

Text mining for the analysis of digital consumer behavior on Twitter

Minería de texto para el análisis del comportamiento del consumidor digital en Twitter

Angel M. Plaza, Msc.¹, Julio Marcillo, MG.¹, José Hidalgo, Msc.¹, Oscar Anchundia, MG.¹, Luis Pilacuan, MG.¹, Aldo Parrales, MG.¹, William Navas, MG.¹

¹Universidad de Guayaquil, Ecuador, angel.plazav@ug.edu.ec, julio.marcillop@ug.edu.ec, jose.hidalgo@ug.edu.ec, oanchundia@gmail.com, luis.pilacuanb@ug.edu.ec, aldo.parrales1@ug.edu.ec, william.navase@ug.edu.ec

Resumen – El presente trabajo tiene como objetivo caracterizar el comportamiento del consumidor digital ecuatoriano en redes sociales, por lo que se propone analizar los comentarios de los tweets publicados por empresas ecuatorianas que ofrecen servicios de telecomunicaciones. El uso intenso de las redes sociales por parte del consumidor ecuatoriano, y el cambio de modelo de negocio en los comercios ecuatorianos por efecto de la crisis sanitaria exige a las empresas desarrollar estrategias de marketing digital, publicidad o mejorar productos y servicios que ofrecen para satisfacer las necesidades del denominado consumidor digital. Dentro de las actividades se propone extraer los datos mediante los token de acceso que ofrece el Api oficial de Twitter, de los nueve procesos determinados para el tratamiento de grandes volúmenes de datos, solo se van a utilizar seis procesos que se adaptan muy bien al requerimiento del análisis de datos, en el Jupyter Notebook con el uso de Python 3 se aplica el análisis de frecuencia de palabras mediante el uso de algoritmos automáticos. Los resultados analizados permitirán mostrar características positivas y negativas de la interacción del consumidor en redes sociales relacionados a la calidad del servicio.

Keywords— Minería de datos, Twitter, API, tokens, big data, Jupyter Notebook, Python.

Abstract– The present work aims to characterize the behavior of the Ecuadorian digital consumer in social networks, for which it is proposed to analyze the comments of the tweets published by Ecuadorian companies that offer telecommunications services. The intense use of social networks by the Ecuadorian consumer, and the change in the business model in Ecuadorian businesses due to the health crisis, requires companies to develop digital marketing strategies, advertising or improve products and services they offer to satisfy the needs of the digital consumer. Within the activities, it is proposed to extract the data through the access tokens of the Twitter API, of the nine processes determined for the treatment of large volumes of data, only six processes that adapt to the requirement of data analysis will be used. In the Jupyter Notebook with the use of Python 3, a word frequency analysis is developed using automatic algorithms. The analyzed results will allow to show positive and negative characteristics of the consumer's interaction in social networks related to the quality of the service.

Keywords-- Data mining, Twitter, API, tokens, big data, Jupyter Notebook, Python.

I. INTRODUCCIÓN

Debido a los diversos escenarios que se han implementado para comercializar productos que se han desarrollado a través de las redes sociales, es de vital importancia en los procesos de marketing entender al consumidor digital en su comportamiento y características debido a las particularidades heterogéneas de gustos, expectativas y capacidad de compra. El propósito es satisfacer las necesidades que tengan con relación a su capacidad de evaluar la compra de servicios o productos. Un gran número de autores han tratado de definir y entender el comportamiento del consumidor (sentimiento, confianza, emociones, percepción), frente al uso de las páginas web, redes sociales, publicaciones, etc. De esta manera ayudar a comercios electrónicos a tener una perspectiva adecuada del consumidor digital y que es lo que en realidad busca cuando realiza las compras online. Pero el modelamiento de la conducta del usuario digital requiere ampliar una variedad de características y definiciones de comportamiento que limiten su perfil.[1]

El comportamiento del consumidor digital es cada vez más exigente, generando complicaciones al momento de caracterizar su conducta en una manera exacta, debido a los cambios constante que presenta en su preferencia, el usuario normalmente cuenta con una amplia variedad de características que definen su comportamiento de consumo dentro del mundo digital, debido a su estado activo que demuestra en estare importante como lo es el internet.

Según la Conferencia de las Naciones Unidas sobre comercio y desarrollo (UNCTAD) y la participación de tres entidades empresariales conexas con el ambiente cibernético, declaran que, debido a la crisis sanitaria, más de la mitad de personas encuestada realizan con frecuencia compras online y acuden a las noticias que presenta el internet en diferentes plataformas para estar informados de acontecimientos que suceden a diario y usar como ocio digital. [2]

Actualmente, como respuesta a las restricciones de movilización impuestas como parte de la cuarentena y luego las restricciones de cantidad de usuarios en un espacio físico

Digital Object Identifier (DOI):

<http://dx.doi.org/10.18687/LACCEI2022.1.1.772>

ISBN: 978-628-95207-0-5 ISSN: 2414-6390

determinado, muchos de los negocios como tiendas, almacenes, boutique, entre otros que se desenvolvían tradicionalmente solo con ventas presencial, han tenido que cambiar o modificar su modelo de negocio, trasladando el proceso de compra y venta de sus productos a nivel de la web. Este cambio ha incrementado el volumen de transacciones, publicidad y publicaciones, desde el negocio hasta los usuarios que se mantienen conectados mayor cantidad de tiempo en Internet.

