

APLICACIÓN DE MINERÍA DE TEXTO PARA EL ANÁLISIS DE SENTIMIENTOS DEL SERVICIO DE TELEFONÍA MÓVIL EN EL ECUADOR

F. J. R. ZAMBRANO¹, B. F. V. FLORES², W. J. P. MENDOZA³, R. A. R. ZAMBRANO⁴, S. M. C. RIVADENEIRA⁵
Universidad Laica Eloy Alfaro de Manabí^{1,2,3}, Unidad Educativa Julio Pierregrosse⁵, Universidad Técnica de Manabí⁴

ORCID ID: <http://orcid.org/0000-0002-5384-7243>
e1311434961@live.uleam.edu.ec¹

Submetido em 29/11/2018 - Aceito em 18/12/2020

DOI: 10.15628/holos.2020.7994

RESUMO

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 el análisis de sentimientos, implementando la metodología CRISP-DM para la extracción y manipulación de la información

generando reportes estadísticos. En base a las comparaciones realizadas de los términos comunes más frecuentes entre empresas, se determinó que la empresa Movistar cuenta con el mejor servicio de telefonía móvil a nivel nacional porque obtiene un 17% en términos negativos, siendo el porcentaje más bajo generado por parte de las opiniones de los usuarios seguidores, dando una ventaja considerable con respecto a los altos porcentajes negativos de las otras empresas.

PALAVRAS-CHAVE: Análisis de sentimientos, CRISP-DM, minería de datos, minería de texto, telefonía móvil.

APPLICATION OF TEXT MINING FOR THE ANALYSIS OF FEELINGS OF THE MOBILE TELEPHONE SERVICE IN ECUADOR

ABSTRACT

This project applies text mining techniques to the extraction of comments on the social network Facebook of the official profiles of the companies Movistar, Claro and CNT at a national level, from January to May 2018, on the mobile telephone service, with the purpose of analyzing the positive, negative or neutral criteria expressed by Internet users through the analysis of feelings, implementing the CRISP-DM methodology for the extraction and manipulation of information,

generating statistical reports. Based on the comparisons made of the most frequency common terms among companies, it was determined that the Movistar company has the best mobile telephone service at national level because it obtains 17% in negative terms, being the lowest percentage generated by the opinions of the followers' users, giving a considerable advantage with respect to the high negative percentages of the other companies.

KEYWORDS: Feelings analysis, CRISP-DM, data mining, text mining, mobile telephony.



1 INTRODUCCION

El análisis de sentimientos o *sentiment analysis* es el estudio donde 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.

El presente trabajo comprende el estudio del análisis de posts y comentarios de la red social Facebook sobre el servicio de telefonía móvil de las empresas Movistar, Claro y CNT de la República del Ecuador, aplicando técnicas de minería de texto (*text mining*), con un aproximado de 14.000 opiniones públicas, que fueron extraídos por medio de la herramienta Facepager durante el periodo enero del 2018 hasta mayo del 2018, debido a su alto nivel de usabilidad y dependencia de los usuarios al momento de interactuar y de emitir comentarios que son públicos en torno al ámbito del servicio de telefonía móvil. Estos posts y comentarios no se encuentran estructurados ni semiestructurados por lo cual, realizando el análisis de datos oportuno se genera información sobre los comentarios en dicha red social, generando una oportunidad de indagar sobre el comportamiento de la opinión de los usuarios a cerca de las diferentes empresas de telefonía móvil, siendo un pilar de mejora en el crecimiento o desuso del servicio para los clientes y una visión a las empresas de telefonía móvil sobre las 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 por parte de alguna empresa de 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 a conocer sus productos o servicios.

Por esta razón se planteó el estudio y análisis del presente trabajo, aplicando las fases de la metodología CRISP-DM hasta la evaluación de resultados, siendo una metodología estándar, estructurada, con fases bidimensionales y específica para análisis de datos, mediante el cual se analiza y extrae la información más relevante al aplicar técnicas de minería de texto y modelos estadísticos como los gráficos de pirámides que permitan determinar cuál es el modelo más eficaz para el análisis de sentimientos y determinar qué empresa de telefonía móvil posee o tiene mejores criterios en su estatus sobre su servicio a nivel nacional.

