

 **Ángel Patricio Fajardo Cárdenas**
Facultad de Ingeniería, Universidad de Cuenca,
Ecuador
patricio.fajardo96@ucuenca.edu.ec

 **Néstor Ariel Bravo Chuqui**
Facultad de Ingeniería, Universidad de Cuenca,
Ecuador
ariel.bravo@ucuenca.edu.ec

 **Andrés Vinicio Auquilla Sangolqui**
Facultad de Ingeniería, Departamento de Ciencias
de la Computación, Universidad de Cuenca,
Ecuador
andres.auquilla@ucuenca.edu.ec

 **Paúl Fernando Vanegas Peña**
Facultad de Ciencias Químicas, Departamento de
Espacio y Población, Universidad de Cuenca,
Ecuador
paul.vanegas@ucuenca.edu.ec

Revista Tecnológica ESPOL - RTE
vol. 33, núm. 2, Esp. p. 134 - 146, 2021
Escuela Superior Politécnica del Litoral, Ecuador
ISSN: 0257-1749
ISSN-E: 1390-3659
rte@espol.edu.ec

Recepción: 11 julio 2021
Aprobación: 28 septiembre 2021

DOI: <https://doi.org/10.37815/rte.v33n2.839>

Resumen: En la actualidad, en especial en tiempos de pandemia, el comercio electrónico se convierte en la manera predominante de comercializar productos y servicios en el mundo. Un estudio realizado por la Cámara Ecuatoriana de Comercio Electrónico, en el año 2020, demuestra que las compras y ventas a través de canales digitales se han incrementado al menos 15 veces desde el inicio de la pandemia. Por lo tanto, para realizar estudios de mercado las empresas deben buscar nuevas formas de extraer información para luego desarrollar un análisis de la misma y así obtener una ventaja competitiva. La extracción de datos es un proceso complejo y poco escalable, de manera que, esta investigación presenta una metodología para la extracción de información de un sector industrial determinado. La metodología consiste en dos pasos fundamentales, primero se realiza un ranking de las principales fuentes de información disponibles y más utilizadas en un determinado sector de la industria, se consideran varias características y opinión de expertos. Segundo, se propone una plataforma, la cual integra las fuentes de información mejor rankeadas y realiza la extracción de datos. Finalmente, estos datos se presentan en un Dashboard con la disponibilidad de poder descargar y hacer uso en estudios posteriores. Se concluye que las 4 plataformas que mayor beneficio ofrecen para esta investigación son: Google Trends, Facebook, YouTube y Twitter. También existen fuentes de información que al aplicar el análisis propuesto tienen una calificación alta, sin embargo, la extracción de los datos es difícil debido a sus políticas de seguridad.

Palabras clave: API, Dashboard, Comercio Electrónico, Metodología de extracción de datos, MiPymes.

Abstract: Nowadays, and more with the pandemic, e-commerce is becoming the predominant way of marketing products and services in the world. A study conducted by the Ecuadorian Chamber of Electronic Commerce in 2020 shows that purchases and sales through digital channels has increased at least 15 times since the beginning of the pandemic. Therefore, to conduct market research companies must seek

new ways to extract information and then then carry out its analysis and thus obtain competitive advantage. Data extraction is a complex and not very scalable process; therefore, this research presents a methodology for the extraction of information from a given industrial sector. The methodology consists of two fundamental steps, first, a ranking of the main sources of information available and most used in the country in a given industry sector is made, several characteristics and expert opinion are considered. Second, a platform is proposed that integrates the best ranked information sources and performs data extraction. Finally, these data are presented in a Dashboard with the availability to be downloaded and used in subsequent studies It is concluded that the 4 platforms that offer the greatest benefit for this research are: Google Trends, Facebook, YouTube and Twitter. There are also sources of information that have a high rating when applying the proposed analysis, however, data extraction is difficult due to their security policies.

Keywords: API, Dashboard, Data Extraction Methodology, E-Commerce, SMEs.

Introducción

Hoy en día, el estudio de mercado es una herramienta de vital importancia antes de comenzar un proyecto de estrategia comercial de cualquier empresa. Además, contribuye en la evaluación de la rentabilidad del proyecto y aumenta la probabilidad de éxito, permite conocer mejor a los potenciales clientes, a la competencia, también puede ayudar a detectar nichos de mercado y obtener así una ventaja competitiva en las empresas (Barrientos Felipa, 2017). En la actualidad, y debido a la pandemia, las ventas y compras se han trasladado mayormente a la modalidad online mediante internet. En Ecuador, 2020, el comercio electrónico alcanza un volumen de negocio de \$ 2.300 millones; un crecimiento de \$ 700 millones (43,75%) respecto a 2019; esto según informe de la Cámara Ecuatoriana de Comercio Electrónico. Según (Barrientos Felipa, 2017), el internet se ha convertido en un insumo intangible de toda estrategia comercial para llegar al cliente; es, a su vez, un medio que puede ser utilizado tanto por grandes y pequeñas empresas.

Estudios como el de la Cámara de Comercio Electrónico del Ecuador 2019-2020 muestran que, a raíz de la pandemia, las compras por canales digitales aumentaron al menos en 15 veces en el país desde el inicio del distanciamiento social. Según esta misma investigación, la pandemia ha impulsado a ese 34% de usuarios que usaban las plataformas digitales como medio de compra secundario (Ecommerceday, 2017) a usar constantemente; así también los usuarios que las usaban esporádicamente o nunca ahora son usuarios de este medio para evitar salir de casa. La COVID-19 ha generado un futuro incierto en varias industrias, pero en el caso del comercio electrónico, las ha impulsado grandemente (E-commerce, 2020).

El comportamiento del mercado ha cambiado en los últimos años, ahora debido a la pandemia las empresas mayormente realizan sus ventas mediante plataformas digitales como redes sociales y páginas de E-Commerce permitiendo así a los clientes tener un medio para expresar su opiniones positivas o negativas sobre un producto. Por este motivo, la industria ha reconocido la importancia de las redes sociales y el contenido generado por el usuario para poder realizar análisis como estudios de mercado (Tripathi & S, 2015). Entonces es necesario implementar nuevos métodos para la recolección de información de las diferentes plataformas de manera efectiva. Esta recolección o adquisición de grandes cantidades de información digital se puede realizar de diferentes maneras, por ejemplo, Web Scrapping, Web Crawler, pero principalmente se utiliza el método de las API (Applications Program Interface) de cada plataforma (Perriam et al., 2020). Para iniciar con el empleo de este método es

necesario realizar un análisis de las limitaciones de cada API, debido a que muchas plataformas están continuamente cambiando sus políticas de seguridad de acceso a los datos. En algunos casos, estos cambios de políticas son más restrictivas por lo que -en este trabajo- se propone una metodología que permita evaluar y rankear las plataformas disponibles considerando varias características con el fin de elegir la plataforma más adecuada para la extracción de datos, y además permitir que este proceso sea escalable.

