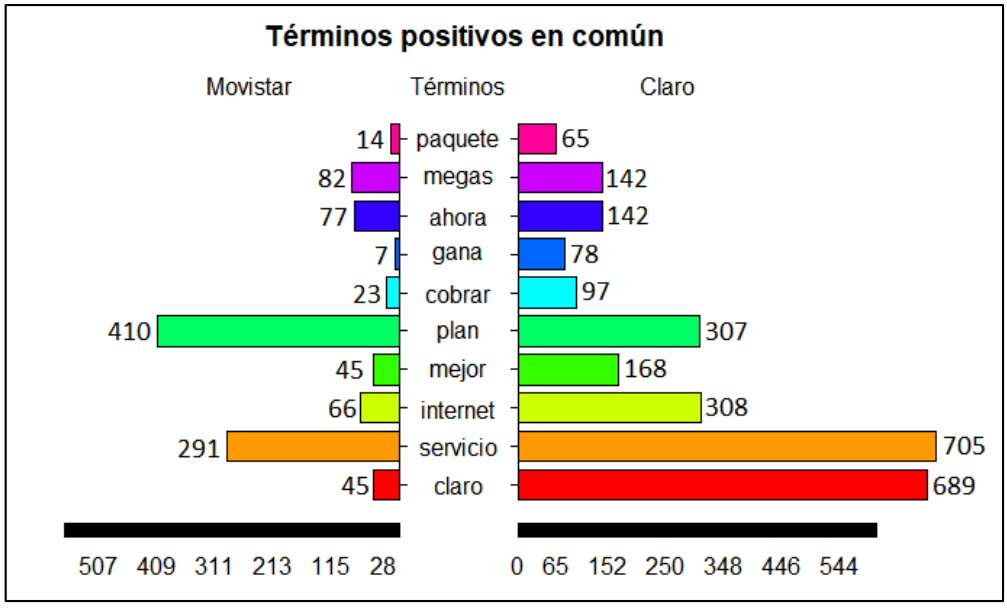


Gráfico 58. Términos positivos en común entre empresas Movistar y Claro.

En el gráfico 58, se observan los términos positivos que se relacionan entre Movistar y Claro, destacando las palabras 'claro', 'servicio', 'internet' y 'plan'. En el cual determinando que en Claro el servicio de internet es mucho más frecuente que en movistar pero que los usuarios realizan más contratos en planes en Movistar que en Claro.

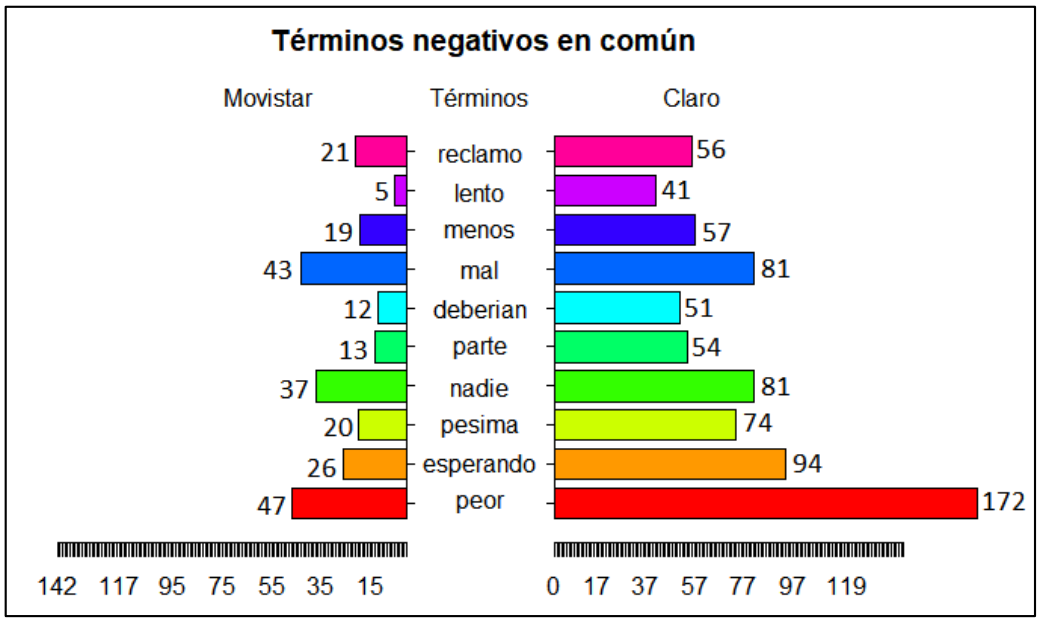


Gráfico 59. Términos negativos en común entre empresas Movistar y Claro.

En el gráfico 59, se observan los términos negativos que se relacionan entre Movistar y Claro, destacando las palabras 'peor', 'esperando', 'mal' y 'reclamo'. En el cual determinamos que en Claro las personas tienden a realizar muchos comentarios negativos de reclamos con respecto a problemas que presentan con el servicio que ofrece la telefonía móvil, mientras que en Movistar no existe un nivel elevado sobre problemas y conflictos en servicios y contrataciones de planes móviles. En el cual determinamos que las personas prefieren contratar Movistar por encima de Claro por el bajo nivel de frecuencia referente a términos negativos.

