



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

**Métodos de análisis para identificar público objetivo que
consume productos de artesanía a través de Big Data Piura -2019**

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
Ingeniero de Sistemas

AUTOR:

Carrasco Cruz, Luis Germán (ORCID: 0000-0002-0000-7590)

ASESOR:

Mg. More Valencia, Rubén Alexander (ORCID: 0000-0002-7496-3702)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

PIURA-PERÚ

2020

Dedicatoria

Esta tesis es dedicada en primer lugar a Dios por hacer que esta primera meta sea posible.

A mis padres y hermano por acompañarme en este camino dándome su apoyo incondicional.

A mis abuelos, tíos, primos y demás familiares que han sido testigos de este largo proceso.

Agradecimiento

Agradezco a Dios y familiares en el cielo que guían mi camino.

Agradezco mis padres y hermano por el esfuerzo hecho para poder lograr este objetivo.

A mis abuelos, tíos, primos y demás familiares que han sido parte importante de esta primera meta.

Índice de Contenidos

Carátula.....	i
Dedicatoria.....	ii
Agradecimiento	iii
Índice de Contenidos	iv
Índice de tablas.....	v
Índice de gráficos y figuras	vi
Resumen.....	vii
Abstract	viii
I.INTRODUCCIÓN	1
II. MARCO TEÓRICO	3
III. METODOLOGÍA	16
3.1 Tipo y Diseño de la Investigación.....	16
3.2 Variables y Operacionalización	17
3.3 Población y Muestra.....	18
3.4 Técnicas e Instrumentos de recolección de datos	19
3.5 Método de Análisis de Datos	21
3.6 Aspectos Éticos.....	24
IV. Resultados.....	25
V. DISCUSIÓN.....	35
VI. CONCLUSIONES.....	37
VII. RECOMENDACIONES	38
REFERENCIAS	39
ANEXOS	41

Índice de tablas

Tabla 1: Operacionalización de Variables	17
Tabla 2: Métodos de extracción y análisis de datos para Facebook y Twitter	25
Tabla 3: Resultados de primer objetivo	27
Tabla 4: Resultados de primer objetivo	27
Tabla 5: Resultados de método descriptivo en Facebook	30
Tabla 6: Resultados en cada clúster	30
Tabla 7: Resultados de método descriptivo en Twitter	31
Tabla 8: Resultados en cada clúster	31

Índice de gráficos y figuras

Figura 1: Resultados de extracción de datos de Twitter	26
Figura 2: Resultados de extracción de comentarios de Facebook	26
Figura 3: Grafico de los tweets usuarios	28
Figura 4: Grafico de comentarios de facebook.....	29
Figura 5: Clasificación de Tweets.....	32
Figura 6: Palabras positivas de los tweets	33
Figura 7: Clasificación de comentarios de Facebook	34
Figura 8: Palabras positivas de comentarios de Facebook.....	34

Resumen

Esta investigación, nació con la finalidad de dar un impulso a la artesanía piurana con la ayuda de la tecnología, se buscó en primer lugar métodos que permitan extraer datos de las redes sociales en este caso se tomó en cuenta Facebook y Twitter, para luego describir que nivel de interés tienen los usuarios y que tan útil puede ser un modelo de minería de datos.

El diseño de esta investigación fue no experimental ya que sólo se basó en identificar y describir los métodos de extracción y procesamiento de los datos de las redes sociales ya mencionadas, y llegó hasta un nivel descriptivo de tipo transversal. La población que se tomó en cuenta para llevar a cabo esta tesis estuvo formada por los usuarios que interactuaron en Facebook y Twitter entre el 1 y 10 de octubre del 2019 y la unidad de análisis son solo los usuarios que hablaron de artesanía en esas fechas.

Para reunir los datos se trabajó con un análisis documental como técnica y como instrumento una ficha de registro de datos, donde se analiza y compara la extracción y el procesamiento de los datos de las redes sociales.

Se obtuvo como resultados que para la extracción de datos de Facebook es más complejo que para la extracción que se hace en Twitter ya que se generan de manera automática y el mayor nivel de interés sobre la artesanía se obtuvo de los tweets extraídos.

En general es factible hacer uso del Big Data con una gran cantidad de datos para hacer una mejor predicción sobre algún tema en específico para que luego se genere un aprendizaje automático en la máquina y trabaje por si solo con los nuevos datos.

Palabras Clave: Big Data, Extracción de Datos, Procesamiento de Datos, Predicción.

Abstract

This research was born with the purpose of giving a boost to Piura's crafts with the help of technology. Methods were searched in the first place, to extract data from social networks, in this case, Facebook and Twitter were taken into account and then describe what level of interest users have and how useful a data mining model can be.

The design of this research was non-experimental since it was only based on identifying and describing the methods of the extraction and processing of data from the aforementioned social networks. And it reached a descriptive level of transversal type. The population that was taken into account to carry out this thesis was formed by users who interacted on Facebook and Twitter between October 1 and 10, 2019 and the users who spoke about crafts on those dates are the unit of analysis.