Las API permiten a terceros la capacidad de consultar y obtener acceso a porciones de información a partir de los datos de usuarios creados al usar y experimentar plataformas digitales. Cada API especifica las reglas mediante las cuales el software se comunica entre sí, articulando qué elementos se pueden consultar, con qué frecuencia y cómo aparecen los resultados (Acker & Kreisberg, 2020). Los desarrolladores e investigadores pueden acceder a las API mediante scripts para recuperar, almacenar y manipular rastros digitales dejados por los usuarios para su respectivo análisis. Parte del atractivo de la investigación basada en las API es que la recopilación, organización, limpieza, preservación, y análisis de los datos se pueden automatizar, logrando que las API sean una herramienta de investigación altamente eficiente (Lomborg & Bechmann, 2014).

Investigaciones como las de (Aakash & Aggarwal, 2020; Amarouche et al., 2015; Nguyen et al., 2019) presentan algunos estudios relacionados a plataformas digitales y que son de utilidad para el análisis de mercado, es el caso de la detección de Hot Topics y Análisis de sentimientos. Para alimentar los modelos se realiza una extracción de datos a través de las API de las diferentes plataformas; sin embargo, estos estudios no se enfocan en la extracción de datos, sino en otros tipos de análisis. Otras investigaciones como (Kim et al., 2016) realizan una extracción de datos de redes sociales para encontrar el conocimiento del mercado comparando las opiniones de los consumidores con el desempeño de ventas de una empresa, donde se demuestra que los datos de las redes sociales contienen conocimiento valioso para obtener una ventaja competitiva. Por último, un estudio presentado en (Xue et al., 2018) utiliza datos extraídos de YouTube para encontrar posibles competidores de IBM y así obtener una ventaja competitiva. Todos estos estudios presentados tienen en común que la extracción de datos es mediante las API y como resultado final, se obtiene conocimiento para hacer un estudio de mercado. No obstante, la extracción de datos se maneja de una manera secundaria y no se cuenta con una metodología o proceso para priorizar las distintas fuentes de información, de forma que ayuden a tomar decisiones más informadas.

A base de lo mencionado, este artículo tiene dos objetivos, i) proponer una metodología para obtener datos de plataformas

digitales que sean de utilidad para un estudio de mercado de un sector determinado que analice qué fuentes de información son relevantes y especifique parámetros que ayuden a crear un ranking de fuentes de información; y, ii) crear una plataforma en la cual las empresas puedan extraer fácilmente información como comentarios, publicaciones, entre otros, de las fuentes de datos que se han determinado en la primera fase. El resto del artículo está organizado de la siguiente manera: en la sección 2 se describe la metodología propuesta para la creación de un ranking de las fuentes de información y la extracción de la información, además de un caso de estudio en el sector textil, y, por último, en esta sección también se explica detalladamente los pasos para la implementación del prototipo realizado. Finalmente, la sección 3 sintetiza las principales conclusiones y trabajos futuros.

Metodología

Para la implementación de una plataforma digital de extracción de datos de un sector definido se realizó dos pasos: i) selección y generación de un ranking de fuentes de información para el sector determinado y, ii) extracción de datos a las fuentes de información mejor rankeadas. En la Figura 1 se esboza la metodología propuesta, la cual se divide en dos partes: Ranking de fuentes de información y Extracción de datos. A continuación, se explica cada parte de la metodología con su respectivo proceso.