CAPÍTULO IV

EVALUACIÓN DE RESULTADOS

La fase de evaluación de la metodología CRISP-DM, se enfoca en analizar los resultados obtenidos de los modelos seleccionados en la fase anterior para verificar el cumplimiento de los objetivos del negocio en este caso asociado a los objetivos de minería de texto establecidos en el proyecto como medida de prevención y alternativa para mejorar el servicio en las empresas de telefonía móvil del país Movistar, Claro y CNT Ecuador. Además, se describe de forma objetiva que modelo es más útil para su aplicación de estudio sobre qué empresa posee un mejor servicio donde el usuario sea el beneficiado al momento de inclinarse por una de las anteriormente mencionadas.

4.1 Evaluar los resultados

- *Objetivo 1*

Clasificar las opiniones de los usuarios seguidores en términos positivos, negativos y neutros en base al servicio que ofrecen las telefonías móviles Movistar, Claro y CNT Ecuador. Para cumplir con este objetivo se usó los gráficos de barras generales donde se realizó una comparativa del modelo presentados en el capítulo anterior de las empresas de telefonía móvil sobre la cantidad de términos como participación de los usuarios en la red social Facebook sobre el servicio que ofrece las empresas antes mencionadas.

Tabla 16. Resultados de los gráficos de barras.

Resultados de participación (Cantidad de términos catalogados por sentimiento)				
Empresas	Positivo	Negativo	Neutro	Total
Movistar	319	134	145	598
Claro	395	137	218	750
CNT Ecuador	296	87	160	543
Total	1010	358	523	1891

Como se puede observar en la tabla 16, se define la cantidad total de términos y la cantidad catalogados por sentimiento, con un total de 1891 términos entre ellos los frecuentes y los no tan frecuentes. Existe una mayor participación positiva de los usuarios en la interacción con las empresas telefónicas, destacando a Claro con una cantidad total de 395 términos positivos. A su vez Claro también se destaca en la cantidad de términos negativos con un total de 137 catalogándolo como una empresa con mayor interacción de usuarios en la red social Facebook, pero no de carácter favorable. Por otro lado, Movistar cuenta con un total de 598 términos del cual se destaca en el sentimiento positivo con un total de 319 términos, mientras que CNT cuenta con un total de 543 términos siendo la empresa con menos cantidad de términos en el ámbito de interacción en redes sociales sobre telefonía móvil.

Determinando en base al gráfico de barras 35 del capítulo anterior y la tabla 16, de manera general, la empresa Claro se posiciona por ser la de mayor interacción, teniendo una cantidad elevada en participación de los internautas, mientras que Movistar se posiciona por tener una mejor frecuencia en términos positivos, tal como se identificó en el gráfico de barras 38 y CNT queda por debajo

de las dos empresas anteriormente mencionadas por tener la menor cantidad de términos en base a la interacción que presenta por parte de los usuarios seguidores.

- *Objetivo 2*

Realizar una comparación de los términos comunes positivos y negativos con mayor frecuencia existente en las empresas telefónicas Movistar, Claro y CNT Ecuador. Para ello se utilizaron los gráficos de barras de términos más frecuentes, siendo estos: Movistar gráfico 38, Claro gráfico 39 y CNT Ecuador gráfico 40, las nubes de palabras y los gráficos de pirámides del capítulo anterior.

- Resultados de los gráficos de barras términos más frecuentes

En la siguiente tabla se realiza una comparativa del modelo de barras en base a los términos más frecuentes, presentados en el capítulo anterior de las empresas de telefonía móvil sobre el nivel de frecuencia en términos de la participación de los usuarios en la red social Facebook sobre el servicio que ofrece Movistar, Claro y CNT Ecuador.

Tabla 17. Resultados de los gráficos sobre los términos más frecuentes.

Resultados de participación (Base de términos catalogados por frecuencia)						
Términos	Movistar	Freq1	Claro	Freq2	CNT Ecuador	Freq3
Término1	plan	410	servicio	705	cnt	1748
Término2	movistar	403	claro	689	internet	1183
Término3	servicio	291	internet	308	servicio	1170
Término4	dias	129	plan	307	plan	728
Término5	puedo	121	solo	296	quiero	389
Término6	quiero	119	senal	192	senal	384

Como se puede observar en la tabla 17, se define los primeros seis términos por frecuencia de manera decreciente por empresa. A pesar de que CNT es la empresa con la menor cantidad de términos etiquetados anteriormente por sentimiento, cuenta con términos que son muchos más mencionados que los de Movistar y Claro, como el término 'cnt', 'internet' y 'servicio', seguido de la empresa Claro y por último Movistar. De las tres empresas entre los términos más frecuentes tenemos la palabra 'servicio', siendo una palabra muy mencionada por parte de los internautas, pero de mayor mención en la empresa CNT con una frecuencia de 1170, determinando que en la empresa CNT es donde más se habla e interactúa sobre el servicio de telefonía móvil.

Cabe recalcar que el nivel de frecuencia nos permite identificar qué clase de términos son más utilizados por los usuarios seguidores de las empresas de telefonía móvil en la red social Facebook, pero sin determinar si esas menciones afectan o no a la reputación de las anteriormente mencionadas.