To collect the data, a documentary analysis was used as a technique and as an instrument a data registration form, where the extraction and processing of data from social networks are analyzed and compared.

It was obtained as result that in order to the extraction of data from Facebook it is more complex than for the extraction that is done on Twitter because it is generated automatically and the highest level of interest about crafts was obtained from the extracted tweets.

In general, it is feasible to make use of Big Data with a large amount of data to make a better prediction on a specific topic to generate automatic learning in the machine and it works by itself with the new data.

Keywords: Big Data, Data Extraction, Data Processing, Prediction.

I.INTRODUCCIÓN

La artesanía es una de las actividades principales que se realizan en el Perú, la Oficina Comercial del Perú en Beijing emitió una ficha de mercado producto de Artesanías, donde habla sobre la producción y exportación de artesanías de China que junto con la India son una importante proporción de la producción de artesanía en el mundo, por otra parte dice que los consumidores chinos optan por comprar o adquirir artesanías de marcas emergentes con estilos propios y únicos pero debido a la demanda de sus productos gran parte es producida por máquinas y solo usan el trabajo manual para los retoques finales.

En Perú el Ministerio de Comercio Exterior y Turismo mediante una ficha de mercado producto propone algunas ideas para promover el producto y captar nuevos consumidores en los mercados chinos.

Las ideas son la apertura de showrooms (sala de exposición) en hoteles de lujo en alianza con empresas peruanas para entrar poco a poco en el mercado chino.

Obtener más visibilidad de los productos en las redes sociales, blogs y sitios webs para la atracción del público.

Organizar algún tipo de feria o evento para la exhibición de la artesanía peruana.

El uso del big data contribuye en muchos aspectos y si usamos ese análisis en función a las exportaciones de algún tipo de producto, se puede obtener información por ejemplo que país tiene mayor demanda sobre ese producto, que cantidades se han comprado en el último año o que público esté interesado en conocer más de la artesanía.

Debido a la problemática que hay a la hora de exportar o dar a conocer la artesanía de Piura se pretende identificar un método para identificar un público objetivo en las redes sociales que esté interesado en la artesanía, obteniendo como resultado la difusión de la producción piurana a un público objetivo utilizando un análisis profundo de datos y tener ya un contacto directo con el futuro cliente u organización interesada.

Según la realidad problemática se busca la solución mediante la formulación de una pregunta al problema: ¿Cuáles son los métodos que permiten analizar público

objetivo que consuma productos de artesanía a través de big data?, luego se describe los problemas específicos: ¿Qué nivel de predicción tienen los modelos de minería de datos a usar?, ¿Qué tan útil es el modelo de minería de datos para obtener los resultados requeridos?

Luego de analizar las preguntas de la realidad problemática se presenta la justificación del estudio, el cual se justifica de manera científica, considerando que permite explorar la ciencia en cuanto a los grandes volúmenes de datos para hacer la propuesta de un método hacia público objetivo donde se pueda mostrar la artesanía Piurana y obtener un beneficio para la región. Asimismo, se justifica de manera teórica donde se realizó con el propósito de generar una propuesta de método para captar nuevo público objetivo que consuma artesanía Piurana, mediante el análisis de los grandes volúmenes de datos que existen en la web, donde se identificará cuál es el origen de los datos, como es su procesamiento y que modelo seguir para obtener los resultados que se requieren. También se justifica de manera social la cual permite explorar que métodos son efectivos para la toma de decisiones en las diferentes realidades y esto servirá para beneficio de la región en el aspecto económico y social de Piura dando a conocer aún más la producción de artesanía para su futura exportación a nuevos clientes que serán previamente seleccionados, a través de un análisis profundo de datos. Finalmente se justifica de manera personal donde la investigación me permite generar conocimiento sobre el análisis masivo de datos que se realiza para realizar una buena toma de decisiones en algún tipo de realidad que se desea aplicar, además que se realiza con el fin de obtener mi título de Ingeniero de Sistemas.

Siguiendo con la investigación se plantea como objetivo general: Identificar los métodos de extracción de datos: Público objetivo que consuma productos de Artesanía a través de Big Data. En seguida se presenta los objetivos específicos: Describir el nivel de interés de artesanía con el modelo usado en cuanto al análisis de datos y Describir que tan útil es el modelo de minería de datos para obtener los resultados requeridos. Y como hipótesis general se planteó que existen métodos de análisis para público objetivo que consuma productos de artesanía a través de big data.