A medida que pasa el tiempo las compras en internet crecen de una manera sorprendente, se puede decir, que aún más por motivo de la crisis sanitaria que está pasando el planeta entero, la mayoría de las personas efectúan sus compras online por miedo a contraer el virus debido a que si van a una tienda física pueden contagiarse y por esta razón se ha elevado el consumo digital en el internet. El cambio de modelo de negocio implementado actualmente por los comercios ecuatorianos, tienen como objetivo de permitir que diferentes empresas desarrollen estrategias ya sea de marketing digital, publicidades, o permitir mejorar sus productos o servicios que ofrecen a los clientes, buscando de esa manera satisfacer las necesidades del consumidor digital. [3]

Se puede notar que la crisis sanitaria que está atravesando actualmente el mundo entero, ha dejado un incremento de las interacciones del consumidor digital en la participación de diferentes actividades; por teletrabajo, compras online, acceso a las redes sociales, entre otras. Todos estos usuarios navegan desde casa sin tener que salir de ella. En varias ocasiones el consumidor digital, se cansa de observar el mismo contenido, debido a que pasa mucho tiempo conectado en sitios digitales aumentando su deseo de querer observar más contenido que le parezca interesante. [4]

Una característica importante del consumidor digital es su intensa participación en las secciones de comentarios y publicaciones sobre la satisfacción de los productos o servicios que adquiere en línea, deja una gran cantidad de información almacenados en las bases de datos de las páginas web o redes sociales de ventas. Estos sitios guardan importantes cantidades de información que pueden ser usados para saber acerca de temas de preferencia, que sentimientos o emociones expresan, por lo tanto, a través de un proceso de minería de textos se puede realizar un análisis de estos datos textuales con el fin de conocer patrones y tendencias de interés del consumidor digital. [3] [5]

Es de gran importancia llevar a cabo este análisis porque debido a esto se podrá conocer las características más adecuadas sobre el comportamiento del consumidor digital, sin que este sea afectado por su alta capacidad de exigencia. Sino más bien llegar a acuerdos que lo beneficien y para esto será necesario aplicar técnicas de minería de datos, mediante un proceso en fases que permitirá obtener resultado sobre análisis del comportamiento del consumidor digital.

Para el desarrollo de este análisis del comportamiento del consumidor digital se aplicará la técnica de minería de datos, con el uso de una metodología expuesta en fases: Identificación de datos, adquisición y filtrado de datos,

extracción de datos, validación y limpieza de datos, análisis de datos (Minería de texto), visualización de datos. Y para esto se usará la API de Twitter, para tener acceso a los datos de esta red social con los que se trabajará mediante el programa Jupyter con Python con el uso de varias librerías que permitirán la extracción de los datos y aplicar el análisis de la Data mining para finalmente presentar los resultados en gráficos interactivos. [6] [7]

I. ANTECEDENTES

Se pretende trabajar sobre la red social Twitter, durante un periodo determinado de tiempo en el cual se utilizará herramientas que permitan recopilar datos de esta red social, y llegar a obtener las características y comportamiento del consumidor digital de los comentarios, publicaciones, lo que permitirá establecer un perfil más definido del consumidor digital, aportando con esta investigación a proyectos futuros que quieran realizar cualquier estrategia de marketing efectivas para ayudar a fortalecer las ventas de los comercios en línea.

A. Consumidor digital

El comportamiento del consumidor ya no es el mismo desde hace algunos años, está más conectado, con más acceso, y tiene un compromiso más profundo con el contenido y las marcas, gracias a la innovación tecnológica de dispositivos y plataformas digitales. El internet está siendo una innovación transformadora realizando cambios generacionales en la forma en que viven y compran estos consumidores [8]. Por eso es de suma importancia estar en constante investigación para conocer las tendencias del consumidor digital y ayudar a empresas a ajustar las estrategias que usan para, para estar más cerca de sus clientes manteniéndolos fieles a la marca.

Los consumidores digitales son personas con características propias y expectativas diferentes, que se encuentran activas en internet, día y noche, accediendo a todo tipo de información, con relación a los productos y servicios. Estos utilizan diferentes dispositivos tecnológicos para mantenerse conectados en cualquier sitio de internet. Los mercados digitales presentan productos y servicios a los consumidores, de manera que generar una necesidad de comprar, pero esto actualmente está cambiando, debido a que existen cambios constantes en los consumidores digitales y por ende las tendencias de mercados dependen de estos cambios; comportamiento, actitudes y las expectativas del consumidor.

A continuación, se describe algunas características que definen al consumidor digital. [9] [10]

- Es experto: tiene más conocimiento que los propios vendedores.
- Produce información: comparte sus experiencias mediante, comentarios, publicaciones en las redes sociales, acerca de los productos o servicios que adquieren.
- Siempre está activo en internet.