- Resultados de las nubes de palabras

En cuanto a las nubes de palabras o wordcloud tienen la finalidad de mostrar las palabras y las frases más importantes que usaron los internautas, en una nube de palabras. Cuanto más grande sea la letra, más importante o significativa será la palabra. Resaltando las de mayor intensidad y dejando a un lado las de menor intensidad. De manera que destacan por el nivel de frecuencia de términos de cada empresa haciendo visible y de fácil entendimiento para el receptor.

En el caso de Movistar en la nube de palabras del gráfico 46, determinamos que se habla mucho sobre los planes telefónicos, pero de manera negativa ya que

resalta el término problema y del cual los clientes esperan una solución a este inconveniente.

Por otro lado, en la nube de palabras del gráfico 47 de la empresa Claro, determinamos que se habla mucho sobre el servicio de internet, pero hay un debate de que cierta cantidad de internautas opinan de manera positiva sobre el servicio de internet y la otra parte de manera negativa.

En el gráfico 48 de la empresa CNT, resalta los términos 'servicio' e 'internet' con una frecuencia mayor por encima de Claro, determinando que los internautas comentan mucho más sobre el servicio de internet.

Los términos con sentimiento positivo más utilizados por los internautas de manera agrupada entre las empresas son: 'servicio', 'internet', 'claro', 'plan' y 'quiero', de manera que los usuarios realizan un sin número de referencias sobre el servicio de internet. Por otro lado, los términos con sentimiento negativo más utilizados por los internautas de manera agrupada son: 'peor', 'mal', 'problema', 'esperando' y 'lento', es decir que se habla directamente sobre el servicio de internet en las empresas de telefonía móvil, pero con muchas quejas de que el servicio es malo y lento y esperan a que solucionen esos inconvenientes.

De manera general las nubes de palabras ayudan a identificar y relacionar los términos que más utilizan los internautas, pero no brinda una determinación de que telefonía posee un mejor servicio.

- Resultados de los gráficos de pirámides

Los gráficos de pirámides, del gráfico 54 al 59 del capítulo anterior, permitieron relacionar términos comunes por el nivel de frecuencia y separando términos positivos de negativos de las empresas de telefonía móvil.

Referencia pirámide gráfico 54 y gráfico 55 Movistar-CNT Ecuador

Haciendo referencia al gráfico 54, determinamos que la empresa CNT cuenta con términos más frecuentes que Movistar destacándose en términos como 'internet', 'servicio' y 'plan'. Con respecto al gráfico 55 pero en términos negativos entre Movistar y CNT, se destacaron los términos 'problema', 'lento' y 'cambiarme', determinando que son más frecuentes en CNT. Es decir, la empresa CNT Ecuador cuenta con una mayor frecuencia y domina temas sobre el servicio de planes de internet, pero de manera negativa, los usuarios se quejan mucho de los problemas del servicio, determinan que es demasiado lento a diferencia de Movistar el servicio de internet es más rápido y posee menos problemas por lo que los internautas prefieren elegir Movistar por encima de CNT.

Referencia pirámide gráfico 56 y gráfico 57 Claro-CNT Ecuador

Haciendo referencia al gráfico 56, en la comparación de términos comunes entre Claro y CNT, se determinó que CNT cuenta con una mayor frecuencia de términos positivos en base al servicio, planes e internet. Con respecto al gráfico 57 pero en términos negativos entre Claro y CNT, se determinó que CNT brinda un pésimo servicio y presenta muchos problemas, pero no tan lejos de la empresa Claro ya que los internautas consideran que, entre CNT y Claro, la empresa Claro

es la peor a pesar de que no cuenta con la misma cantidad de problemas que CNT o superior a ella.

Referencia pirámide gráfico 58 y gráfico 59 Movistar-Claro

Haciendo referencia al gráfico 58, en la comparación de términos comunes entre Movistar y Claro, se determinó que Claro cuenta con una mayor frecuencia en términos positivos en base al servicio de internet, haciendo mención directamente a la empresa Claro como tal. Con respecto al gráfico 59 pero en términos negativos entre Movistar y Claro, la empresa Claro es la peor a pesar de contar con una gran frecuencia de términos positivos, las críticas negativas son de mayor diferencia que en Movistar. Determinando que entre Movistar y Claro las personas prefieren más Movistar dado que Claro presenta un pésimo servicio y existen muchos reclamos por parte de los internautas que esperan una solución.

4.1.1 Modelos aprobados

Luego de evaluar los modelos en base a los objetivos de la minería de datos para el análisis de sentimientos se ha procedido aprobar el gráfico de pirámides como el óptimo, porque facilita la interpretación de forma comparativa por la frecuencia de términos comunes en grupos de dos, por ejemplo: CNT y Movistar, CNT y Claro, y por último Movistar y Claro.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

En base a los resultados obtenidos del análisis de sentimientos aplicado a los posts y comentarios de usuarios de la red social Facebook, se llegaron a las siguientes conclusiones:

- La herramienta Facepager extrae una gran cantidad de datos, pero con limitantes de información de usuarios como, por ejemplo: la ubicación de la persona que realiza el comentario, siendo este un inconveniente en el caso de querer conocer a detalle la frecuencia de las opiniones siendo estas positivas o negativas por ciudades o regiones del país.