A continuación, el presente artículo se encuentra organizado de la siguiente manera: en la sección 2 se establecen conceptos y trabajos referentes a este artículo, donde se analizaron revisiones de otros autores sobre el estudio de comentarios públicos en redes sociales, resaltando la importancia de aplicar técnicas de minería de texto y la forma en la que ayuda a la toma de decisiones en base al análisis de sentimientos, dando paso al concepto principal de minería de texto y sus etapas, en la sección 3 se establece la metodología aplicada CRISP-DM, siendo específica para el análisis de datos en el concepto de minería de datos y las fases que lo corresponden, en la sección 4 se presentan los resultados en base a los reportes generados de los modelos estadísticos que se utilizaron en la sección 3, como último punto en la sección 5 se presentan las conclusiones y recomendaciones del presente artículo.



2 MARCO TEÓRICO

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 información de manera automática y de forma continua, las cuales pueden tener estructuras creadas por bases 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 donde se puede adquirir lo más relevante mediante el análisis de datos. Existe una gran complejidad para realizar 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, considerada como una herramienta útil para el análisis de grandes volúmenes de información.

2.1 Análisis de opiniones en redes sociales

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 formularios en papel, entrevistas y encuestas hasta formularios electrónicos. El problema se resuelve mediante una metodología y se desarrolla una aplicación que permite el análisis de textos cortos de opinión, clasificándolos por sentimiento positivo, negativo y sin opinión o sentimiento.

2.2 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 de 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 el cual aplicando varias técnicas de minería determinaron 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.

Por otro lado, las técnicas de minería de datos se clasifican en supervisados y no supervisados. En la tabla 1 se muestra la clasificación, siendo supervisados algoritmos que tienen

entradas de datos cambiantes por lo que el resultado va en función de dichas entradas, mientras que los no supervisados son algoritmos que no necesitan de la intervención humana para interpretar resultados, es decir que 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.

Clasificación	Técnicas
Supervisados	Red neuronal, árboles de decisión, regresión.
No supervisados	Agrupamiento ("clustering"), reglas de asociación, minería de texto ("text mining").

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).

2.3.1 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.

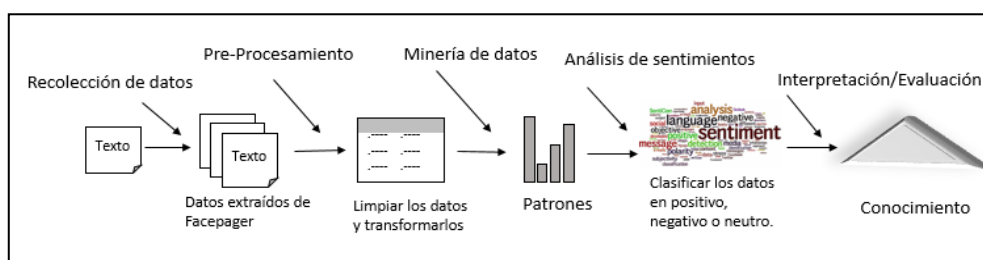


Figura 1: Etapas del proceso de minería de texto

En la figura 1, se define el proceso de minería de 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 los perfiles oficiales de las empresas telefónicas Movistar, Claro y CNT Ecuador, usando la herramienta Facepager.

- Preprocesamiento, en esta etapa los textos se transforman en un 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 simetrías 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.

2.3.2 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 las siguientes:

- Sistema preestablecido reelaborado. Los pre-ajustes básicos se actualizan para reflejar los cambios en la API de Facebook.
- 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.
- Corrección de errores y mejoras menores.

2.3.3 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, soporta procedimientos y técnicas requeridas para análisis de calidad (RStudio, 2018).

2.3.4 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 & Luján, 2016).

Según (Piatetsky, 2014) en base a datos de la última encuesta realizada, 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. Siendo una metodología favorable para proyectos de minería de datos, 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.

3 METODOLOGÍA

Tomando como referencia el enfoque teórico de la metodología CRISP-DM, se tiene a considerar el desarrollo para el presente proyecto aplicar cinco de las seis fases de la metodología, siendo estas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado y evaluación de resultados.

3.1 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 desde una perspectiva no técnica.