- Están bien informado
- Espera honestidad de parte de las marcas.
- Busca personalización en las marcas, que le ofrezcan lo que quiera y cuando quiera
- Espera ser escuchado por la marca
- Es impaciente, sino obtiene rápidamente lo que desea busca otras alternativas.
- Busca contenidos interesantes para compartirlos en sus redes.
- Es infiel: debido a que pueden comparar precios, marcas, calidad de los productos o servicios, interfiriendo en las decisiones del consumidor.
- Es poderoso tiene el internet en sus manos y puede expresar cualquier tipo de opinión y ser escuchado por todo el mundo, por ejemplo; si esta insatisfecho por la marca, con un simple comentario o publicación negativo puede dañar la reputación de la empresa.
- Usa nuevos lenguajes, por ejemplo; emoticones o abreviaturas.
- Le gusta probar el producto o servicio antes de comprarlo.
- Es multitarea porque mientras navega por internet puede realizar otras tareas

B. Análisis de comportamiento en línea

El análisis del comportamiento del consumidor en línea, en estos últimos años y aún más por la pandemia, está siendo una cuestión muy estudiada por las empresas o negocios que venden productos y servicios en línea. [11]

Se define el termino comportamiento del consumidor; como el comportamiento que los consumidores exponen al buscar, utilizar, comprar, evaluar y desechar los productos o servicios que consideran que satisfacen sus necesidades. Por otra parte, el análisis del comportamiento se enfoca en la forma que las clientes toman decisiones para gastar sus recursos como los son; dinero, tiempo y esfuerzos en los artículos para su consumo. [12][13]

El comportamiento en línea de las personas es exigente y a su vez, se torna no lineal ya que buscan la mejor opción a la hora de adquirir el servicio o producto, revisa diferentes sitios de comprar e indaga entre los comentarios, para saber las opiniones de otras personas que han comprado el producto o servicio de esa marca. Con la finalidad de elegir el mejor producto o servicio para satisfacer su necesidad.

Existen varios objetivos y beneficios que se pueden obtener a partir de análisis del comportamiento del consumidor en línea, ya que son de importancia para las empresas o negocios que tienen sus tiendas en internet.

El objetivo de analizar el comportamiento del consumidor digital [13] es el siguiente:

- Estudiar las conductas de las personas a la hora de obtener, usar y consumir algún producto o servicio.
- Estudiar el porqué, dónde, con que frecuencias y en qué condiciones se producen los consumos de productos y servicios que se ofrecen en línea.

- Trata de comprender, explicar y predecir las conductas humanas relacionadas con el consumo digital.

Además, el análisis del comportamiento en línea del consumidor deja múltiples beneficios que ayudan al fortalecimiento de la marca y a satisfacer las necesidades de los clientes [14]:

- Se puede tener un mejor conocimiento de las necesidades de los consumidores y a la vez oportunidades de mercado para las empresas.
- Analizar el comportamiento es importante para plantear estrategias de marketing.

C. Análisis de sentimiento

El análisis de sentimientos se realiza para saber cómo el público percibe o acoge una marca o actividad en línea. Para esto se evalúan los comentarios de los usuarios, si se han escrito con intenciones positivas o negativas referente a la marca. Para esto se usa la minería de texto, es decir, el análisis automático de textos, escritos en lenguaje natural. Los métodos de análisis más sencillo buscan en el texto palabras que tengan un significado positivo o negativo, a partir de un diccionario adecuado, según el tema y que ha sido creado previamente. Los resultados obtenidos permiten dar una idea general, pero no es útil para analizar el ambiente real, además que no arrojan resultados de frecuencias de ciertas palabras en positivas o negativas con relación a la evaluación del producto. Es factible que para el análisis de sentimientos se usen herramientas de inteligencia artificial, por motivo de que existen textos que los consumidores escriben en redes sociales como si estuvieran hablando con un amigo personalmente y no tiene normas gramaticales, tampoco buena ortografía, ya que las herramientas de análisis deben detectar esos matices. [15]

En este proceso, es adecuado usar métodos de aprendizaje automático, con herramientas de sentimientos que reconozcan de una forma precisa el público objetivo y el entorno del producto que se está analizando. Con esto se va entrenando poco a poco las herramientas para mejorar la calidad de los resultados a largo plazo. El análisis de sentimientos tiene como objetivo, obtener una idea general de la opinión y estado de ánimo de un público objetivo con relación a un producto o servicio. No solo se recomienda analizar las reseñas del producto de la página web o de las tiendas online, sino también las publicaciones en Facebook, Twitter y otras redes sociales. Este análisis se realiza con el fin de descubrir las emociones que están detrás del texto y que quiere expresar realmente en el texto, el individuo. [15]

D. Minería de datos

La minería se refiere a la recopilación de toda aquella información valiosa que se encuentra almacenada en grandes bases de datos, es un término que actualmente está sonando mucho por su capacidad de utilizar la tecnología computacional, emplean varias técnicas para recopilar grandes volúmenes de datos con la ayuda de algoritmos, inteligencia artificial y programas estadísticos, entre otros.