- El uso de la metodología CRISP-DM permitió desarrollar la investigación de manera estructurada y sistemática destacando en ella las características de cada etapa de la minería de texto como técnica de minería de datos.

- Se concluye mediante el gráfico de barras un total de 1010 términos positivos, 358 negativos y 523 términos neutros entre las tres empresas siendo este más útil si se desea conocer la cantidad de términos por sentimiento. A diferencia de las nubes de palabras que no muestra la cantidad sino el término, por ejemplo: la palabra ´servicio´ se repite en las tres telefonías móviles con mayor frecuencia.

- Al analizar las opiniones de los usuarios con respecto al servicio que ofrecen las telefonías móviles a nivel nacional, se determina que, el gráfico más útil para realizar una comparación de los términos comunes más frecuente entre empresas es el de pirámides.
- Se determinó que la empresa Movistar cuenta con el mejor servicio de telefonía móvil a nivel nacional, porque es la que tiene una mayor frecuencia de los términos comunes positivos que negativos.
- A pesar de que la empresa CNT tiene una menor cantidad de comentarios, es la que posee mayor frecuencia en los términos negativos haciendo que tenga el nivel más bajo de las tres empresas consideradas, a diferencia de Claro que se la coloca en un nivel intermedio entre Movistar y CNT.

Recomendaciones

- Realizar evaluaciones periódicas a la información de la red social Facebook de las empresas de telefonía móvil a nivel nacional, mediante el análisis de sentimientos con el fin de mejorar el servicio de telefonía móvil sobre el criterio de los usuarios para tomar a tiempo los correctivos necesarios pensando en la satisfacción y comodidad de los usuarios.
- Utilizar la metodología CRISP-DM en futuras investigaciones relacionadas con análisis de datos que permitan identificar de manera rápida, oportuna y económica las opiniones de los internautas debido a su alto nivel de aplicabilidad en las diferentes redes sociales.
- Se recomienda utilizar la red social Twitter para un análisis más detallado ya que su política de privacidad a la extracción de información no es tan limitada como la red social Facebook.

- Usar los gráficos de pirámides para una mejor comparación de frecuencia de datos entre entidades, porque permite facilitar la comprensión del análisis de sentimientos en base a relaciones comunes.

BIBLIOGRAFÍA

1. Acevedo Miranda Carlos; Clorio Rodriguez Ricardo; Zagal Flores Roberto; García Mendoza Consuelo V. (2014). Arquitectura Web para análisis de sentimientos en Facebook con enfoque semántico. Obtenido de http://www.rcs.cic.ipn.mx/2014_75/Arquitectura%20Web%20para%20analisis%20de%20sentimientos%20en%20Facebook%20con%20enfoque%20semantico.pdf
2. Alcazar, J. P. (22 de febrero de 2017). Ranking Redes Sociales, Sitios Web y Aplicaciones Móviles Ecuador 2017. Obtenido de http://blog.formaciongerencial.com/ranking-redes-sociales-sitios-web-aplicaciones-moviles-ecuador-2017/
3. Arcotel. (13 de septiembre de 2017). Agencia de Regulación y Control de las Telecomunicaciones (Arcotel). Obtenido de http://www.arcotel.gob.ec/arcotel-15055-240-lineas-de-telefonia-celular-existen-en-el-ecuador/
4. Arcotel. (junio de 2018). Agencia de Regulación y Control de las Telecomunicaciones. Obtenido de Servicio Móvil Avanzado: http://www.arcotel.gob.ec/wp-content/uploads/2018/07/1.1.1-Lineas-activas-por-servicio_y_Densidad_May-2018_R.xlsx
5. Arcotel. (junio de 2018). Agencia de Regulación y Control de las Telecomunicaciones (ARCOTEL). Obtenido de Servicio de Telefonía Fija: http://www.arcotel.gob.ec/wp-content/uploads/2018/07/2.1.1-LINEAS-TELEFONICA-POR-TIPO-DE-ACCESO_mayo2018_R.xlsx
6. Becerra, C. M. (16 de junio de 2016). Análisis de sentimiento en Twitter: lo bueno y lo malo. Obtenido de

https://rdu.unc.edu.ar/bitstream/handle/11086/3751/Becerra%202016_analisis-de-sentimiento.pdf?sequence=1&isAllo wed=y