Facebook es un medio, donde las empresas de telefonía móvil 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 Ecuador.

Con respecto a las empresas de telecomunicación móvil en el mercado ecuatoriano: Movistar, Claro y CNT, hasta el mes de mayo del 2018 se puede observar que la empresa Movistar tiene el 30,32% de líneas activas, Claro 53,17% y CNT Ecuador consta con el 16,51%, 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 la figura 2.





Figura 2: Posicionamiento de las diferentes telefonías móviles en el mercado,

FUENTE: (ARCOTEL, 2018)

Dentro de los objetivos del negocio está en garantizar la masificación de las telefonías móviles, permitiendo mejorar la infraestructura de telecomunicaciones para brindar productos y servicios de la más avanzada tecnología para ello se establecieron los siguientes objetivos:

- Clasificar las opiniones de los usuarios seguidores en términos positivos, negativos y neutros en base al servicio que ofrecen las telefonías móviles anteriormente mencionadas.
- Realizar comparación de términos comunes positivos y negativos con mayor frecuencia existentes en las empresas telefónicas.

3.2 Comprensión de los datos

En esta fase de la metodología 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 de telefonía móvil y su popularidad en la red social Facebook.

La recolección inicial de datos se realiza mediante la herramienta Facepager que tiene la función de recolectar todos los posts (publicaciones) y comentarios de personas que han participado en la red social Facebook de los perfiles oficiales @MovistarECU, @claroecu, @CNTEcuador. Recolección que tuvo un aproximado de 14.000 opiniones públicas en el periodo de enero del 2018 hasta mayo del 2018 tal como se muestra en la tabla 2 en detalle.

Tabla 2: Datos obtenidos de Facepager.

Total de posts y comentarios: (enero 2018 – mayo 2018)			
Entradas	Movistar	Claro	Cnt
Publicaciones	41	278	475
Comentarios	2721	4022	7867

Total	2762	4300	8342
-------	------	------	------

3.3 Preparación de los datos

Esta fase se concentra en la selección de los datos para el análisis de sentimientos, los cuales serán utilizados en la aplicación de técnicas de minería de texto. Luego se realiza la respectiva limpieza, construcción si es necesario de los datos utilizando la herramienta RStudio, finalmente se integra y da paso a la fase de modelado en base al estudio realizado.

Para la selección de los datos se utilizó la base de comentarios de cada empresa de telefonía móvil, que se encuentra almacenados en archivos csv por empresa, que fueron generados por medio de la herramienta Facepacer mencionada en la sección 2.

En la figura 3 se muestra el contenido de cada archivo csv, tomando como ejemplo a la empresa Claro, donde se identifican los campos: ID personal, nivel de nodo, tipo de objeto, entre otros.

	A	B	C	D	E	F	G	H	I	J
1	id	parent_id	level	object_id	object_query_status	query_time	query_type	message	created_time	
2	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	
3	2983	2	1	207588849262306_1_data	fetched (200)	2018-06-11 18:52:07.8	Facebook:page/posts	¡Hoy nos visitó Mike Bah!	2018-05-31T00:18:27+0000	
4	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	
5	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	
6	3274	2984	2	1861010243920150_data	fetched (200)	2018-06-11 19:09:03.2	Facebook:post/comment	Solo claro te roba el sald	2018-06-01T12:18:27+0000	
7	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	
8	2985	2	1	207588849262306_1_data	fetched (200)	2018-06-11 18:52:07.8	Facebook:page/posts	¿Te ha pasado? ¡Cuéntan	2018-05-30T23:30:00+0000	
9	3277	2985	2	1840336719320836_data	fetched (200)	2018-06-11 19:09:03.5	Facebook:post/comments		2018-05-30T23:52:56+0000	
10	3278	2985	2	1840336719320836_data	fetched (200)	2018-06-11 19:09:03.5	Facebook:post/comment	??????	2018-05-31T04:08:37+0000	

Figura 3: Datos de Facepacer de la empresa de telefonía móvil Claro

Para la selección de los datos se utilizó la base de comentarios de cada empresa de telefonía móvil, que se encuentra almacenados en archivos csv por empresa, que fueron generados por medio de la herramienta Facepacer mencionada en la sección 2.