La minería de datos es usada para analizar lo que sucede en un escenario específico e intenta predecir lo que sucederá en un futuro con relación a las empresas o negocios, conllevando a tomar buenas decisiones gracias al manejo de la información. Lo que hace la minería de datos es buscar patrones de coincidencia de palabras entre los datos que tienen las empresas, brindándoles información valiosa para el desarrollo de ventajas competitivas para sobresalir en el mercado. [16] [17]

Por ejemplo si una empresa tiene una tienda digital, esta puede acceder a los datos de sus consumidores acerca de los comentarios o publicaciones que realizan después de obtener los productos o servicios, dicha información es llevada al procesamiento de datos, aplicando técnicas de minería de datos y con los resultados obtenidos servirles para mejorar las estrategias de la empresa, de tal manera que sus ventas incrementen y ayuden a ofrecer un mejor servicio a los clientes con todo el conocimiento que tienen acerca de los gustos, preferencias y comportamientos del consumidor. [18]

La minería de datos tiene bases en la inteligencia artificial y en el análisis estadístico, ya que esto permite llevar a cabo el análisis de los datos.

Un conjunto de empresas en 1996, estuvieron de acuerdo en crear un proceso estándar para la Minería de Datos, este es un proceso independiente de los productos y técnicas de softwares el cual describen detalladamente el ciclo de vida de un proyecto de Data Mining y esta dividido en seis fases, siendo un proceso de alto nivel. [19]

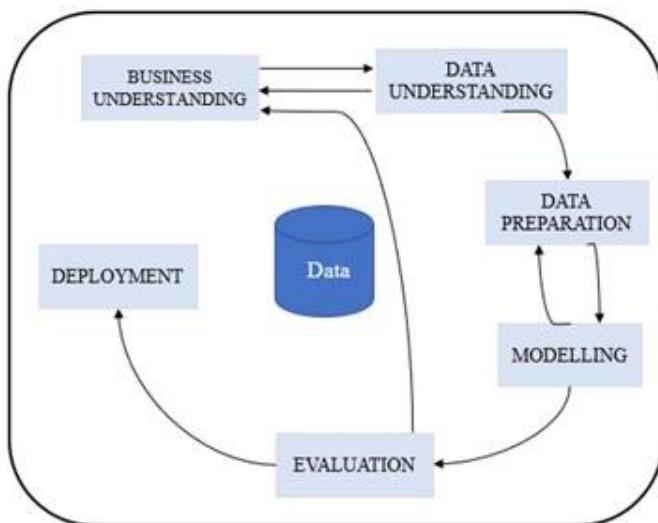


Figura 1 - Proceso de la Minería de Datos

II. MATERIALES Y MÉTODOS

La metodología que se utilizará se basa en el ciclo de vida de analítica de datos, propuesto por Erl Khattak en su libro Big Data Fundamental, Big Data [20] se define como el

campo multidisciplinario dedicado al análisis de grandes volúmenes de datos no estructurados.



Figura 2 - Fases de la metodología de analítica de datos [21]

Para cumplir con el análisis de Big data, se necesita aplicar paso a paso la metodología, en la que consiste en organizar las actividades y tareas con relación a la adquisición, procesamiento, análisis y reutilización de datos. Sin embargo, no se pretende cumplir con el Big data, debido a que se escogerán solamente ciertas fases que se adapten al proceso de análisis de datos con la técnica de minería de datos.

La metodología propuesta por Erl, Khattak & Buhler, está conformada por nueve fases, para el análisis de grandes volúmenes de datos, características de velocidad y variedad de datos que se están procesando, pero en el desarrollo de la propuesta solo se usará seis fases, previo a esta decisión se realizó una comparativa de las metodologías más usadas en la minería de datos y se llegó a la conclusión de escoger, la que mejor se adapta al ciclo de vida de analítica de datos, adecuándola a las necesidades del proyecto, por ende solo se eligieron seis fases que van acorde al proceso de minería de datos. [20]

De acuerdo a las fases propuestas para el desarrollo del proyecto, se inicia con la identificación de una red social como lo es Twitter, donde se usarán los tweets publicados por los perfiles de las empresas, que ofrecen comercio online en dicha red, se han elegido; CNT, Claro y Netlife, además se tomarán los comentarios que realizan los usuarios en los tweets, a partir de estos se recolectan los datos, mediante un proceso de adquisición y filtrado, usando la API de Twitter para extraer los datos con la plataforma web Jupyter Notebook, que usa el lenguaje de programación de Python, los datos se guardaran en un formato CSV en Excel. Posterior a esto se procede con el procesamiento de datos, donde, a través del uso de Jupyter Notebook Python 3, se ejecuta la extracción, validación y limpieza de los datos. Las fases antes mencionadas, preparan los datos para el análisis.