II. MARCO TEÓRICO

Según Nancy Heredia (2017) propone diseñar un plan de exportación de artesanías mexicanas, en México las artesanías dependen mucho del turismo y para dar a conocer estas artesanías no solo se les debe mostrar a los turistas sino llevarlas a otros países, en México 12 millones de habitantes se dedican a la producción de artesanías y 600,000 comercializan su producción mediante una institución gubernamental que diseña y ejecuta políticas de promoción y comercialización de la actividad artesanal (Fonart) y su propósito es hacer que la economía crezca y se genere empleo, por ello esta puede ser una oportunidad para dar propuestas y estrategias de comercialización internacional, con el objetivo de mejorar la situación económica y social, dando a conocer sus artesanías en otra parte del mundo. (Heredia 2018)

En Ecuador se hizo un estudio de rentabilidad en cuanto a la exportación de artesanías, debido a la búsqueda constante de nuevos ingresos se quiere aprovechar las ventajas que tiene Ecuador en cuanto a la producción de artesanía de tipo textil, esto ayudaría al aspecto social y económico por qué incrementaría la oferta de trabajo y, además, se daría a conocer al mercado extranjero su cultura artesanal. Este proyecto sugiere una mejora en cuanto a la forma en que se comercializan los productos de manera local, la idea es incrementar la comercialización a nivel internacional. (López et al. 2010)

En Loreto se hizo un análisis sobre las exportaciones de artesanía en un período de tiempo (2009-2013) la artesanía de exportación en el Perú es importante por qué es una fuente de empleo para los pobladores de las zonas rurales y urbanas. Esta exportación a pesar de lo difícil que pueda ser se puede desarrollar y alcanzar niveles de crecimiento y comercialización a los mercados del exterior, lo que conlleva a la necesidad de poder analizar la evolución que ha tenido este tipo de comercio en el período de tiempo establecido. (DIANA LLANOS RODRÍGUEZ, 2015)

PROMPERU en el año 2014 muestra un catálogo de las exportaciones y dice que en la región Piura el tipo de exportación no tradicional es la artesanía, los finos trabajos de cerámica de Chulucanas y los tejidos de paja toquilla de Catacaos se destacan por dar la vuelta al mundo y conquistan mercados nacionales y

extranjeros, sin embargo, cada vez más los comerciantes optan por hacer diseños modernos que no se despeguen de sus raíces para dar paso a la modernidad en los diseños y detalles. (PROMPERU, 2014)

En cuanto a la exportación de artesanías en Piura, un informe realizado habla sobre algunos de los problemas de los artesanos a la hora de exportar sus productos, no pueden ofrecer su producción de manera directa, no cuentan con una vitrina digital para la exhibición de sus productos por lo que tienen que contactarse con terceros y están en constante búsqueda de nuevos mercados por ello han optado por diseños más modernos. (Guevara, 2012)

Se encontró el estudio realizado por Diana y Fiorella (2015) titulado “Análisis de las exportaciones de Artesanías de la Región Loreto, Periodo 2009 - 2013”, en el cual se abordó el tema de la exportación en cuanto a las artesanías, para saber cómo fue su desarrollo en los años 2009 y 2013, esto permitió determinar que las exportaciones de artesanía en la región Loreto mostró un comportamiento que variaba. En el año 2010 creció 77.37% pero en el año 2011 disminuyó a -35.46%. Está investigación ayudaría para hacer una selección previa de que productos de artesanía que se comercia más o es más fácil para su comercio. (DIANA LLANOS RODRÍGUEZ, 2015)

Otra investigación es la realizada por Katya Garvich (2017) titulada “PROPUESTA DE ANÁLISIS DE DATOS NO ESTRUCTURADOS PARA GENERAR DECISIONES OPORTUNAS EN LA EMPRESA GMD”, su propósito fue determinar cuál es el impacto al momento de implementar una propuesta de análisis de datos no estructurados, para generar decisiones oportunas cuando se implemente algún proyecto. Esta investigación además se basa en un enfoque cualitativo que busca analizar el impacto en tiempos y costos al realizar una propuesta de datos no estructurados. (Garvich, 2017)

Finalmente, concluyen que gracias a la propuesta de arquitectura de análisis de datos no estructurados ayudado mediante la plataforma de Big Data en IBM, van a permitir una reducción de tiempos en cada proceso automatizado, como son los procesos de extracción, análisis y visualización de la empresa. La postura de esta investigación coincide en algunos puntos con lo que se desea plantear, se hará la

implementación de un análisis de datos para tener en cuenta que es lo que prefiere el consumidor o ver cuál es su comportamiento y llegar a una buena toma de decisiones para impulsar más el comercio de artesanías y tener buenos resultados.

La investigación titulada "Análisis de Sentimientos y predicción de eventos en Twitter" hecha en 2014 por Montesinos García, tiene como objetivo desarrollar un método que extraiga la opinión de usuarios chilenos en Twitter, donde los tweets sean clasificados en negativos o positivos acerca de los candidatos de dos partidos políticos para al final predecir cual sería el resultado de las elecciones, este tipo de análisis de sentimientos me ayudará a aplicar a los datos recogidos de las redes sociales para luego hacer una clasificación y predicción sobre los usuarios que hablan de artesanía. (Montesinos García 2014)

Miguel Grández en el 2017 hizo una investigación llamada "Aplicación de Minería de Datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales" que tuvo como objetivo general la aplicación de minería de datos a través de un Software que pueda encontrar patrones de consumo en una distribuidora de suplementos nutricionales y luego implementar políticas que ayuden a subir el nivel de ventas, donde aplicó una entrevista a los administradores y vendedores de la tienda e hizo una recolección de datos acerca de las últimas ventas que se realizaron.