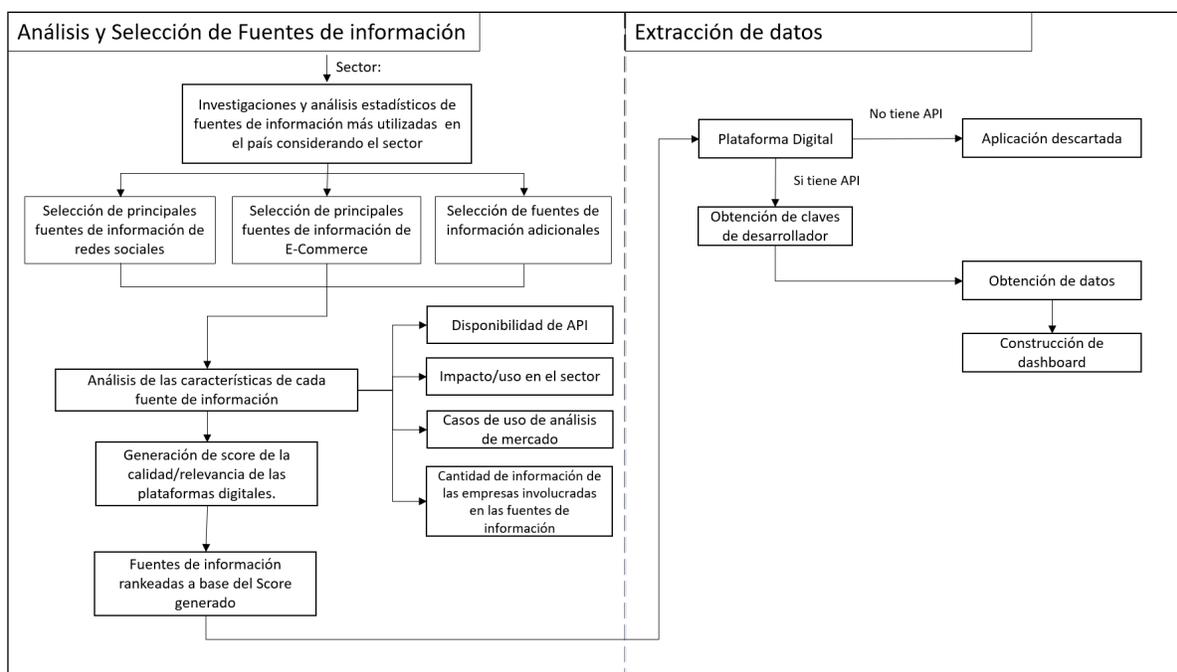


Figura 1
Metodología utilizada para la extracción de información

En el análisis y selección de fuentes de información se eligió el sector a estudiar. Posteriormente, se efectuó una selección de fuentes de información utilizando: i) investigaciones relacionadas y ii) análisis estadísticos digitales de empresas que presentan estudios como: plataformas más visitadas en el país, medios digitales mayormente utilizados en un sector, páginas propias de empresas más visitadas, etc. Luego de esta primera selección se realizó una clasificación de fuentes de información en redes sociales, sitios Web de comercio electrónico y otras fuentes adicionales. Posteriormente, se desarrolló el análisis con mayor profundidad sobre cada fuente de información considerando los siguientes parámetros: i) si tuvo disponibilidad de una API; ii) el impacto/uso en el sector analizado; iii) caso de uso de análisis de mercado; y iv) la cantidad de información disponible de las empresas involucradas en el proyecto. Por último, se generó un score (calificación) final para cada plataforma donde cada parámetro considerado tuvo un peso de acuerdo a la importancia en este estudio para finalmente obtener la tabla con un ranking de las fuentes de información seleccionadas.

El segundo paso fue la extracción de datos de las fuentes de información seleccionadas siguiendo los siguientes pasos: i) se realizó la obtención de las claves de desarrollador o token de acceso de cada API a utilizar; ii) luego se implementó un programa utilizando las

claves y se procedió a la obtención de los datos; y por último iii) se derivó a integrar todos los datos extraídos en una plataforma.

Caso de estudio

La aplicación de la metodología para la extracción de información se realiza con dos empresas MiPymes del sector textil pertenecientes al proyecto “Incorporating sustainability concepts to management models of textile Micro, Small and Medium Enterprises (SUMA)”. Según Karen Lucero en la Revista Gestión menciona que el sector textil es muy importante para el país debido a que representa 5,9% del sector industrial y aporta cerca de un punto porcentual al PIB (0,8%), pero debido a la pandemia este sector disminuyó sus ventas al menos un 36% en 2020 (Karen Lucero, 2021), motivo por el cual se selecciona este sector para analizar en este caso de estudio.

Análisis y Selección de Fuentes de Información.

Primero, se seleccionan las principales fuentes de información de acuerdo a investigaciones y/o análisis estadísticos para este sector dentro del contexto ecuatoriano. Participan estudios realizados por Similar Web, SemRush, Kepios y también la Cámara Ecuatoriana de Comercio Electrónico. Como resultado de este análisis se obtiene que las redes sociales más visitadas en Ecuador son Facebook, Instagram, Twitter y YouTube; los sitios web E-Commerce más visitadas en Ecuador son Mercado Libre y OLX, y se determina que el buscador preferido es Google.

Segundo, se toma en cuenta cuatro investigaciones en distintas provincias de Ecuador, y una investigación de Bogotá-Colombia, donde en cada uno se muestra la importancia de cada plataforma digital para el sector textil en el análisis de mercado.

- El primer estudio considerado se presenta en (Díaz, 2018), donde se realiza un análisis para determinar qué tanto utilizan las redes sociales las microempresas del sector textil. En este estudio se realizan entrevistas a 29 microempresas del sector textil en la ciudad de Quito, donde se determina que las fuentes de información más usadas por las microempresas son: Facebook, Google, Twitter, YouTube e Instagram, en ese orden. Por su parte en (Árcos;Valverde, 2016) se realiza una investigación para analizar la incidencia del uso de redes sociales aplicado en el sector textil en la ciudad de Guayaquil en 2016. En este estudio se realizan entrevistas a 385 clientes del sector textil, donde se dispone que las fuentes de información más usadas por los clientes son Facebook y Twitter.

- El segundo estudio considerado se presenta en (Muñoz Bustos, 2019), donde se realiza una investigación de una empresa del sector

textil de la provincia de Tungurahua, la cual ofrece ropa interior masculina y femenina. Para este estudio los autores realizan una encuesta a 168 personas de diferentes ciudades como son Quito y Guayaquil, donde el objetivo principal es determinar los medios eficaces para realizar publicidad y encontrar posibles clientes. Después de implementar las encuestas y tabular los resultados se encuentra que las principales fuentes de información para los clientes son las redes sociales Facebook, YouTube, Instagram y Twitter, en ese orden. Además, se destaca que el medio preferido por las personas para comunicarse es Facebook.

- El tercer estudio considerado se presenta en (Avila, 2019), aquí se realiza un estudio sobre la comercialización de productos haciendo uso del comercio electrónico y su incidencia en las Pymes del sector textil de Atuntaqui provincia de Imbabura. Para este estudio se toma una muestra de 163 pequeñas y medianas empresas de la ciudad, a las cuales se le implementa una encuesta sobre la utilización de herramientas de comercio electrónico. En este estudio se concluye que las Pymes que tienen presencia en redes sociales consideran que sus productos son vendidos en un nivel bajo alcanzando un 55,3% y en nivel medio en 42,16% por lo que consideran que las redes sociales son de gran utilidad para las empresas. Además, efectúa una prueba de chi-cuadrado y se determina que existe una relación entre las ventas de redes sociales y un nivel bajo de interactividad (nivel I, etapa II del modelo eMICA -Extended Model of Internet Commerce Adoption-), por lo tanto, al implementar estas plataformas las ventas aumentaron. Del mismo modo se realizan con la página de E-Commerce y determinan que las Pymes que tienen implementado página de E-Commerce generan más ventas y por lo tanto están dispuestas a seguir invirtiendo en esta tecnología. Por último, este estudio menciona la importancia de las redes sociales, pero no dice cuáles son las más importantes, lo mismo sucede con el E-Commerce.
- El cuarto estudio se presenta en (Neira, 2017), donde se realiza un análisis de la confianza e intención de compra en consumidores a través de redes sociales y comercio electrónico en el sector textil en Bogotá - Colombia. Se desarrolla una encuesta a 384 personas, donde el 32,6% de los encuestados usan con más frecuencia Instagram, 30,2% Facebook, 27,1% YouTube y 7,8% Twitter.

- Y el último estudio considerado se presenta en (Yánes, 2017), donde se presenta una estrategia de marketing digital para la empresa textil Andelas Cía. Ltda. de la provincia del Tungurahua. En este estudio se toma una muestra de 137 clientes potenciales conformado por distribuidores y confeccionistas al por mayor y menor, a los cuales se aplica una encuesta la que tiene como objetivo conocer hábitos, comportamientos, gustos y preferencias dentro del entorno digital. Como resultados respecto a la importancia y el uso que le dan a las redes sociales son: Facebook con una calificación de 32,8, YouTube con 27, y Twitter con el 14,6 sobre 35.