Luego se llevará los datos para el análisis, donde se utilizará el mismo software antes mencionado. Esta fase se enfoca en usar las técnicas de minería de datos, en la se

aplicará la técnica, de análisis de frecuencia usando códigos de lenguaje de programación que permitan desarrollar estos algoritmos automáticamente. Por otra parte, los datos serán guardados en archivos de textos planos en formato CSV, para su fácil comprensión de lectura y almacenamiento. Y, por último, la fase de visualización de datos que presentará los resultados mediante gráficos ilustrativos para puedan ser entendidos sin problemas, para eso se usará tendrán en cuenta las nubes de palabras y barras estadísticas.

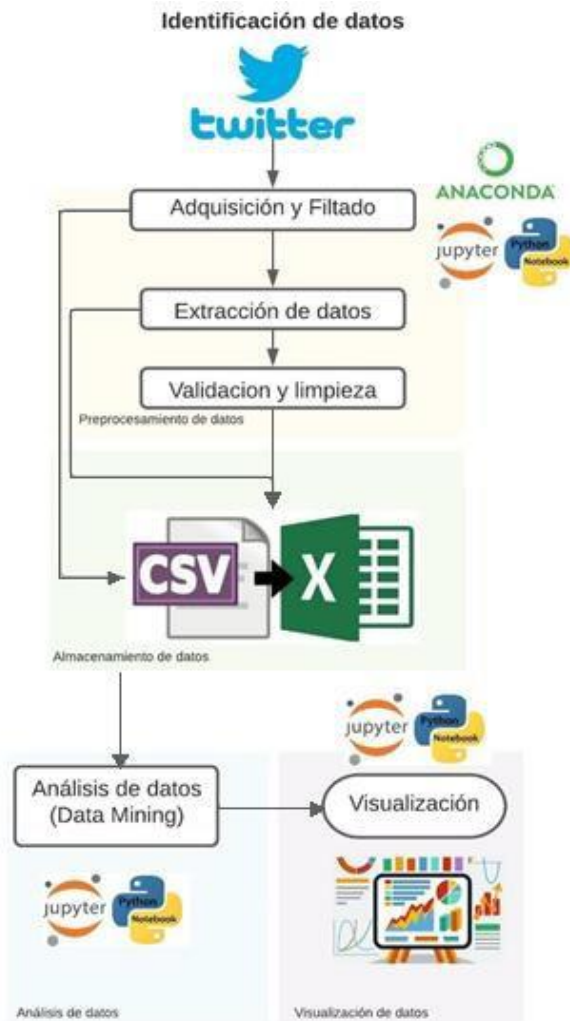


Figura 3 - Esquema del proceso de análisis de datos

A. Identificación de datos

Como fuente de datos se ha identificado la red social Twitter de opinión pública, siendo esta calificada como una de las redes sociales más usadas por las personas para emitir comentarios. La ventaja que posee esta red social es que la mayoría de sus cuentas son de acceso público (se hace uso de los tweets públicos, generados por los clientes de las cuentas de; Claro Ecuador, CNT Ecuador y Netlife Ecuador), esto permite extraer una gran cantidad y diversidad de datos para analizarlos.

B. Adquisición y filtrado de datos

Los datos son adquiridos de los tweets generados como comentarios de las empresas que prestan servicio de telecomunicaciones. Para adquirir los datos se procede a conectarse con los servidores de la red social Twitter mediante la API. En el cuaderno de Jupyter se importan las librerías a usarse para el análisis de datos, después se escriben las credenciales generadas en la cuenta de developer; las cuales son; consumer_key, consumer_secret, Access_token y por último el Access_token_secret.

```
In [67]: #Revisor con hashtags
for status in tweepy.Cursor(api.search, q = "#@netlifeecuador").items(3000):
    print(status.text)

#@netlifeecuador DM enviado. Me urge su solución a este tema
No tengo internet desde las 2pm , llamo a #@netlifeecuador y me citan a las 3pm pero de mañanaaaaaaaaaa. @
#@netlifeecuador el día de ayer quedó en venir el técnico Cristian Guaraca para mover el internet a mi nueva casa y. https://t.co/0AATLE1E1E
#@marcelochever: #@netlifeecuador @caamichelena @DEFENSORIAEC Marcelo por supuesto que es un atropello al bolsillo p. https://t.co/51u152u08
#@netlifeecuador buenas noches, por favor a que hora restablecen el servicio? lleva tiempo la reparación desde la tarde, gracias
#@netlifeecuador buenas noches su ayuda por favor con el DM que les envié. Gracias.
RT @abustamanteofic: Señores #@netlifeecuador van 22 minutos tratando de contactarlos, pero no hay respuesta. como servicio al cliente es un...
RT @marcelochever: RESPONDAN #@netlifeecuador O LOS SON MUY SABIDOS O UNO ES MUY COJUDO PARA PAGARLES LO QUE OJICHA LES DA LA GANA!!!! XQ E...
RESPONDAN #@netlifeecuador O LOS SON MUY SABIDOS O UNO ES MUY COJUDO PARA PAGARLES LO QUE OJICHA LES DA LA GANA!!!!. https://t.co/YV1C1xyg5V
#@fabrisanchezb #@netlifeecuador Estimado usuario:
Para reclamos, ARCTEL agradece registrarse e ingresar en el sigul. https://t.co/mjgsejrtm
@jsemejlliones #@netlifeecuador Estimado usuario:
Para reclamos, ARCTEL agradece registrarse e ingresar en el sig. https://t.co/24gTvf0KH
#@netlifeecuador usan el mejor servicio de internet...hasta ahí al nlan u ni al servicio. favor nos !!
```