Luego de aplicar un modelo de minería de datos llegó a la conclusión que el producto de más demanda tanto para hombres como para mujeres es el NECTAR ISOLATE que contiene proteínas y es más probable que clientes entre 25 y 32 años compren el producto. (Grandez Márquez [2017])

Como teorías relacionadas al tema se tomaron en cuenta algunas que ayudaron en la investigación.

Metodología De Análisis De Datos, es el proceso de examinar la situación de alguna organización con la finalidad de mejorar con la ayuda de métodos y procedimientos adecuados.

Para el análisis de datos se hace uso de la minería de datos que involucra un proceso iterativo e interactivo de búsqueda de modelos, patrones o parámetros y es preciso evaluar la validez, utilidad y simplicidad de los patrones que se obtienen, haciendo el uso de alguna técnica de Minería de Datos, tomando en cuenta que la finalidad es aplicar el conocimiento obtenido en algún tipo de realidad para tomar decisiones a partir de los resultados o administrar la información.

La minería de datos puede ser aplicada a cualquier modelado de datos sobre clientes, todo el proceso debe estar encaminado a soportar una futura toma de decisiones estratégicas que es el principal objetivo. Se puede analizar, compra de productos, patrones secuenciales de compra, efectividad de promociones, segmentación de clientes. (Fandiño, 2005)

Primero se debe identificar los datos que sean útiles para los objetivos propuestos y se procede a la minería de datos, se seleccionan herramientas y técnicas adecuadas, tras ese proceso llega el análisis de resultados y se obtiene el conocimiento deseado.

En la aplicación de minería de datos se consideran cinco etapas:

La primera etapa de Selección, se crea un conjunto de datos objetivo, o se enfoca en un subconjunto de variables o muestras de datos, en las que se debe realizar el descubrimiento. En la segunda etapa de Procesamiento previo, se hace una limpieza y pre-procesamiento de datos de destino para obtener datos consistentes. La tercera etapa de Transformación, se empieza la transformación de los datos usando métodos de reducción de dimensionalidad o transformación. En la cuarta etapa de Minería de Datos, consiste en la búsqueda de patrones de interés, dependiendo del objetivo de minería de datos (por lo general, la predicción). Y en la quinta etapa de Interpretación / Evaluación, consiste en la interpretación y evaluación de los patrones.

El Método Predictivo es utilizado para la minería de datos y se basa en entrenar un modelo o un método por medio de diferentes datos para obtener una predicción del resultado futuro, se hace uso de la máquina de aprendizaje como las redes neuronales o como los modelos estadísticos.

Un ejemplo claro son los datos de ingeniería, como los datos provenientes de sensores, instrumentos y sistemas conectados al mundo real o se pueden incluir los datos de las transacciones de los sistemas empresariales para tomar decisiones basadas en estos datos.

El Método Descriptivo es un método no supervisado que forma grupos de datos de manera rápida y usando como técnica el análisis de clúster que crea los grupos en función a la naturaleza de los datos, con este análisis descriptivo se pretende aumentar la comprensión de los datos disponibles.

Por otro lado, el Big Data es un análisis aplicado a toda aquella información que no puede ser procesada o analizada usando las herramientas tradicionales (Zikopoulos, Eaton y Deroos, 2012), las entidades de cualquier tipo, cuentan con acceso a un gran volumen de información, pero no saben cómo obtener valor de ella, por qué la información obtenida no está estructurada y por lo tanto no saben si esa información les podría ser útil. (Garvich, 2017)

En el libro Harness the Power of Big Data (Zikopoulos, Eaton y Deroos, 2012) dice que los vendedores o comerciantes no ven más allá de la capacidad de almacenar grandes volúmenes de datos, si se desea liderar, debes saber analítica. (Paul, 2013)

Según (Zikopoulos, Eaton y Deroos, 2012) el Big Data se define en tres características (Volumen, Velocidad, Variedad), estas conforman el llamado Big Data que ha dado paso a un nuevo enfoque de análisis y toma de decisiones. (Paul, 2013)

La primera característica del Big Data que es el volumen implica realizar un análisis de grandes cantidades de datos que navegan diariamente en el mundo digital y que se dan gracias a la interacción de los usuarios en la web o también al ya conocido internet de las cosas.

Katya Garvich San Martin nos dice que al hablar a través de un Smartphone se genera un evento o cuando se hace una descarga se genera nueva información a cada segundo, ante esto las organizaciones no saben cómo administrar grandes volúmenes de datos, pero mediante las herramientas correctas se puede llegar a tener un mejor entendimiento del negocio y de los clientes. (Garvich, 2017)

Finalmente, el tema sobre los volúmenes de datos está evolucionando de “Terabytes” hacia “Petabytes” y sin duda seguirá hacia “Zettabytes” y toda esta información ya no se podrá almacenar ni procesar en los sistemas que usualmente se usan para el análisis de datos, pues no cuentan con las capacidades necesarias. (Paul, 2013)

La velocidad es uno de los factores del Big Data por qué el análisis se basa con datos en tiempo real, las organizaciones buscan ganar ventaja ante su competencia mediante el análisis que le dan a su información para hacer una buena toma de decisiones.