Tercero, se realiza un análisis de las API de cada fuente de información seleccionados considerando: i) la información disponible que se puede extraer de cada fuente y la personalización de la información obtenida, ii) las restricciones/limitaciones de la API ofrecida por la empresa, iii) las características adicionales de la API, por ejemplo, es necesario el uso del token de desarrollador y cuál es el proceso para obtenerlo.

Cuarto y última característica considerada es la información disponible de las empresas involucradas en las diferentes plataformas, donde se realiza una investigación de la cantidad de información disponible de cada empresa en las fuentes de información seleccionadas, se toma en cuenta si la empresa posee o no una cuenta en las distintas redes sociales y si tiene presencia o no en páginas de comercio electrónico o buscadores como Google.

Con todas estas características mencionadas se procede a calcular un score (calificación a una fuente de información) final para realizar el ranking. Cabe mencionar que para el caso de Google se logra identificar una herramienta llamada Google Trends, la cual muestra las búsquedas más frecuentes que se han realizado, y al contar con una API que se puede integrar fácilmente a cualquier plataforma, se la considera en este estudio. Se detectan plataformas en las cuales su API no proporciona información de utilidad, es el caso de OLX. El score dado a cada plataforma se presenta en la Tabla 1.

Tabla 1
Cálculo del Score de la Calidad/Relevancia de las Fuentes de Información

PLATAFORMA	IMPACTO/USO EN EL SECTOR TEXTIL (SOBRE 40)		CASOS DE USO DE ANÁLISIS DE MERCADO (SOBRE 20)		CANTIDAD DE INFORMACIÓN DE LAS EMPRESAS DEL PROYECTO (SOBRE 10)		ACCESIBILIDAD DE LOS DATOS MEDIANTE API (SOBRE 30)		SCORE FINAL (SOBRE 100)
	CALIFICACIÓN	PESO	CALIFICACIÓN	PESO	CALIFICACIÓN	PESO	CALIFICACIÓN	PESO	
Facebook	9,75	4	7,5	2	10	1	5	3	70
Instagram	6,75	4	6,25	2	8,25	1	5	3	62
Twitter	7	4	8,75	2	0	1	8,75	3	71
YouTube	7,75	4	6,25	2	0	1	10	3	70
Mercado Libre	0	4	6,25	2	0	1	6,25	3	31
Google Trends	8	4	10	2	10	1	7,5	3	80

Para el cálculo del score, cada parámetro tiene una escala de 0 a 10 puntos y con un grupo de expertos se asigna un peso a cada característica seleccionada de acuerdo a la importancia en este estudio, donde la suma total da un valor de 100. La metodología utilizada para cada uno fue la siguiente:

- Para el cálculo del “Impacto/uso en el sector textil” se toman en cuenta investigaciones realizadas en Ecuador sobre el uso de fuentes de información en el sector textil. En cada estudio se tiene un score máximo de 10 y para obtener el score final se realiza un promedio de los 5 estudios por fuente de información, donde se toma en cuenta el orden de importancia de cada fuente de información que presenta cada estudio. El primer lugar obtiene el máximo puntaje, el segundo lugar se penaliza 1.25 puntos, el tercer lugar 2.5 puntos y así sucesivamente. Para el caso de Google que solo se menciona en un estudio, se realiza una penalización de 0.1875 puntos por cada estudio en el que no se encuentra.

- Para el cálculo del segundo parámetro “Caso de uso de análisis de mercado”, de igual manera que el cálculo del “Impacto/uso en el sector textil”, se ordena de acuerdo al número de estudios encontrados que se tomen en cuenta en el análisis de mercado. La plataforma que está en el primer lugar es la que mayor puntaje tiene, el segundo lugar se penaliza con un valor de 1.25 que es el mismo del primer parámetro, el tercer lugar se penaliza con 2.5 y así sucesivamente.

- El parámetro “cantidad de información” considera la disponibilidad de información que hay de las empresas involucradas en la fuente de datos, y de acuerdo al número de seguidores y “me gusta” en las redes sociales se asigna un valor. También se toma en cuenta una penalización de 1.25 proporcional al resto de parámetros. Cabe mencionar que con respecto a Google se toma un valor de 10 puntos si tiene presencia caso contrario se toma un valor igual a 0.

- Con respecto a la “accesibilidad de los datos mediante API” se toma en cuenta la información disponible, restricciones/limitaciones, características de la API, personalización que se puede realizar y con un grupo de expertos se determina cuan fácil es el acceso a los datos de cada fuente de información. De igual forma, se toma como máximo puntaje a la plataforma que es más fácil el acceso a los datos y al resto se penaliza con un valor de 1.25 como en los casos anteriores.

Después de realizar el análisis se obtiene el siguiente resultado ver Figura 2, se determina que las 4 plataformas que mayor beneficio ofrecen a este estudio son: Google Trends, Facebook, YouTube y Twitter. Adicionalmente, se realiza otro análisis sin tomar en cuenta el parámetro “Accesibilidad de los datos mediante API” debido a que se puede acceder a los datos mediante otras técnicas como Web Scraping. A base de este análisis se determina que las 4 plataformas

que mayor beneficio ofrecen son: Facebook, Google, Instagram y Twitter. Con respecto a las limitaciones de las API, la de Twitter tiene una limitación en el número de consultas por día y se pueden extraer datos de los últimos 7 días únicamente. De ser necesario más datos, se tiene que realizar un pago. Así también, la API de YouTube tiene una limitación en el número de consultas por día que puede recibir, aunque para la mayoría de aplicaciones, este límite es casi imposible de superar.