Figura 4 - Visualización de datos de tweets

C. Extracción de datos

Se va a extraer la base de datos en archivo CSV, mediante la librería panda y se le asignan los mismos nombres a la fila principal de la tabla, pero de forma abreviada referente a los atributos asignados anteriormente, así como también el delimitador para obtener una separación entre las columnas. Los datos que se visualizan en la ilustración 47, son presentados en tablas como archivo de data frames, para aquello se usa la función “tweets_df”, visualizado los datos de manera ordenada y filtrado por atributos.

```
In [511]: tweets_df = pd.read_csv('@netlifeecuador17.21.17:22-Nov.csv', names=['Fecha', 'user.name', 'user.location', 'place', 'retweet_count',
# ...
In [511]: :22-Nov.csv', names=['Fecha', 'user.name', 'user.location', 'place', 'retweet_count', 'retweeted', 'full_text.replace'], delimiter='|')
```

Figura 5 - Proceso data frames. Adaptada de Jupyter Notebook

Para este estudio se usó el texto de los tweets de la columna llamada “full_text.replace”. Los cuales se juntan en un solo texto para ser llevados a la fase de limpieza y validación. Se crea una variable llamada “textos” en la que se almacenan y juntan todos los tweets en un solo texto.

D. Validación y limpieza de datos

La limpieza de datos y validación es una de las fases importantes de la propuesta de proyecto ya que establece reglas de validación y elimina datos no válidos conocidos que

que se comprendan rápidamente los datos analizados y de esta manera tomar decisiones o aplicar mejoras en los negocios de servicios de telecomunicaciones que presentan en las redes sociales (Twitter).

Se usó la librería de WordCloud, como herramienta para visualizar las palabras más nombradas en el texto en tamaño grande y las menos nombradas en palabras pequeñas. Con el texto limpio y validado las palabras con más frecuencias del texto. Se procede a crear la nube de palabras; en la cual se utiliza la función con los parámetros, que permite ajustar la altura, anchura, el color del fondo y palabras que deseo que se presenten (“collacation_threshold”). Estos parámetros se pueden ir ajustando manualmente. Se puede observar la nube de palabras generada a partir de los textos analizados, según sus frecuencias. (palabras grandes: son las más nombradas y las pequeñas: menos nombradas en los comentarios de los tweets)

III. RESULTADOS

Como resultados se presenta el análisis de datos públicos, obtenidos de comentarios en las cuentas de: ClaroEcu, Cnt_Ecu, NetlifeEcuador de la red social Twitter.

A continuación, se presentan una parte de la información recopilada de cada cuenta por separado, las palabras más nombradas de la información pública obtenida de la red social en tablas e ilustraciones con sus respectivos conteos por palabras, además del conjunto de palabras negativas y positivas relacionadas a cada empresas. También se ilustran las nubes de palabras que dan una idea más clara de las palabras más repetidas.



Figura 9 - Nube de palabras. Claro

Con relación a la cuenta de ClaroEcu, la nube de palabras no muestra una tendencia significativa de alguna palabra sobre las demás, en cuanto a la clasificación de los términos negativos para conocer con qué frecuencia son nombrados, en la tabla 1 se visualiza la palabra “pésimo”, “problema” y ”peor” con mayor frecuencia que otros términos

de un conjunto aproximado de 3000 tweets filtrados y depurados a lo largo de una semana. Como se muestra en la tabla el conteo de las tres palabras suman un total de 89.

Tabla 1 - Frecuencia de palabras negativas CLARO

Palabras (-)	Frecuencia
Pésimo	35
Problema	30
Peor	24
total	89

Para llevar el análisis de la empresa de CNT, se extrajeron y validaron aproximadamente 1222 tweets, como se visualiza en la nube de palabras, la mayor cantidad de menciones se refieren a los siguientes términos; “servicio”, “internet” y “lenin”. Con respecto a la clasificación de los términos negativos, en la tabla 2 se visualiza las palabras “Problema”, “Pésimo” y ”Arcotelec” con mayor frecuencia que otros términos. Como se muestra en la tabla el conteo de las tres palabras suman un total de 198.



Figura 10 - Nube de palabras. CNT Tabla

2 - Frecuencia de palabras negativas CNT

Palabras (-)	Frecuencia
Problema	55
Pésimo	48
Arcotelec	95
total	198

Para analizar está cuenta se han extraído aproximadamente 1000 tweets, como resultado se puede observar en la nube de palabras correspondiente, las tres primeras palabras son los requerimientos fijos por la cual están comentados o publicando los tweets. Se puede observar que existen palabras como “obsequio”, “gratis” y “oferta”.