Lidiar efectivamente con Big Data requiere de la realización de un análisis contra el volumen y variedad de data mientras ésta está en movimiento no después de que está en reposo. (Paul, 2013)

La variedad del Big Data se refiere en cómo se obtienen los datos que serán analizados, hay datos que son generados de manera semi estructurada y no estructurada que provienen de páginas web, redes sociales, e-mail, sensores, etc. Y gran parte de estos datos recogidos no pueden ser procesados ni administrados en una base de datos tradicional. Las empresas comenzaron a darse cuenta del valor y la oportunidad del Big Data y para poder capitalizar esta oportunidad, las empresas deben estar en la capacidad de hacer un análisis de todo tipo de datos, ya sea relacional o no relacional. (Paul, 2013)

En las tesis de Katya Garvich dice que en el libro de Mitchell, Locke y Wilson (2012), existe una cuarta “V” que es el: “Valor”. Esto es lo que le interesa a una empresa, en obtener más valor en sus negocios, una forma de dar mayor valor es no solo operar en datos históricos, sino que también se puede analizar información en tiempo real. (Mitchell, 2012)

Big Data les da a las empresas la oportunidad de usar la información ya existente con el objetivo de extraer valor adicional para la toma de decisiones.

Tipos de Datos, según Mitchell, Locke y Wilson (2012), existen 3 tipos básicos de datos: La data estructurada, que son registros que están distribuidos en campos y filas de las bases de datos tradicionales, esto permite hacer una búsqueda y ordenamiento fácilmente a los criterios que creamos convenientes. La Data no Estructurada que no tiene un formato definido y no puede ser almacenada, como los videos, imágenes, audios, etc. La Data Semiestructurada que es una combinación de los datos estructurados y no estructurados, estos datos semiestructurados no se encuentran en una base de datos relacionales, pero pueden ser documentos XML y podrían facilitar su manipulación.

Tipos de Fuentes de Big Data, en el libro “Big Data, análisis de grandes volúmenes de datos en organizaciones” hace una clasificación en diferentes categorías, cuales contienen a su vez un buen número de fuentes diversas que recolectan, almacenan, procesan y analizan que a continuación se mencionan:

Web y Medios Sociales: Datos de flujos de clicks, Feeds de Twitter, Entradas de Facebook, Contenido Web.

Máquina a Máquina: Lecturas medidores inteligentes, Lecturas RFID, Señales GPS.

Datos de Transacciones Grandes: Demandas de salud, Demandas de telecomunicaciones, Registro de detalles, Registros de Facturación.

Biometría: Reconocimiento Facial, Genética.

Generado por los Humanos: Registros de voz, Correo Electrónico, Registros Médicos.

La información de Web y Medios Sociales es obtenida de los medios sociales como Facebook, Twitter, blogs, Wikipedia; los datos se capturan, almacenan o distribuyen teniendo en cuenta que son procedentes de los flujos de clics, tweets, retweets o también de sitios de almacenamiento de información como Dropbox. (Aguilar 2016)

El Análisis en Big Data hace referencia a la acumulación de un gran volumen de datos y a los procedimientos que se utiliza para encontrar patrones dentro de esos datos, dada su complejidad para almacenar y analizar se aplica un conjunto de técnicas (Mariño, 2015), entre las más utilizadas son:

Asociación: Trata de buscar relaciones entre las diferentes variables, para pretender encontrar una predicción en el comportamiento de otras variables. (Mariño, 2015)

Minería de Datos: El objetivo de este método es encontrar el comportamiento predictivo. Engloba un conjunto de métodos estadísticos y aprendizaje automático con almacenamiento en bases de datos.

Agrupación: Este análisis divide grandes grupos de datos en unos más pequeños y tiene como propósito encontrar similitudes entre estos grupos, y la identificación de nuevo conocimiento.

El Análisis de Sentimientos es el análisis de información no estructurada que puede ser recogida de las redes sociales. Usa técnicas de lingüística y de aprendizaje para obtener conocimientos que no se ven de manera explícita en ningún texto, pero que se pueden generar a partir de una relación o un aprendizaje automático. (Javier, 2017)

El análisis de sentimientos cubre todas las áreas centrales, donde incluye temas emergentes, se puede analizar las intenciones o detectar opiniones falsas y esto es adecuado para investigadores o para aplicaciones en ciencias de la gestión y ciencias sociales. (Liu, 2015) Este análisis tiene sus etapas que son:

Preparar texto para el análisis: Se hace uso de dos métodos utilizados uno para extraer comentarios de Facebook y el otro para extraer los tweets de Twitter

Extraer conceptos o palabras importantes: Se analiza que palabras son las más repetidas en los comentarios o tweets para tenerlas luego en cuenta al momento de hacer una predicción para los nuevos datos.