En la figura 3 se muestra el contenido de cada archivo csv, tomando como ejemplo a la empresa Claro, donde se identifican los campos: ID personal, nivel de nodo, tipo de objeto, entre otros.

La etapa del preprocesamiento también conocido como limpieza de datos permite transformar los datos y separar o quitar caracteres innecesarios. Siguiendo el proceso tal como se muestra en la figura 4 de los scripts para la limpieza del texto donde se remueve todo tipo de caracteres, utilizando funciones como *content_transformer*, *removePunctuation*, *removeNumbers*, entre otros. Seguido de ello se procede a la transformación del código ASCII a UTF-8 porque existen comentarios que no pueden ser interpretados de manera correcta por el tipo de formato en RStudio, luego se repite el proceso de remover signos de puntuación y culmina con el texto limpio para dar paso a la etapa de clasificación.

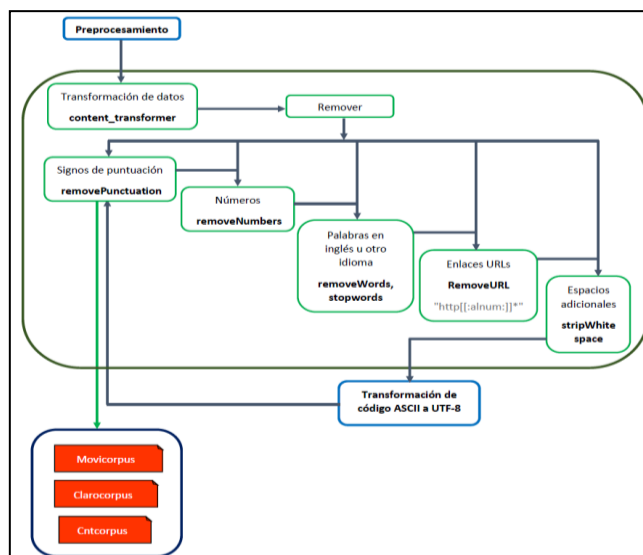


Figura 4: Funciones del preprocesamiento de texto

Para la etapa de clasificación dividida por funciones se sigue una secuencia ordenada de subprocesos como se muestra en la figura 5, el cual empieza creando la matriz de documentos y eliminando los términos poco frecuentes con la función *removeSparseTerms*, luego suma las veces que se repiten cada término aplicando la función *colSums* por cada documento, mostrando como salida la suma de la frecuencia de los términos más importantes, ordenados de manera descendente en un nuevo data frame por empresa, separando en dos variables el término de la frecuencia.

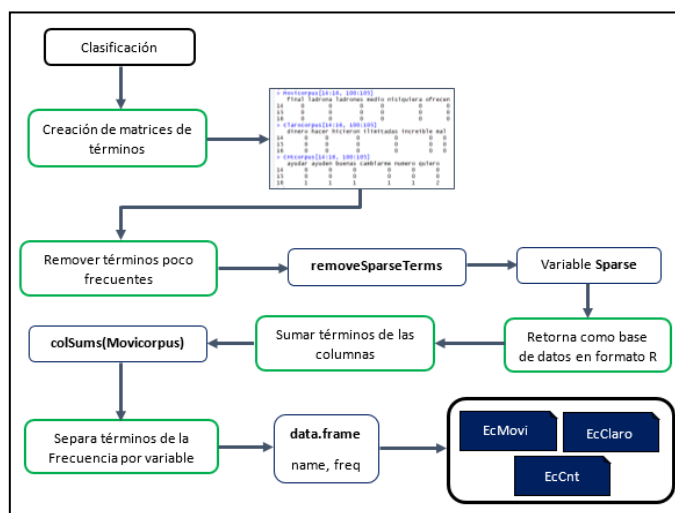


Figura 5: Funciones de la clasificación de términos

En la figura 6, se muestran las diez palabras más frecuentes por empresa de telefonía móvil siendo estas: *frecuenciasMovi* (Movistar), *frecuenciasClaro* (Claro), *frecuenciasCnt* (CNT Ecuador).