7. Belinchón, Y. (2015). MINERÍA DE DATOS. Obtenido de https://es.scribd.com/document/308398381/15mem-pdf
8. Buenaño, D., & Luján, S. (2016). Repositorio Institucional de la Universidad de Alicante. Obtenido de https://rua.ua.es/dspace/bitstream/10045/61852/1/2016_Buenao_Lujan_Tecnologia-innovacion.pdf
9. Censo. (2010). Instituto Nacional de Estadísticas y Censos (INEC). Obtenido de http://www.ecuadorencifras.gob.ec/resultados/
10. Coria, S. R. (10 de marzo de 2016). ResearchGate. Obtenido de https://www.researchgate.net/publication/266280867_Introduccion_a_la_Mineria_de_Datos_y_el_Data_Warehousing
11. CRISP-DM (2000). The modeling Agency. Obtenido de CRISP-DM 1.0: https://www.the-modeling-agency.com/crisp-dm.pdf
12. ElUniverso. (16 de Julio de 2015). Tuenti es una nueva marca de la operadora Otecel. Obtenido de https://www.eluniverso.com/noticias/2015/07/16/nota/5021320/tuenti-es-nueva-marca-operadora-otecel
13. García, L. (agosto de 2014). Repositorio U Chile. Obtenido de http://repositorio.uchile.cl/bitstream/handle/2250/130479/cf-montesinos_lg.pdf?sequence= 1
14. Intelligent. (19 de Julio de 2017). Análisis de sentimientos, ¿qué es, ¿cómo funciona y para qué sirve? Obtenido de http://www.itelligent.es/es/analisis-de-sentimiento/

15. Jacobo, H. L. (enero de 2016). Análisis automático de opiniones. Obtenido de https://www.gelbukh.com/thesis/Hugo%20Librado%20Jacobo%20-%20MSc.pdf
16. Jácome, P. H. (25 de abril de 2017). Repositorio ESPE. Obtenido de http://repositorio.espe.edu.ec/jspui/bitstream/21000/13003/1/T-ESPE-053863.pdf
17. Jünger, J., & Keyling, T. (2017). Github. Obtenido de https://github.com/strohne/Facepager
18. Microsoft. (05 de mayo de 2018). Obtenido de https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts?view=sql-analysis-services-2017
19. Mikel, N., & Arantza, I. (25 de mayo de 2017). Dyna New Technologies. Obtenido de https://www.dyna-newtech.com/Recursos/Controles/descarga.aspx?IdDocumento=7835&Tipo=1&CodIdioma=&IdWeb=fea752ad-ffaa-40db-a4c7-1ab3b2c7d55a
20. Ministerio de Telecomunicaciones y Seguridad de la Información (2012). Los servicios de telefonía son más inclusivos en el Ecuador. Obtenido de https://www.telecomunicaciones.gob.ec/telefonia-en-el-ecuador/
21. MINTELZ/RLBA. (26 de marzo de 2013). Ministerio de Telecomunicaciones y de la sociedad de la información. Obtenido de https://www.telecomunicaciones.gob.ec/telefonia-en-el-ecuador/
22. Moreno, A. I. (2017). Técnicas estadísticas en Minería de Textos. Obtenido de https://idus.us.es/xmlui/bitstream/handle/11441/63197/Valero%20Moreno%20Ana%20Isabel%20TFG.pdf?sequence=1

23. Narvaez, M. S. (2017). Repositorio ESPE. Obtenido de https://repositorio.espe.edu.ec/bitstream/21000/13528/1/T-ESPE-053887.pdf
24. Piatetsky, G. (2014). Kdnuggets. Obtenido de https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html
25. Rosado, D. M. (agosto de 2016). Repositorio.ug.edu.ec. Obtenido de http://repositorio.ug.edu.ec/bitstream/redug/15703/1/TESIS_DENNYS_ZAMBRANO_MAETEL.pdf
26. Rstudio. (2018). Rstudio. Obtenido de https://www.rstudio.com/products/RStudio/
27. Strohne. (08 de 01 de 2018). Facepager. Obtenido de https://github.com/strohne/Facepager/releases
28. Villena, J. (13 de octubre de 2015). Introducción al análisis de sentimientos (minería de opiniones). Obtenido de https://www.meaningcloud.com/es/blog/introduccion-al-analisis-de-sentimientos-mineria-de-opinion
29. Viteri, R. J. (2016). Repositorio UG. Obtenido de Universidad de Guayaquil: http://repositorio.ug.edu.ec/bitstream/redug/11493/1/PTG-BCISC%20889%20%20VITERI%20ALVARADO%20RICARDO%20JAVIER.pdf

ANEXOS

Anexo 1: Librerías

a. Script de importación de librerías en RStudio.

```
library(plyr)
library(maps)
library(lubridate)
library(dplyr)
library(MASS)
library(stringr)
library(qdap)
library(dendextend)
library(httr)
library(caret)
library(ggplot2)
library(NLP)
library(RColorBrewer)
library(wordcloud)
library(sentimentr)
library(SentimentAnalysis)
library(ROAuth)
library(SnowballC)
library(tm)
library(readr)
library(cluster)
library(slam)
library(Matrix)
library(plotrix)
library(ggthemes)
library(RWeka)
library(tidyverse)
library(tidytext)
library(zoo)
library(scales)
library(syuzhet)
library("nycflights13")
library(lattice)
library(e1071)
library(devtools)
```

Anexo 2: Exploración de datos

a. Script en RStudio para la importación de datos de las empresas de telefonía móvil de la red social Facebook, enero del 2018 hasta mayo del 2018.

```
Datos <- read.csv("Movistar1.csv", sep=";")
Datos2 <- read.csv("Claro1.csv", sep=";")
Datos3 <- read.csv("Cnt1.csv", sep=";")

#revisión de tablas

table(Datos$level)
table(Datos2$level)
table(Datos3$level)
```