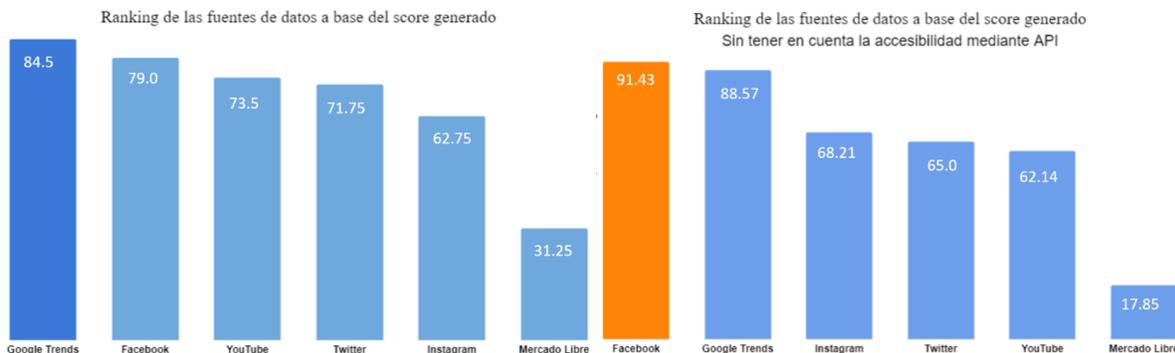


Figura 2
Ranking de las plataformas digitales

Desarrollo de la Plataforma

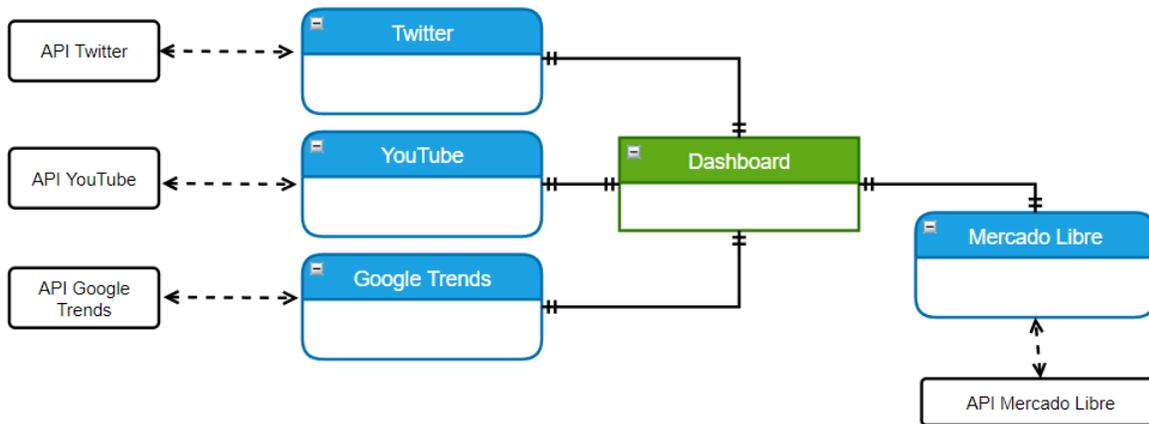


Figura 3
Arquitectura de la Aplicación

Para presentar de manera visual la información extraída: total de registros, número de datos extraídos de cada plataforma, se procede a implementar una aplicación web tipo dashboard de nombre Dashboard SUMA. Esta aplicación se desarrolla con el Framework Django y se utiliza la librería Chart.js para los gráficos. La plataforma cuenta con varios subsistemas, donde cada uno extrae datos de una

plataforma digital. La arquitectura implementada se presenta en la Figura 3.

En la Figura 4 se presenta la estructura de la aplicación de Django y en la Tabla 2 se presenta una descripción de los componentes de la aplicación.

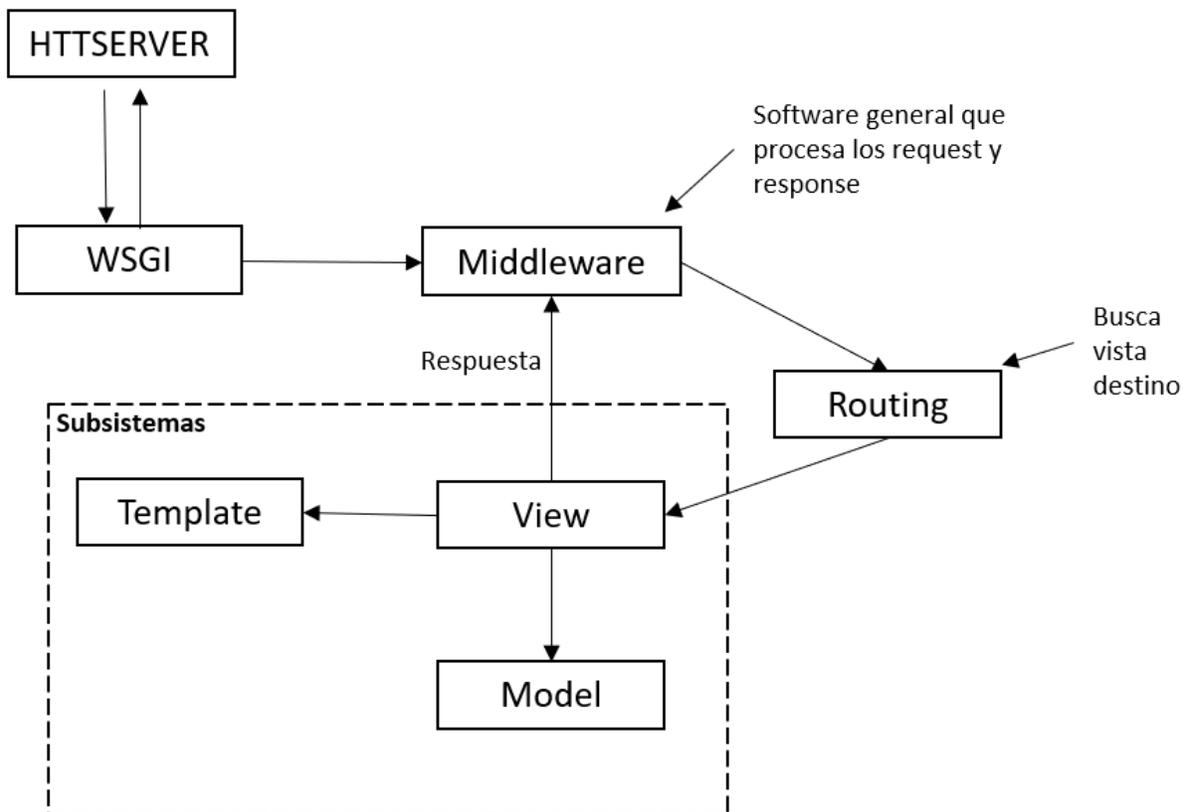


Figura 4
Estructura de la Aplicación Django

Tabla 2
Descripción de las Partes de la Estructura de la Aplicación

COMPONENTE	DESCRIPCIÓN
HttpServer	Es el servidor que escucha las peticiones. Entre los más conocidos están nginx, apache, iis.
WSGI	Modelo de interacción entre servidores web y aplicaciones.
Middleware	Software python dentro de django, definido en settings y que permite alterar la request y/o la response.
Routing	Módulo de django que usa las definiciones en el fichero urls.py, para seleccionar la vista.
View	Código python que recibe la request y la procesa y usando las templates genera la respuesta.
Model	La vista usa las clases que representan el modelo de datos.
Template	En conjunción con la vista, describe la salida, que será principalmente html. Usará datos proporcionados por la vista, que vendrán de la base de datos y las consulta a la API.