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



Figura 6: Diez palabras más frecuentes por empresa

3.4 Modelado

En esta fase de la metodología CRISP-DM se evalúan las técnicas de minería más apropiadas para el procesamiento del texto en base al estudio realizado sobre el servicio que ofrecen las diferentes empresas de telefonía móvil: Movistar, Claro y CNT de la República del Ecuador, donde se establecerá si el sentimiento es positivo, negativo o neutro en base al uso de un diccionario de datos denominado léxico *bing* que contiene una serie de términos etiquetados por sentimiento, luego se procede a generar los reportes de modelos estadísticos y evaluar los resultados para determinar qué empresa posee un mejor servicio en base a la opinión de las experiencias de los usuarios sobre el servicio de telefonía móvil en la red social Facebook.

Para la construcción del modelo se basó en los siguientes reportes estadísticos:

3.4.1 Reportes de términos en gráfico de barras

En la figura 7 se muestra el gráfico de barras de las empresas de telefonía móvil: Movistar, Claro y CNT Ecuador, donde el eje horizontal representa las clases de sentimiento y el eje vertical el porcentaje de términos. Donde se determina que el sentimiento positivo de la empresa Movistar es equivalente al 32%, Claro con un equivalente del 39% y CNT Ecuador con el 29%, siendo Claro la empresa con un mayor porcentaje en términos positivos.

Por el lado del sentimiento negativo la empresa Movistar cuenta con un equivalente del 38%, mismo porcentaje que la empresa Claro y la empresa CNT Ecuador con un 24%, siendo Claro y Movistar las empresas con el mismo nivel de porcentaje elevado en torno al ámbito negativo.

Respecto al sentimiento neutro la empresa Movistar cuenta con un equivalente del 28%, Claro con el 42% y CNT Ecuador con el 30%, siendo Claro la empresa con el porcentaje más elevado en torno a términos neutros.

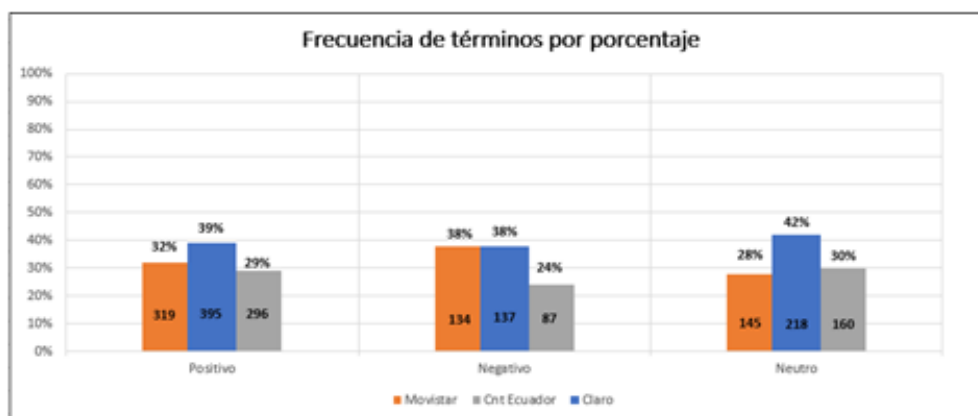


Figura 7: Gráfico de frecuencia de términos por porcentaje

3.4.2 Reportes de términos en pirámides

El gráfico de pirámides permite realizar comparaciones entre dos empresas, en este caso comparar términos comunes y su frecuencia. Donde el eje horizontal representa el porcentaje de frecuencia y el eje vertical los términos más comunes.

En la figura 8, se observa del lado izquierdo Movistar y CNT Ecuador del lado derecho, donde el volumen de barras depende de la frecuencia que tenga cada palabra, destacando los términos positivos 'internet', 'servicio' y 'plan', siendo estos más frecuentes en CNT que en Movistar es decir que entre las dos empresas las personas prefieren interactuar más sobre este tema en CNT referente a los términos positivos. En el caso de los términos negativos se destacan 'lento', 'problemas' y 'mal', siendo estos más frecuentes en CNT perjudicando a la empresa, relacionando los términos positivos y negativos de ambas empresas se concluye que las personas interactúan mucho en CNT, pero la mayor parte de manera negativa sobre el servicio de telefonía móvil siendo este lento, con muchos problemas y malo para el uso.