Anexo 3: Fase de preprocesamiento

a. Script en RStudio para la limpieza de datos.

```
#PREPROCESAMIENTO MOVISTAR

Movicorpus <- Corpus(VectorSource(Datos$message))
Movicorpus <- tm_map (Movicorpus, content_transformer(tolower))
Movicorpus <- tm_map(Movicorpus, removePunctuation)
Movicorpus <- tm_map(Movicorpus, removeNumbers)

#PREPROCESAMIENTO CLARO

Clarocorpus <- Corpus(VectorSource(Datos2$message))
Clarocorpus <- tm_map (Clarocorpus, content_transformer(tolower))
Clarocorpus <- tm_map(Clarocorpus, removePunctuation)
Clarocorpus <- tm_map(Clarocorpus, removeNumbers)

#PREPROCESAMIENTO CNT

Cntcorpus <- Corpus(VectorSource(Datos3$message))
Cntcorpus <- tm_map (Cntcorpus, content_transformer(tolower))
Cntcorpus <- tm_map(Cntcorpus, removePunctuation)
Cntcorpus <- tm_map(Cntcorpus, removeNumbers)
```

```
#stopwords de ingles, espaÃ±ol MOVISTAR

Movicorpus <- tm_map(Movicorpus, removeWords, c (stopwords("english"), ("my")))
Movicorpus <- tm_map(Movicorpus, removeWords, c (stopwords("spanish"), ("si"), ("porque")))

#stopwords de ingles, espaÃ±ol CLARO

Clarocorpus <- tm_map(Clarocorpus, removeWords, c (stopwords("english"), ("my")))
Clarocorpus <- tm_map(Clarocorpus, removeWords, c (stopwords("spanish"), ("si"), ("porque")))

#stopwords de ingles, espaÃ±ol CNT

Cntcorpus <- tm_map(Cntcorpus, removeWords, c (stopwords("english"), ("my")))
Cntcorpus <- tm_map(Cntcorpus, removeWords, c (stopwords("spanish"), ("si"), ("porque")))

#REMOVEDOR DE URL FUNCION

removeURL <- function(x) gsub("http[[:alnum:]]*", "", x)
#MOVISTAR
Movicorpus <- tm_map(Movicorpus, content_transformer(removeURL))
#CLARO
Clarocorpus <- tm_map(Clarocorpus, content_transformer(removeURL))
#CNT
Cntcorpus <- tm_map(Cntcorpus, content_transformer(removeURL))

#Espacios adicionales entre los textos
#MOVISTAR
Movicorpus <- tm_map(Movicorpus, stripWhitespace)
#CLARO
Clarocorpus <- tm_map(Clarocorpus, stripWhitespace)
#CNT
Cntcorpus <- tm_map(Cntcorpus, stripWhitespace)
```

b. Script en RStudio para transformación de código ASCII a UTF-8.

```
#MOVISTAR
Movicorpus <- tm_map(Movicorpus, iconv, from = "utf-8", to = "ASCII//TRANSLIT")
Movicorpus <- tm_map(Movicorpus, removePunctuation)
#CLARO
Clarocorpus <- tm_map(Clarocorpus, iconv, from = "utf-8", to = "ASCII//TRANSLIT")
Clarocorpus <- tm_map(Clarocorpus, removePunctuation)
#CNT
Cntcorpus <- tm_map(Cntcorpus, iconv, from = "utf-8", to = "ASCII//TRANSLIT")
Cntcorpus <- tm_map(Cntcorpus, removePunctuation)
```

Anexo 4: Fase de clasificación

a. Script en RStudio para la creación de matriz de documentos.

```
#Matrices de terminos Movistar
frequenciesMovi <- DocumentTermMatrix(Movicorpus)
#Matrices de terminos Claro
frequenciesClaro <- DocumentTermMatrix(Clarocorpus)
#Matrices de terminos Cnt
frequenciesCnt <- DocumentTermMatrix(Cntcorpus)
```

b. Script en RStudio para la reducción de términos.

```
#Reducir las palabras que se repiten muy poco o son poco frecuentes
sparseMovi <- removeSparseTerms(frequenciesMovi, 0.999)
sparseClaro <- removeSparseTerms(frequenciesClaro, 0.999)
sparseCnt <- removeSparseTerms(frequenciesCnt, 0.999)
```

c. Script en RStudio para retornar base de datos y suma de términos.

```
# retornar la variable sparse como una base de datos en formato R
Movicorpus <- as.data.frame(as.matrix(sparseMovi))
Clarocorpus <- as.data.frame(as.matrix(sparseClaro))
Cntcorpus <- as.data.frame(as.matrix(sparseCnt))

#suma las veces que se repite una palabra
frequenciesMovi <- colSums(Movicorpus)
frequenciesClaro <- colSums(Clarocorpus)
frequenciesCnt <- colSums(Cntcorpus)

#ordena de manera decreciente
frequenciesMovi <- sort(frequenciesMovi, decreasing = T)
frequenciesClaro <- sort(frequenciesClaro, decreasing = T)
frequenciesCnt <- sort(frequenciesCnt, decreasing = T)
```

d. Script en RStudio para la separación de variables en nueva base de datos en formato R.

```
#SEPARA FRECUENCIA MOVISTAR
EcMovi <- data.frame(word = names(frequenciesMovi),freq=frequenciesMovi)
#SEPARA FRECUENCIA CLARO
EcClaro <- data.frame(word = names(frequenciesClaro),freq=frequenciesClaro)
#SEPARA FRECUENCIA CNT
EcCnt <- data.frame(word = names(frequenciesCnt),freq=frequenciesCnt)
```