Aplicar el análisis de enlace de texto: Se crea una variable que almacene el tweet o comentario con el sentimiento, ya sea positivo o negativo.

Construir patrones de aprendizaje: Se aplica la creación de patrones o matrices de aprendizaje para determinar de manera automática el sentimiento de cada tweet o comentario.

Desplegar los modelos predictivos: Hacer uso de la matriz con los nuevos datos que ingresen para que la precisión tenga una mayor aceptación.

Los beneficios del análisis de sentimientos en el ámbito del Marketing, este análisis se empieza a usar cada vez más en la comercialización para tener un contacto más cercano con el cliente acerca de sus intereses.

El Aprendizaje Computacional tiene como objetivo hacer la búsqueda del análisis de la información de manera automática de una manera supervisada, con una base en pruebas de entrenamiento, una de las técnicas que se usan son: SVM (Support Vector Machine), Naive Bayes y de esta manera se hace uso de una categoría gramatical de las palabras y la frecuencia de algunos términos. (Paltoglou, 2012)

El Support Vector Machine (SVM) se usa para clasificar sentimientos, este método se basa utilizando vectores para dar uso a un aprendizaje automático y lograr tomar decisiones entre dos categorías. (Paltoglou, 2012)

Los Datos de las Redes Sociales, según Aguilar (2016) el acceso a redes sociales se realiza por numerosos dispositivos, teléfonos inteligentes, tabletas, televisores, etc. Cuando se almacenan los datos que produce cada red usuario en una red social, ya sea el número de mensaje que envía, el tiempo de permanencia en el sitio o cuantos amigos, amigos de amigos, y amigos de amigos de amigos, esto requiere una gran capacidad de proceso.

Si comparamos todos los datos producidos en la red social con los datos que tienen las empresas, ya sea de sus clientes, proveedores, social, costumbres, en sus transacciones diarias, tales como: facturas, llamadas de teléfonos, pedidos, servicios de atención al cliente, correos electrónicos o mensajes. Todos estos datos si son debidamente analizados, conduce a un nivel de conocimiento estratégico de la empresa, que permitirá optimizar en la resolución de problemas mediante una correcta toma de decisiones.

Acerca de la Obtención de Datos de Twitter, esta red social tiene unas reglas que explican cuáles son los comportamientos que se admiten en Twitter a la hora de hacer una búsqueda sobre algún tema en específico. (Twitter) Se debe tomar en cuenta algunos consejos para que los resultados que se obtengan sean los más relevantes: Evitar publicar contenido duplicado, ya sean tweets o enlaces y evitar el uso de bots para publicar mensajes automáticos.

El Análisis de Datos, según Katya nos dice que Judd, McClelland y Ryan (2011) dijo que el análisis de datos es un proceso que se basa en inspeccionar, limpiar y hacer una transformación de datos con la finalidad de obtener información útil y apoye la toma de decisiones. (Garvich, 2017)

En los bancos se analizan los retiros y los patrones de gasto para prevenir el fraude o robo de identidad, para el caso del comercio electrónico se examina el tráfico existente en el sitio web o los patrones de navegación para saber que clientes son los más cercanos a realizar la compra, tomando en cuenta las compras previas o patrones de visualización. Para el análisis de datos moderno se usa tableros de información basados en flujos de datos en tiempo real.

Los métodos y capacidades para el análisis de datos evolucionan, es decir los datos ya no son simples números que se encuentran en una base de datos tradicional, los archivos tales como texto, audio, video también pueden aportar conocimiento.

Metodología para el Análisis del comportamiento del consumidor, la minería de datos ha ido aumentando de forma gradual y con respecto a un consumidor por ejemplo en un supermercado, cada compra hecha se considera como un byte, se analiza cuanto gasta un consumidor, que días hace sus compras, que compra más a menudo y todo eso se almacena en una base de datos para ello el consumidor ni siquiera es consciente de todo lo que hay detrás de una sola compra ().

El Comercio Electrónico, según Ydones (2015) el comercio electrónico es toda transacción comercial de compra-venta o prestación de servicios, se hace un intercambio de valores, entre empresas, personas. Esto se lleva a cabo a través de medios digitales como un computador. (Ydones et al. [2015])

Con el avance del comercio electrónico el cliente o consumidor no quiere ir a una tienda física para comprar lo que desea, sino que, con esta nueva modalidad de

comercio, prefiere hacer sus compras mediante un sitio web y con alguna forma de pago segura. Esto genera mayor rapidez y con respecto a los negocios, mejora la rentabilidad y el manejo del stock de los productos.

Un estudio realizado por GFK en el 2015 sobre las compras por internet, su principal desventaja con un 21% es que el consumidor teme a que no llegue el producto, otra desventaja es que el producto no sea como se muestra en línea con un 19% y además del temor a un mal uso de los datos personales con 19%.