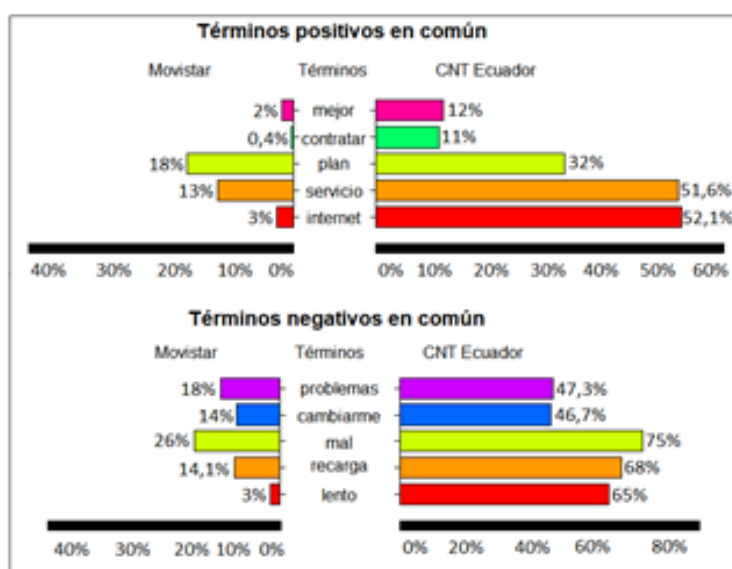


Figura 8: Gráfico de pirámides de términos comunes positivos y negativos Movistar y Cnt

En la figura 9, se observa del lado izquierdo Claro y CNT Ecuador del lado derecho, donde el volumen de barras depende de la frecuencia que tenga cada palabra, destacando los términos positivos 'internet', 'servicio' y 'plan', siendo estos más frecuentes en CNT que en Claro es decir que entre las dos empresas las personas prefieren interactuar más sobre este tema en CNT referente a los términos positivos. En el caso de los términos negativos se destacan 'peor', 'problema' y 'mal', siendo estos más frecuentes en CNT perjudicando a la empresa, relacionando los términos positivos y negativos de ambas empresas. Se concluye que las personas interactúan más en CNT que en Claro, pero la mayor parte de manera negativa sobre el servicio de telefonía móvil siendo este lento, con muchos problemas y malo para el uso.

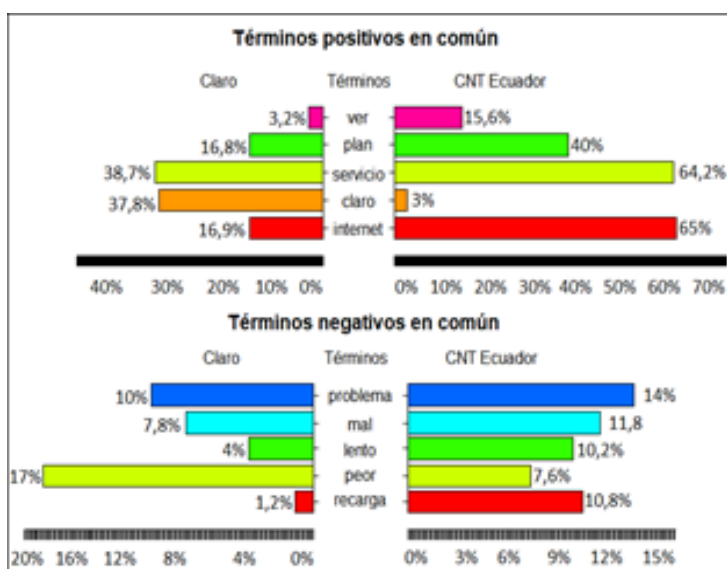


Figura 9: Gráfico de pirâmides de términos comunes positivos y negativos Claro y Cnt

En la figura 10, se observa del lado izquierdo Movistar y Claro del lado derecho, donde el volumen de barras depende de la frecuencia que tenga cada palabra, destacando los términos positivos 'internet', 'servicio' y 'plan', siendo estos más frecuentes en Claro a excepción del término 'plan' que es más frecuente en Movistar, es decir que las personas interactúan más sobre el servicio de internet en Claro pero Movistar se destaca sobre los planes de telefonía móvil referente a los términos positivos. En el caso de los términos negativos se destacan 'peor', 'esperando' y 'pésima', siendo estos más frecuentes en Claro perjudicando a la empresa. Relacionando los términos positivos y negativos de ambas empresas, se concluye que las personas interactúan más en Claro sobre el servicio de internet, pero de una manera negativa, lo que en Movistar no es muy frecuente con respecto al uso de términos negativos.