Anexo 5: Diccionario de datos

a. Script en RStudio para importar diccionario de datos, función merge y renombre de variables.

```
##DICCIONARIO DE TÉRMINOS TRADUCIDO
espanish <- read.csv("diccionarioespañol.csv", sep=";")

#Función merge()

EcMovi = merge(EcMovi, espanish)
EcClaro = merge(EcClaro, espanish)
EcCnt = merge(EcCnt, espanish)

#CAMBIAR NOMBRE DE VARIABLES
#MOVISTAR
colnames (EcMovi) [colnames (EcMovi) == "sentiment"] <- "sentimentMovi"
colnames (EcMovi) [colnames (EcMovi) == "freq"] <- "freqMovi"
#CLARO
colnames (EcClaro) [colnames (EcClaro) == "sentiment"] <- "sentimentClaro"
colnames (EcClaro) [colnames (EcClaro) == "freq"] <- "freqClaro"
#CNT
colnames (EcCnt) [colnames (EcCnt) == "sentiment"] <- "sentimentCnt"
colnames (EcCnt) [colnames (EcCnt) == "freq"] <- "freqCnt"
```

b. Script en RStudio para reemplazo cadena de caracteres

```
#MOVISTAR
EcMovi$sentimentMovi <- factor(EcMovi$sentimentMovi, levels = c("positivo","negativo","neutro"))
EcMovi$sentimentMovi[is.na(EcMovi$sentimentMovi)] <- "neutro"
#CLARO
EcClaro$sentimentClaro <- factor(EcClaro$sentimentClaro, levels = c("positivo","negativo","neutro"))
EcClaro$sentimentClaro[is.na(EcClaro$sentimentClaro)] <- "neutro"
#CNT
EcCnt$sentimentCnt <- factor(EcCnt$sentimentCnt, levels = c("positivo","negativo","neutro"))
EcCnt$sentimentCnt[is.na(EcCnt$sentimentCnt)] <- "neutro"
```

Anexo 6: Modelado

a. Script en RStudio para la creación y visualización de gráfico de barras.

```
#GRÁFICO DE BARRAS MOVISTAR
ggplot(EcMovi, aes(x = sentimentMovi, fill = Etiquetas)) +
  geom_histogram(aes(fill = sentimentMovi),stat = "count") +
  xlab("Sentimiento") + ylab("Cantidad") +
  labs(title = "Gráfico de barras Movistar",
       subtitle = "Cantidad de términos divididos por sentimiento",
       caption = "Datos obtenidos de Facebook")
##############################################
#GRÁFICO DE BARRAS CLARO
ggplot(EcClaro, aes(x = sentimentClaro, fill = Etiquetas)) +
  geom_histogram(aes(fill = sentimentClaro),stat = "count") +
  xlab("Sentimiento") + ylab("Cantidad") +
  labs(title = "Gráfico de barras Claro",
       subtitle = "Cantidad de términos divididos por sentimiento",
       caption = "Datos obtenidos de Facebook")
##############################################
#GRÁFICO DE BARRAS CNT
ggplot(EcCnt, aes(x = sentimentCnt, fill = Etiquetas)) +
  geom_histogram(aes(fill = sentimentCnt),stat = "count") +
  xlab("Sentimiento") + ylab("Cantidad") +
  labs(title = "Gráfico de barras CNT Ecuador",
       subtitle = "Cantidad de términos divididos por sentimiento",
       caption = "Datos obtenidos de Facebook")
```

b. Script en RStudio para la creación y visualización de gráfico de barras por términos más frecuentes.

```
EcMovi <- EcMovi[order(EcMovi[, 2],decreasing = T),]
#DIEZ TÉRMINOS MAS FRECUENTES DE MOVISTAR
EcMovi[1:10, ] %>%
  ggplot(aes(word, freqMovi)) +
  geom_bar(stat = "identity", color = "black", fill = "dodgerblue4") +
  geom_text(color = "floralwhite",aes(hjust = 1.3, label = sentimentMovi)) +
  coord_flip() +
  labs(title = "Diez términos más frecuentes de Movistar",
       x = "Términos", y = "Frecuencia de uso")

EcClaro <- EcClaro[order(EcClaro[, 2],decreasing = T),]
#DIEZ TÉRMINOS MAS FRECUENTES DE CLARO
EcClaro[1:10, ] %>%
  ggplot(aes(word, freqClaro)) +
  geom_bar(stat = "identity", color = "black", fill = "dodgerblue4") +
  geom_text(color = "floralwhite",aes(hjust = 1.3, label = sentimentClaro)) +
  coord_flip() +
  labs(title = "Diez términos más frecuentes de Claro",
       x = "Términos", y = "Frecuencia de uso")

EcCnt <- EcCnt[order(EcCnt[, 2],decreasing = T),]
#DIEZ TÉRMINOS MAS FRECUENTES DE CNT
EcCnt[1:10, ] %>%
  ggplot(aes(word, freqCnt)) +
  geom_bar(stat = "identity", color = "black", fill = "dodgerblue4") +
  geom_text(color = "floralwhite",aes(hjust = 1.3, label = sentimentCnt)) +
  coord_flip() +
  labs(title = "Diez términos más frecuentes de CNT Ecuador",
       x = "Términos", y = "Frecuencia de uso")
```