Las Características del Comercio Electrónico, según (Laudon y Traver 2016) en su libro “E-commerce: business, technology, society” nos muestra ocho características del comercio electrónico que nos explicaran por qué el interés hacia la implementación del E-commerce. (Laudon y Traver 2016)

Ubicuidad: El comercio electrónico está disponible en cualquier parte, en todo momento. Esto reduce los costos de transacciones desde el punto de vista del consumidor por qué ya no gastará en ir a una tienda física. (Laudon y Traver 2016)

Alcance Global: Las transacciones comerciales traspasan los límites culturales y nacionales, con mucha más efectividad en costos de lo que puede alcanzar el comercio tradicional, este solo puede ser en su mayoría local o regional como por ejemplo los periódicos que solo son locales o regionales, pero el comercio electrónico puede sobrepasar los límites nacionales y ser conocido ante una audiencia global. (Laudon y Traver 2016)

Estándares Universales: El internet y comercio electrónico usan los mismos estándares, el comerciante solo gastará en el envío de sus productos y el consumidor podrá ver y escoger el producto que desea y su gasto solo será el pago establecido de dicho producto. (Laudon y Traver 2016)

Riqueza: Se refiere al mensaje que da un comerciante, en internet tiene un potencial mayor que la televisión o radio, por qué el cliente puede interactuar con la web o puede hablar directamente con el comerciante mediante un foro. (Laudon y Traver 2016)

Interactividad: El comercio electrónico puede interactuar con el cliente, por ejemplo, la televisión no puede hacerles preguntas a los consumidores y para esto el

comercio electrónico permite que el comerciante tenga un contacto directo con el cliente y sepa cuáles son sus necesidades. (Laudon y Traver 2016)

Densidad de la Información: La información crece cada vez más y se encuentra disponible para clientes y comerciantes por igual. Estos pueden saber cuáles son las preferencias del consumidor y sacar una ventaja ante la competencia. (Laudon y Traver 2016)

Personalización/adecuación: Los comerciantes puede enviar mensajes de marketing específicos a los clientes a causa de los interés o compras anteriores del cliente y la adecuación es dar un cambio de producto o servicio para que se adecue a las preferencias del consumidor debido a su comportamiento en la web.(Laudon y Traver 2016)

Tecnología Social: La evolución de la tecnología han permitido a los usuarios puedan crear y distribuir contenido por internet.

Tipos de Comercio Electrónico:

Negocio a Consumidor (B2C): Este tipo se ve más a menudo, los negocios o comerciantes buscan la manera de llegar a los consumidores.

Negocio a Negocio (B2B): Son negocios que buscan tener contacto con otros negocios, ya sean distribuidores electrónicos, mercados de intercambio o consorcios industriales.

Consumidor a Consumidor (C2C): Este tipo de comercio se ponen en contacto dos consumidores mediante un mercado en línea, por ejemplo, un usuario publica un producto en modo de subasta o venta y es adquirido por otro consumidor.

Igual a Igual (P2P): Esta tecnología permite compartir archivos de computadora de manera directa sin necesidad de un intermediario, por ejemplo, BitTorrent que es usada para descargar archivos grandes.

Comercio móvil (m-commerce): Consiste en el uso de redes inalámbricas para conectar teléfonos celulares, los consumidores pueden realizar transacciones, comparar precios en tiendas, etc.

Las Exportaciones son un bien o servicio que es enviado a otra parte del mundo con fines lucrativos. Esto se puede concretar por diversos medios de transporte, aéreo, marítimo o terrestre.

El Arraigo Cultural en Piura donde la mayoría de los productos provienen de varias generaciones, por ejemplo, en Chulucanas los accesorios de paja, cerámica y artículos para el hogar, joyería y orfebrería de filigrana de plata. Estos trabajos tienen una alta demanda en el exterior y se siguen elaborando con técnicas ancestrales y se combinan con los modernos diseños para obtener una buena aceptación por parte del mercado internacional.

La cerámica en Piura da la vuelta al mundo, principalmente Chulucanas y Catacaos reflejan obras costumbristas: las tradiciones, costumbres y modos de vida, en los últimos años estos artistas han optado por hacer diseños más modernos pero que no se desligue de sus raíces. Lo que se destaca son los tejidos de paja toquilla, famosa por su calidad, finura y delicadeza; por su artesanía en madera que asombra por su belleza, perfección y diseño.

III. METODOLOGÍA

3.1 Tipo y Diseño de la Investigación

Tipo

Acerca del tipo de investigación es transversal por qué se realizó en un tiempo ya definido que es el año 2019.

El diseño de esta investigación es no experimental por qué solo identifica y describe los métodos de extracción y procesamiento de los datos para saber el interés de los usuarios acerca de la artesanía.

Nivel

El proyecto llega hasta un nivel descriptivo, primero se analizó los métodos que se usan para la extracción y el análisis profundo de datos, ya sea predictivo o descriptivo de los diferentes tipos de datos que existen.