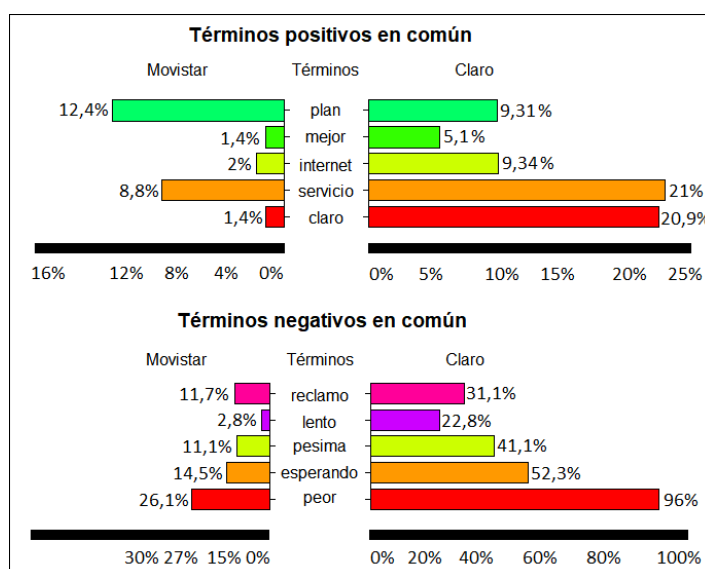


Figura 10 – Gráfico de pirâmides de términos comunes positivos y negativos Movistar y Claro



3.5 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 la fase de comprensión del negocio.

Se clasificó las opiniones según el objetivo 1 de los usuarios seguidores, en términos positivos, negativos y neutros en base al servicio que ofrecen las telefonías móviles Movistar, Claro y CNT Ecuador, haciendo uso de los gráficos de barras generales.

Tabla 3 – Resultados de los gráficos de barras.

Porcentaje de términos catalogados por sentimientos			
Empresas	Positivo	Negativo	Neutro
Movistar	32%	38%	28%
Claro	39%	38%	42%
CNT	29%	24%	30%

Como se puede observar en la tabla 3, se define el porcentaje total de términos catalogados por sentimiento, 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 un total de 39% de términos positivos. A su vez Claro también se destaca en la cantidad de términos negativos con un total de 38% 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 32% de términos positivos, mientras que CNT cuenta con un total de 29% términos positivos quedando 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.

Según el objetivo 2, se realizó una comparación de los términos comunes positivos y negativos con mayor frecuencia existente en las empresas telefónicas Movistar, Claro y CNT Ecuador, mediante los gráficos de pirámides de las figuras 8, 9 y 10.

Tabla 4 – Porcentaje total de términos catalogados positivos.

Porcentaje total de términos catalogados positivos			
Términos positivos	Movistar	Claro	CNT
servicio	291	705	1170
internet	66	308	1183
plan	410	307	728

Total	15%	25%	60%
-------	-----	-----	-----

Tabla 5 – Porcentaje total de términos catalogados negativos.

Porcentaje total de términos catalogados positivos			
Términos positivos	Movistar	Claro	CNT
problema	89	103	144
Lento	5	41	105
esperando	26	94	86
Total	17%	35%	45%

Respecto a las tablas 4 y 5, se seleccionó tres términos positivos y tres negativos más frecuente en las empresas Movistar, Claro y CNT donde los usuarios opinan que el servicio de Movistar es el mejor considerando que cuenta con un 17% en el índice de términos negativos y un 15% en términos positivos, mientras que el servicio de internet y planes de telefonía es mencionado más frecuente en CNT pero de una manera negativa con un porcentaje de 48% de términos negativos y un 60% en términos positivos.