c. Script en RStudio para la creación y visualización de gráfico de barras por términos más frecuentes por porcentaje.

```
EcClaro %>%
  mutate(perc = (freqClaro/sum(freqClaro))*100) %>%
  .[1:10, ] %>%
  ggplot(aes(word, perc)) +
  geom_bar(stat = "identity", color = "black", fill = "darkcyan") +
  geom_text(color = "floralwhite",aes(hjust = 1.3, label = round(perc, 2))) +
  coord_flip() +
  labs(title = "Diez términos más frecuentes de Claro",
       x = "Términos", y = "Porcentaje de uso")
```

d. Script en RStudio para crear y visualizar nubes de palabras.

```
#NUBE DE MOVISTAR
wordcloud(EcMovi$word, max.words = 100,EcMovi$freqMovi,
          random.order = F, colors = brewer.pal(name = "Dark2", n = 8))

#NUBE DE CLARO
wordcloud(EcClaro$word, max.words = 100,EcClaro$freqClaro,
          random.order = F, colors = brewer.pal(name = "PuOr", n = 10))

#NUBE DE CNT
wordcloud(EcCnt$word, max.words = 100,EcCnt$freqCnt,
          random.order = F, colors = brewer.pal(name = "BrBG", n = 11))
```

e. Script en RStudio para la creación de base de datos positivas y negativas.

```
#BASE DE DATOS DE PALABRAS POSITIVAS
Movipositivo=subset(EcMovi,sentimentMovi=="positivo")
Claropositivo=subset(EcClaro,sentimentClaro=="positivo")
Cntpositivo=subset(EcCnt,sentimentCnt=="positivo")
PositivosCC =Reduce(merge, list(Movipositivo, Claropositivo, Cntpositivo))
difference <- abs(PositivosCC[, 2] - PositivosCC[,4] - PositivosCC[,6])
PositivosCC <- cbind(PositivosCC, difference)
PositivosCC <- PositivosCC[order(PositivosCC[, 8],decreasing = T), ]

#BASE DE DATOS DE PALABRAS NEGATIVAS
Movinegativo=subset(EcMovi,sentimentMovi=="negativo")
Claronegativo=subset(EcClaro,sentimentClaro=="negativo")
Cntnegativo=subset(EcCnt,sentimentCnt=="negativo")
NegativosCC =Reduce(merge, list(Movinegativo, Claronegativo, Cntnegativo))
difference <- abs(NegativosCC[, 2] - NegativosCC[,4] - NegativosCC[,6])
NegativosCC <- cbind(NegativosCC, difference)
NegativosCC <- NegativosCC[order(NegativosCC[, 8],decreasing = T), ]
```

f. Script en RStudio para creación y visualización de gráfico de pirámides de términos positivos, misma función para las diferentes empresas solo cambio de variables.

```
Piramide1 = merge(Movipositivo,Claropositivo)
head(Piramide1)

difference <- abs(Piramide1[, 2] - Piramide1[,4])
Piramide1 <- cbind(Piramide1, difference)

Piramide1 <- Piramide1[order(Piramide1[, 6],decreasing = T), ]

#DATA FRAME DE LOS PRIMEROS 10 TÉRMINOS
Piramide1 <- data.frame(x = Piramide1[1:10, 2],
                        y = Piramide1[1:10, 4],
                        labels = Piramide1[1:10,1])

pyramid.plot(Piramide1$x, Piramide1$y,labels = Piramide1$labels,
             main = "Términos positivos en común",gap = 100,
             laxlab = NULL,raxlab = NULL, unit = NULL,
             top.labels = c("Movistar","Términos","Claro"))
```

g. Script en RStudio para creación y visualización de gráfico de pirámides de términos negativos, misma función para las diferentes empresas solo cambio de variables.

```
Piramide6 = merge(Claronegativo,Cntnegativo)

difference <- abs(Piramide6[, 2] - Piramide6[,4])
Piramide6 <- cbind(Piramide6, difference)
Piramide6 <- Piramide6[order(Piramide6[, 6],decreasing = T), ]

#DATA FRAME DE LOS PRIMEROS 10 TÉRMINOS
Piramide6 <- data.frame(x = Piramide6[1:10, 2],
                        y = Piramide6[1:10, 4],
                        labels = Piramide6[1:10,1])

pyramid.plot(Piramide6$x, Piramide6$y,labels = Piramide6$labels,
             main = "Términos negativos en común",gap = 30,
             laxlab = NULL,raxlab = NULL, unit = NULL,
             top.labels = c("Claro","Términos","CNT Ecuador"))
```