Finalmente, se presenta una vista para cada fuente de información donde el usuario puede ingresar su tema de interés y ejecutar la aplicación, en ese momento se realiza la extracción de datos mediante API y se muestra en una tabla con la disponibilidad para poder descargar los datos en formato csv. En las Figura 5 se muestra la página principal de la aplicación con métricas como el número de datos extraídos por cada plataforma, el número total de datos extraídos y un historial de las búsquedas que se han realizado, en la Figura 6 se muestra la interfaz para extraer datos de Twitter y en la Figura 7 se presentan la interfaz con los datos extraídos. De la misma manera que las interfaces de Twitter son para las otras fuentes de información.

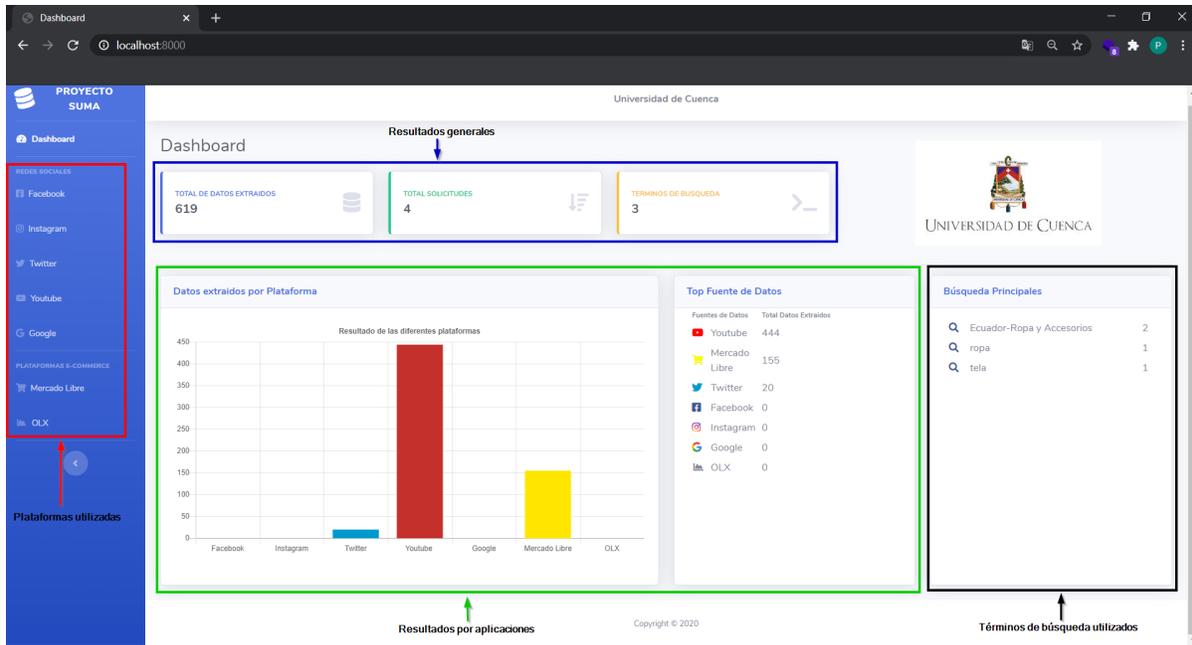


Figura 5
Página Principal de la Aplicación

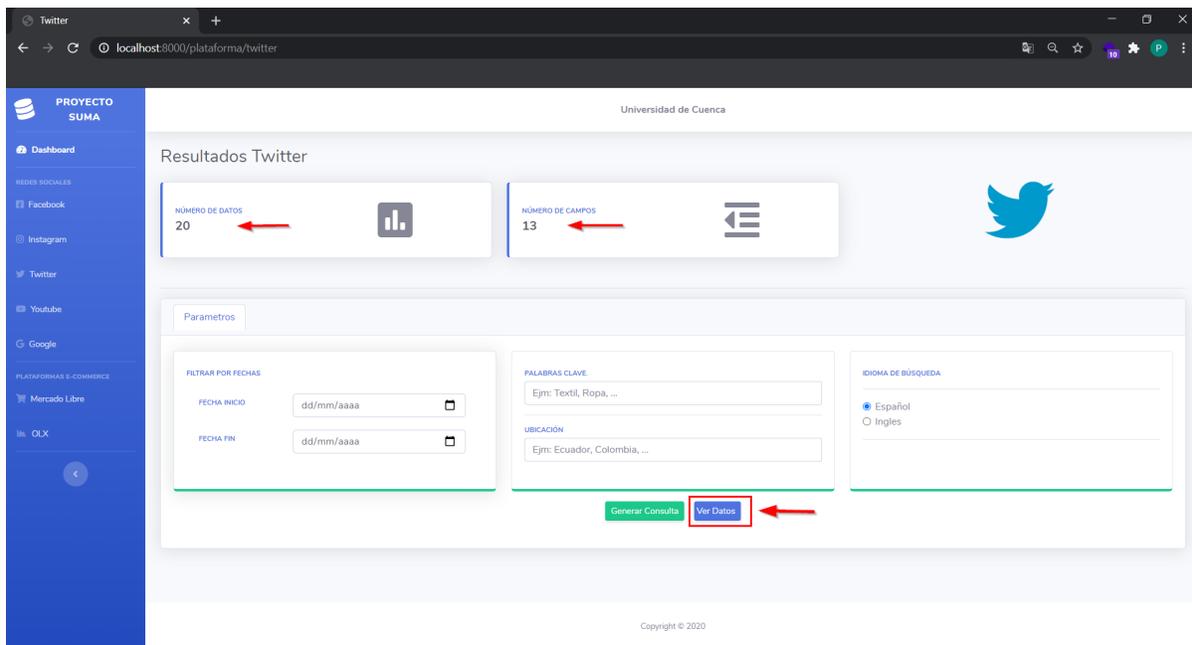


Figura 6
Página para Consulta de Datos de Twitter

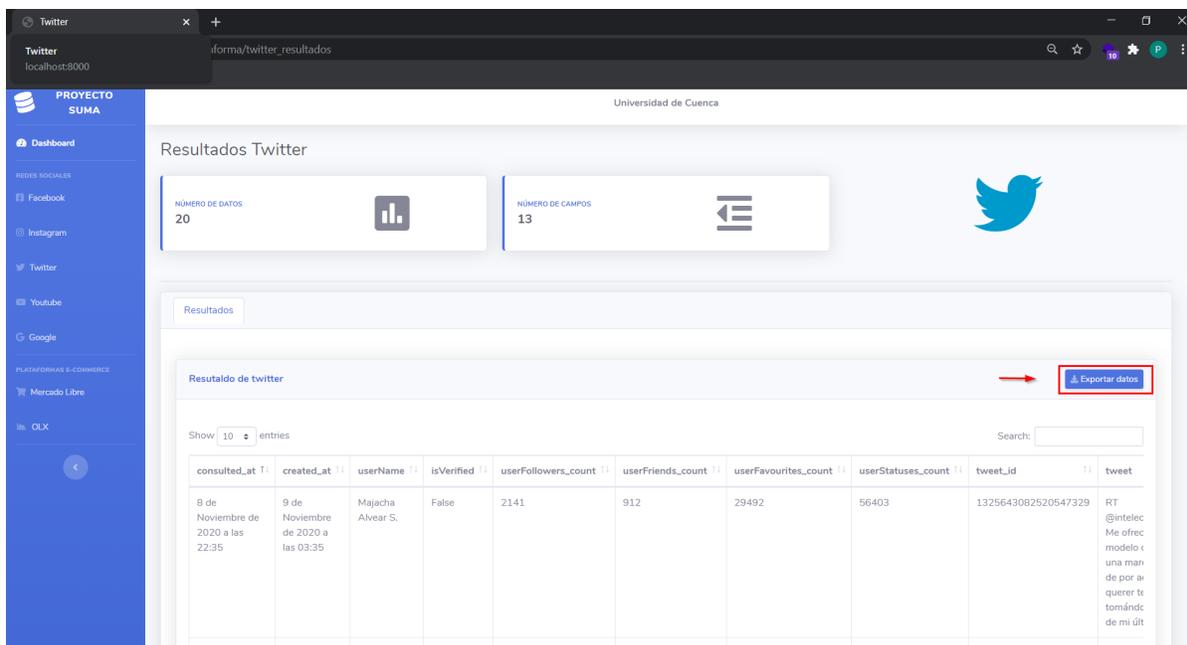


Figura 7
Página de Datos Obtenidos de Twitter

Discusión

A base del análisis que se realiza en esta investigación, se puede determinar que en la actualidad las principales fuentes de información para el sector textil en el contexto de análisis de mercado son las redes sociales. Estas fuentes de información son muy importantes para el análisis de mercado debido a que en dichas plataformas los usuarios pueden expresar comentarios tantos positivos como negativos de un determinado producto. Así, las empresas pueden realizar sus análisis utilizando datos generados por el usuario para luego obtener resultados más cercanos a la realidad sobre éxito o fracaso de sus productos. Sin embargo, debido a problemáticas recientes por mal uso de datos de estas plataformas, cada vez se vuelve más complicado extraer dichos datos ya que las empresas están obligadas a proteger la información y los usuarios confían en ellas. Por lo tanto, el presente estudio presenta un análisis para calificar y priorizar las fuentes de información disponibles considerando la limitante mencionada y también sin tomar en cuenta dicha limitante para que así el lector pueda diferenciar las fuentes de información importantes para el sector. Todas las investigaciones revisadas se centran en el análisis de los datos de una determinada plataforma y que previamente se debe realizar una investigación para la selección de las mismas, pero en cada estudio revisado no presentan el método utilizado por lo tanto esta metodología presentada ayuda a las empresas en esta selección de