4 CONCLUSIONES

En base a los resultados obtenidos del análisis de sentimientos aplicado a los posts y comentarios de usuarios del servicio móvil en la red social Facebook, se evidencia con la rapidez que se pudo extraer una gran cantidad de datos para luego procesarla y convertirla en información, evidentemente existe limitante de información del usuario 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 opiniones por ciudades o regiones del país. Por lo que, mediante el análisis de sentimientos se determinó que la empresa Movistar posee un 32% de sentimiento positivo, Claro cuenta con un 39% y CNT Ecuador con el 29%, siendo Claro la empresa con un mayor porcentaje de términos positivos, mientras que el sentimiento negativo de Movistar y Claro son equivalentes al 38% y de CNT Ecuador de un 24%. Otra de las formas que se analizaron los datos, fue utilizando los términos más frecuentes extraídos mediante gráfico de pirámides, en el cual se realizó una comparación de los cinco términos comunes más frecuentes por sentimiento entre empresas y se concluyó que la empresa Movistar posee una alta frecuencia en planes de telefonía como término positivo, mientras que en Claro y CNT se menciona el servicio de internet con una alta frecuencia en términos positivos, mientras que los planes de telefonía se menciona con una frecuencia alta, en términos negativos.

5 REFERENCIAS

- 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
- 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
- BELINCHÓN, Y. (2015). MINERÍA DE DATOS. Obtenido de <https://es.scribd.com/document/308398381/15mem-pdf>
- 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
- JACOBO, H. L. (enero de 2016). Análisis automático de opiniones. Obtenido de <https://www.gelbukh.com/thesis/Hugo%20Librado%20Jacobo%20-%20MSc.pdf>
- 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>
- NARVAEZ, M. S. (2017). Repositorio ESPE. Obtenido de <https://repositorio.espe.edu.ec/bitstream/21000/13528/1/T-ESPE-053887.pdf>
- PIATETSKY, G. (2014). Kdnuggets. Obtenido de <https://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science-projects.html>
- RSTUDIO. (2018). Rstudio. Obtenido de <https://www.rstudio.com/products/RStudio/>
- STROHNE. (08 de 01 de 2018). Facepager. Obtenido de <https://github.com/strohne/Facepager/releases>

COMO CITAR ESTE ARTIGO:

Zambrano, F. J. R., Flores, B. F. V., Mendoza, W. J. P., Zambrano, R. A. R., Rivadeneira, S. M. C. (2020). Aplicación de minería de texto para el análisis de sentimientos del servicio de telefonía móvil en el Ecuador. *Holos*. 36(7), 1-16.

SOBRE OS AUTORES

F. J. R. ZAMBRANO

Profesor Titular de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí. Ingeniero en Sistemas - Universidad Laica Eloy Alfaro de Manabí, Ecuador. Especialista en Diseño Curricular



por Competencias - Universidad del Mar, Chile. Mestre em Modelação, Analise de Dados e Sistemas de Apoio a Decisão - Universidade do Porto, Portugal. E-mail: fabrijrz@gmail.com
ORCID ID: <http://orcid.org/0000-0002-5384-7243>

B. F. V. FLORES

Ingeniero en Sistemas - Universidad Laica Eloy Alfaro de Manabí, Ecuador. E-mail: e1311434961@uleam.edu.ec.
ORCID ID: <https://orcid.org/0000-0002-6358-6518>

W. J. P. MENDOZA

Ingeniera en Sistemas - Universidad Laica Eloy Alfaro de Manabí, Ecuador. E-mail: e1315606630@uleam.edu.ec
ORCID ID: <https://orcid.org/0000-0001-6077-1796>

RODOLFO ANDRES RIVADENEIRA ZAMBRANO

Profesor de la Facultad de Ciencias Matemáticas, Físicas y Químicas, Universidad Técnica de Manabí. E-mail: andrerz@hotmail.es
ORCID ID: <http://orcid.org/0000-0002-9707-7084>

SILVIA MERCEDES CARVAJAL RIVADENEIRA

Economista - Universidad Técnica de Manabí, Ecuador. Directora Pedagógica - Unidad Educativa Julio Pierregrosse. E-mail: silvitacarvajal@gmail.com
ORCID ID: <http://orcid.org/0000-0003-4587-912X>

Editor(a) Responsável: Francinaide de Lima Silva Nascimento

Pareceristas Ad Hoc: MARCELO DAMASCENO DE MELO E JOSÉ YVAN LEITE