Figura 11 - Nube de palabras. NETLIFE

Tabla 3 - Frecuencia de palabras negativas NETLIFE

Palabras (-)	Frecuencia
Problema	94
Pésimo	51
Peor	34
total	179

Con respecto a la clasificación de los términos negativos, en la tabla 3 se visualiza las palabras “Problema”, “Pésimo” y “Peor”. Como se muestra en la tabla el conteo de las tres palabras suman un total de 179.

A. Características del consumidor digital ecuatoriano

Para lograr determinar una aproximación del perfil del consumidor digital ecuatoriano se revisó el comportamiento en redes sociales (Twitter) tomando en consideración los mensajes publicados relacionados a tres operadores de servicio de Internet, la construcción del perfil se ha basado en características de comportamiento similares que se han podido extraer de la interacción en estos mensajes. Algunas de estas características son:

Actividad: medición de la cantidad de comentarios en un espacio de tiempo. De los datos obtenidos durante una semana se recolectaron aproximadamente 3000 tweets entre las tres operadoras, la hora en que más comentan los usuarios es entre las 15H00 hasta las 17H00 con un número de comentarios de 1729 promedio durante ese lapso, a las 00H00 disminuye la cantidad de comentarios realizados por las personas. Se puede concluir que la hora donde más está activo el consumidor es en la tarde.

Compara los productos con la competencia: veces que menciona a la competencia en los tweets. De acuerdo con los resultados que se muestran en la Tabla 4, los clientes de Netlife y CNT, nombran a la competencia entre 107 hasta 115 veces con el fin de comparar la calidad del servicio, esta interacción pretende llamar la atención para que de esta

manera sus operadoras mejoren o realicen ofertas que los favorezcan.

Tabla 4 - Competencia nombrada por las operadoras

Competencia	Netlife	CNT	Claro
claroecua	107	115	-
movistarec	-	-	51
netlifecuador	-	-	26
total	107	115	77

Exigente: Cantidad de veces que se realiza requerimientos sobre la calidad del servicio.

En la tabla 5, presenta el análisis de las palabras que más son mencionadas en los comentarios, con relación a las exigencias que presentar los consumidores, por motivo de los servicios de internet de las operadoras analizadas.

Tabla 5 - Requerimiento a las operadoras sobre los servicios adquiridos

Palabras	Cnt	Claro	Netlife
Factura	81 mención	-	218 mención
Velocidad	61 mención	-	200 mención
Señal	32 mención	30 mención	-
Respuesta	34 mención	-	-
Solución	-	50 mención	32 mención
Oferta	-	-	189 mención
Total	176 mención	80 mención	639 mención

Ciudad: Lugar donde los usuarios emiten los tweets.

Tabla 6 - Ciudades donde se conecta el consumidor

Ciudad	Frecuencia
Guayaquil	588
Quito	298
Cuenca	22
Manta	16
Machala	10

Dispositivo: tipo de dispositivo (sistema operativo) que utiliza para mantenerse conectado a las redes sociales.

Tabla 7 - Sistemas operativos que usa para mantenerse conectados

Dispositivo	frecuencia
Android	1677
IPhone	632
IPad	17

II. CONCLUSIONES

Como planteamiento inicial para el desarrollo del presente proyecto, se planteo poder caracterizar el comportamiento del consumidor digital ecuatoriano de acuerdo con sus

interacciones a través de las redes sociales, para lo cual se plantearon los pasos en el proceso metodológico, de acuerdo con los resultados obtenidos se puede concluir con el cumplimiento de los siguientes puntos.

Por medio del uso del API de Twitter se logró obtener un conjunto de datos respetables para el procedimiento de análisis, filtrado, extracción, procesamiento, validación, limpieza y visualización de los mismos.

La selección adecuada de la herramienta para el procesado de datos, fue el entorno de trabajo de Jupyter Notebook que permitió desarrollar el código de programación con el uso de Python 3, para de esta manera analizar los datos extraídos de la red social Twitter, mediante el uso de ciertas librerías que facilitaron realizar algoritmos de aprendizaje automático representados en líneas de códigos que permitieron desarrollar la extracción, clasificación, limpieza y análisis de frecuencia de los datos, en tiempos rápidos de ejecución.

Se obtuvieron resultados consistentes en el análisis de frecuencia de palabras a los textos de los tweets, logrando por lo tanto obtener palabras más nombradas, indispensables para conocer de qué forma se comporta el consumidor ante los servicios que adquiere de las operadoras de telecomunicaciones.

Como recomendación se plantea realizar un nuevo estudio basado en el uso de los emoticones, que son usados por muchos usuarios para mostrar situaciones de agrado o desagrado, que son parte de un análisis de sentimientos.