plataformas debido a que permite seleccionar una fuente de información considerando varios parámetros de acuerdo al sector a analizar. Por último, es importante mencionar que el método de las API es una manera de extracción de datos, pero en la actualidad existen muchas otras técnicas que pueden ser implementadas y analizadas en estudios posteriores.

Conclusiones

La metodología propuesta para generar el ranking de fuentes de información ayuda a obtener las plataformas más informativas para la extracción de datos regionales y provee datos de calidad para que las empresas puedan realizar un análisis con los mismos.

El ranking realizado, en el caso de estudio se determina que las 4 plataformas que mayor beneficio ofrecen para esta investigación son: Google Trends, Facebook, YouTube y Twitter. Adicionalmente, se realiza otro análisis sin tomar en cuenta el parámetro “Accesibilidad de los datos mediante API” debido a que se puede acceder a los datos mediante otras técnicas como Web Scraping por lo que también es una alternativa para obtención de datos. A base de este análisis, se determina que las 4 plataformas que mayor beneficio ofrecen son: Facebook, Google, Instagram y Twitter.

Las redes sociales Facebook e Instagram tienen un ranking alto pero debido a la implementación de sus nuevas políticas de acceso a la información altamente restrictivas no puede realizar la extracción de datos. Únicamente se puede obtener datos propios de la cuenta del usuario o página administrada como publicaciones, comentarios de la publicación y reacciones de la publicación. En las plataformas que tienen API y no tienen grandes limitaciones en sus políticas de acceso a la información, se realiza la extracción de datos sin ningún inconveniente. Inclusive, se logra extraer información adicional en forma de metadatos que pueden ser de interés para estudios futuros como la determinación de lugares donde existen problemas como los productos, lugares determinados donde existen algunas necesidades, entre otros.

Como trabajo futuro, se plantea la necesidad de automatizar el proceso para la valoración de los pesos para cada característica y, de esta manera, facilitar la aplicación de esta metodología a cualquier sector. Así también, analizar a profundidad las API de otras plataformas, por ejemplo, Pinterest para determinar qué tipo de información se puede extraer y luego determinar su utilidad para el sector de estudio. Finalmente, considerando las limitaciones de acceso a Facebook e Instagram, se plantea explorar otras opciones para poder extraer datos de dichas plataformas, ya que son estas las que más se utilizan en la región.

Agradecimientos

Este estudio se enmarca en el proyecto de investigación “Incorporating Sustainability concepts to management models of textile Micro, Small and Medium Enterprises (SUMA)”, apoyado por el Consejo Interuniversitario Flamenco (VLIR) y el Vicerrectorado de Investigación de la Universidad de Cuenca (DIUC).

Referencias

- Aakash, & Aggarwal, A. G. (2020). An Analysis of EWOM Text that Contribute to EWOM Helpfulness. In *Lecture Notes in Electrical Engineering* (Vol. 601, pp. 893–900). Springer. https://doi.org/10.1007/978-981-15-1420-3_97
- Acker, A., & Kreisberg, A. (2020). Social media data archives in an API-driven world. *Archival Science*, 20(2), 105–123. <https://doi.org/10.1007/s10502-019-09325-9>
- Amarouche, K., Benbrahim, H., & Kassou, I. (2015). Product Opinion Mining for Competitive Intelligence. *Procedia Computer Science*, 73, 358–365. <https://doi.org/10.1016/j.procs.2015.12.004>
- Árcos;Valverde, D. A. (2016). “Incidencia del uso de redes sociales aplicado en las estrategias de marketing en el sector textil de la ciudad de guayaquil año 2016.”<http://dspace.ups.edu.ec/handle/123456789/14886>
- Avila Fuentes, L. F. (2019). *ARTICULO CIENTÍFICO: Comercialización de productos mediante el uso de herramientas de comercio electrónico y su incidencia en las PYMES del sector textil de Atuntaqui provincia de Imbabura*. <http://181.39.139.68:8080/handle/123456789/249>
- Barrientos Felipa, P. (2017). Marketing + internet = e-commerce: oportunidades y desafíos. *Revista Finanzas y Política Económica*, 9(1), 41–56. <https://doi.org/10.14718/revfinanzpolitecon.2017.9.1.3>
- Diaz Agila, Y. M. (2018). “Análisis De Las Estrategias De Marketing Digital Aplicadas Por Las Microempresas En El Sector Textil De La Provincia De Pichincha.”<http://dspace.ups.edu.ec/handle/123456789/15305>
- E-commerce. (2020). *DC0-Información pública E-Commerce en Ecuador*. <https://www.pwc.ec/es/publicaciones/assets/e-commerce-ecuador.pdf>
- Ecommerceday. (2017). *Estudio de Comercio Electrónico en el País durante el eCommerce Day Ecuador 2017*. 2–4. <http://www.ecommerceday.ec/2018/2017/11/09/se-presento-el-primer-estudio-de-comercio-electronico-en-el-pais-durante-el-ecommerce-day-ecuador-2017/>
- Karen Lucero. (2021, May 2). El sector textil, un puntal de la industria que busca levantarse | Gestión. Economía y Finanzas. <https://www.revistagestion.ec/economia-y-finanzas-analisis/el-sector-textil-un-puntal-de-la-industria-que-busca-levantarse>
- Kim, Y., Dwivedi, R., Zhang, J., & Jeong, S. R. (2016). Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5. *Online*

Information Review, 40(1), 42–61. <https://doi.org/10.1108/OIR-03-2015-0068>

Lomborg, S., & Bechmann, A. (2014). Using APIs for Data Collection on Social Media. *Information Society*, 30(4), 256–265. <https://doi.org/10.1080/01972243.2014.915276>

Muñoz Bustos, D. A. (2019). *Sistema organizacional aplicado al marketing digital en el sector textil de ropa interior de la provincia de Tungurahua*. <https://repositorio.uta.edu.ec:8443/jspui/handle/123456789/29960>

Neira, D. A. (2017). *Las redes sociales y la confianza del consumidor como factores en la intención de compra a través de canales de comercio electrónico en el sector moda colombiano*. <https://repositorio.unal.edu.co/handle/unal/60094>

Nguyen, S., Ngo, B., Vo, C., & Cao, T. (2019). Hot Topic Detection on Twitter Data Streams with Incremental Clustering Using Named Entities and Central Centroids. *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, 1–6. <https://doi.org/10.1109/RIVF.2019.8713730>

Perriam, J., Birkbak, A., & Freeman, A. (2020). Digital methods in a post-API environment. *International Journal of Social Research Methodology*, 23(3), 277–290. <https://doi.org/10.1080/13645579.2019.1682840>

Tripathi, G., & S, N. (2015). Feature Selection and Classification Approach for Sentiment Analysis. *Machine Learning and Applications: An International Journal*, 2(2), 01–16. <https://doi.org/10.5121/mlaj.2015.2201>

Xue, Y., Zhou, Y., & Dasgupta, S. (2018). Mining competitive intelligence from social media: A case study of IBM. *Proceedings of the 22nd Pacific Asia Conference on Information Systems - Opportunities and Challenges for the Digitized Society: Are We Ready?, PACIS 2018*. <https://pennstate.pure.elsevier.com/en/publications/mining-competitive-intelligence-from-social-media-a-case-study-of>

Yánes González, R. M. (2017). *Marketing digital para el sector textil de la provincia de Tungurahua: Caso Andelas CIA. LTDA*. [Universidad Técnica de Ambato. Facultad de Ciencias Administrativas. Carrera de Marketing y Gestión de Negocios.]. <https://repositorio.uta.edu.ec:8443/jspui/handle/123456789/26922>

AmeliCA

Disponible en:

<https://portal.amelica.org/ameli/journal/844/8445114009/8445114009.pdf>

[Cómo citar el artículo](#)

[Número completo](#)

[Más información del artículo](#)

[Página de la revista en portal.amelica.org](#)

AmeliCA

Ciencia Abierta para el Bien Común

Ángel Patricio Fajardo Cárdenas, Néstor Ariel Bravo Chuqui,
Andrés Vinicio Auquilla Sangolqui,
Paúl Fernando Vanegas Peña

Plataforma para Análisis de Mercado a través de Datos de Redes Sociales

Platform for Market Analysis through Social Network Data

Revista Tecnológica ESPOL - RTE

vol. 33, núm. 2, Esp. p. 134 - 146, 2021

Escuela Superior Politécnica del Litoral, Ecuador

rte@espol.edu.ec

ISSN: 0257-1749

ISSN-E: 1390-3659

DOI: <https://doi.org/10.37815/rte.v33n2.839>



CC BY-NC 4.0 LEGAL CODE

Licencia Creative Commons Atribución-NoComercial 4.0 Internacional.