REFERENCES

- [1] Millán, I. S. (2019). El comportamiento del consumidor y las nuevas tendencias de consumo ante las TIC. *Esic market*, (164), 599-642.
- [2] Noticias ONU. (8 de Octubre de 2020). Mirada global historias humanas. Obtenido de Noticias ONU: <https://news.un.org/es/story/2020/10/1482072>
- [3] TONIUT, Hernán Ramón. El uso del modelo de negocio para la innovación empresarial: Aportes de los principales autores. *Palermo Business Review*, 2020, no 22, p. 139-156.
- [4] GARCÍA-MADURGA, Miguel Ángel; GRILLÓ-MÉNDEZ, Ana Julia; MORTE-NADAL, Tamara. La adaptación de las empresas a la realidad COVID: una revisión sistemática. *RETOS. Revista de Ciencias de la Administración y Economía*, 2021, vol. 11, no 21, p. 55-70.
- [5] ZOMEÑO, Daniel; BLAY-ARRÁEZ, Rocío. Big data e inteligencia editorial en el branded content y en los nuevos modelos de negocio de los medios. *Profesional de la Información*, 2021, vol. 30, no 1.
- [6] García, J. (2016). Líneas de investigación en minería de datos en aplicaciones en ciencia e ingeniería: Estado del arte y perspectivas. *Conference on Information Engineering and Computer*, 1-4.
- [7] Aquino, A., Molero, G., & Rojano, R. (2015). Hacia un nuevo proceso de minería de datos centrado en el usuario. *Pistas Educativas Año XXXVI*, 272-291.
- [8] HASRAT, Tasrik; ROSYADAH, Khairina. Usability Factors as Antecedent and Consequence on Business Strategy and SERVQUAL: Nielsen & Mack Approach. *Golden Ratio of Marketing and Applied Psychology of Business*, 2021, vol. 1, no 2, p. 81-92.
- [9] Raúl. (3 de Agosto de 2020). The Technology Savvy. Obtenido de https://thetechnologysavvy.com/characteristics-of-the-digital-consumer/#The_Digital_Consumer
- [10] Mejía, J. (2015). La guía del Community Manager: estrategia, táctica y herramientas. Anaya Multimedia.
- [11] Rodríguez, L., & Sosa, L. (12 de Junio de 2019). Análisis del comportamiento del consumidor digital en Tenerife. Obtenido de <https://riull.ull.es/xmlui/bitstream/handle/915/15000/Analisis%20del%20comportamiento%20del%20consumidor%20digital%20en%20Tenerife.pdf?sequence=1&isAllowed=y>
- [12] Schiffman, L., & Kanuk, L. (2010). *Comportamiento del Consumidor Décima Edición*. México: Pearson Educación.
- [13] Schiffman, L., & Kanuk, L. (2013). *Comportamiento del consumidor*. Pearson Prentice Hall Noticias ONU. (8 de Octubre de 2020). Mirada global historias humanas. Obtenido de Noticias ONU: <https://news.un.org/es/story/2020/10/1482072>
- [14] Rivas, J., & Garde, I. (2015). *Comportamiento del consumidor. Decisiones y estrategia de marketing*. ESIC EDITORIAL.
- [15] IONOS. (5 de Octubre de 2020). Digital Guide. Qué es el análisis de sentimiento. Obtenido de <https://www.ionos.es/digitalguide/online-marketing/analisis-web/analisis-de-sentimiento/>
- [16] PÉREZ-SUASNAVAS, Ana-Lucía; CELA, Karina; HASPERUÉ, Waldo. Beneficios del uso de técnicas de minería de datos para extraer y analizar datos de twitter aplicados en la educación superior: una revisión sistemática de la literatura. Beneficios del uso de técnicas de minería de datos para extraer y analizar datos de Twitter aplicados en la educación superior: una revisión sistemática de la literatura, 2020, p. 181-218.
- [17] CASTRILLÓN-GÓMEZ, Omar D.; SARACHE, William; RUIZ-HERRERA, Santiago. Predicción de las principales variables que conllevan al abandono estudiantil por medio de técnicas de minería de datos. *Formación universitaria*, 2020, vol. 13, no 6, p. 217-228.
- [18] Silva, X. (enero de 2020). Análisis de la minería de datos aplicada en empresas del sector retail. Obtenido de https://repositorio.ucsp.edu.pe/bitstream/20.500.12590/16199/1/DONGO_POZO_ALD_MIN.pdf
- [19] Leventhal, B. (2010). Una introducción a la minería de datos y otras técnicas de análisis avanzado. *Revista de práctica de marketing directo, de datos y digital*, 137-153.
- [20] Erl, T., Khattak, W., & Buhler, P. (2016). *Big Data Fundamentals: Concepts, Drives and Techniques*. New Jersey, United States: Prentice Hall.
- [21] PÉREZ, Miguel A. Luna; ZACATENCO, SEPI ESIME; ÁLVAREZ, Graciela Vázquez. *Metodología de mantenimiento predictivo 4.0. SISTEMAS, CIBERNÉTICA E INFORMÁTICA*, 2019, vol. 16, no 2.
- [22] GÁLVEZ LAO, M^o, et al. Construcción de infraestructura Big Data para el procesamiento y visualización de datos de Twitter. 2019.
- [23] CUQUERELLA DE LORENZO, Juan, et al. Despliegue de motor analítico y herramienta de visualización de datos (Big Data) para el conocimiento del entorno marítimo en tiempo real. 2021.