3.2 Variables y Operacionalización

Tabla 1: Operacionalización de Variables

Variables	Dimensiones	Definición Conceptual	Definición Operacional	Indicadores	Escala de Medición
Método de Análisis (Variable Dependiente)	Minería de Datos Descriptivo	Según Katya Garvich, la minería de datos se define como el proceso de examinar la situación de alguna organización con el propósito de mejorar con la ayuda de métodos y procedimientos adecuados.	Extraer y almacenar la información de la base de datos y seleccionar herramientas y técnicas adecuadas para la rápida agrupación de los datos	-Comprensión de los datos	Escala
	Minería de Datos Predictivo			-Nivel de procesamiento de datos	Escala
Big Data (Variable Independiente)	- Datos Estructurados	Según (Garvich, 2017) Big Data es la información que no se puede procesar o analizar con el uso de las herramientas o procesos tradicionales	Selección de datos útiles, distribución y almacenamiento según la variedad del origen de los datos.	- Nivel de predicción	Escala
	- Datos Semiestructurados			- Nivel de reducción y transformación de datos	Escala
	- Datos no estructurados			- Ordenamiento de los datos	Escala
				-Conocimiento aportado	Escala
				- Ordenamiento de datos	Escala
				-Conocimiento aportado	Escala
				- Ordenamiento de datos	Escala

3.3 Población y Muestra

- Población

Estuvo constituida por todos los usuarios que interactuaron en las redes sociales de Facebook y Twitter entre el 1 y 10 de octubre.

- Unidad de Análisis

Todos los usuarios que hablaron de artesanía del 1 al 10 de octubre.

- Muestreo

No probabilístico – por conveniencia

3.4 Técnicas e Instrumentos de recolección de datos

Técnica: Análisis Documental

Instrumento: Ficha de Registro de Datos

Ficha de Análisis 01

Objetivo del Instrumento: Este instrumento me permitirá obtener conocimiento sobre el método de extracción de datos de las redes sociales, como lo son: Facebook y Twitter. Además, tiene la finalidad de medir indicadores como: Comprensión de Datos, Nivel de procesamiento de datos, Nivel de predicción.

Obtención de Información

Extracción de datos en Facebook: Se selecciona en primer lugar la publicación que se requiere o se necesita analizar, luego se extrae su URL para hacer un análisis de sentimientos sobre los comentarios de la publicación.

Se exporta los comentarios, la fecha, ID y nombre de usuario en un archivo CSV que puede ser descargado para luego procesar sus datos.

Extracción de datos en Twitter: Es una aplicación web que lanza un disparador en Twitter, donde previamente se configura que palabra o Hashtag se desea analizar para que el disparador haga una búsqueda exhaustiva en cada publicación que es utilizada.

Cada vez que la palabra o Hashtag es utilizado por los usuarios, el tweet es grabado automáticamente en Google Drive en una hoja de cálculo, con el usuario y la fecha respectiva.

MÉTODOS	Datos Comprensibles	Búsqueda automática de los datos	Exportación y procesamiento automático
Extracción de datos de Facebook	✓	✓	✗
Extracción de datos de Twitter	✓	✓	✓

Ficha de Análisis 02

Objetivo del Instrumento: Este instrumento me permitirá obtener conocimiento sobre los métodos de procesamiento de datos obtenidos. Además, tiene la finalidad de medir indicadores como: Ordenamiento de datos, Conocimiento aportado, procesamiento de los datos.

Procesamiento de la Información

Método Descriptivo: Al hacer un análisis descriptivo de los datos obtenidos, se necesita hacer un análisis no supervisado o se realiza un análisis clúster, que es una técnica de clasificación automática de datos, donde se clasifica los datos en positivos o negativos para un próximo análisis predictivo.

Método Predictivo: Para hacer un análisis predictivo previamente se aplica un aprendizaje automático de datos para que en un futuro los nuevos datos que ingresan estén clasificados automáticamente en una o más variables.

Procesamiento de la información	Fácil procesamiento de los datos	Nivel de conocimiento Aportado	Alto nivel de predicción de datos
Método Descriptivo	✓	✓	✗
Método Predictivo	✓	✓	✓

3.5 Método de Análisis de Datos

Análisis de Sentimientos aplicado a los datos recolectados de Facebook y Twitter.

Este análisis de sentimientos es un estudio que determina una opinión de los usuarios en internet sobre algún tema en específico, se aplica un algoritmo de clasificación de palabras para determinar si la opinión es negativa o positiva y esta predicción tendría un 74% de precisión en los resultados. (Montesinos García 2014)

Este análisis nos permitirá tener una idea clara de la opinión pública general sobre ciertos temas y en este caso la artesanía.

Tweets y comentarios recolectados con una API

➤ Recolección de datos

- library(httr)
- library (ROAuth)
- setup_twitter_oauth ("LRhacLCUyzv56iZwrx8tzygER", "j5lQtWaksL8NVLJ0hBFFodvSYpRBdi3Qdugv9Em9V7H0IFUieE")
- apple <- searchTwitter ("apple", n=250) //Numero de tweets
- getwd ()
- tweets <- read.csv ("tweets3.csv", sep=",") // Llamar la base de datos de los tweets
- table(tweets\$Sentiment) //Mostrar los tweets positivos y negativos

Limpiar los datos y transformar al formato que necesitamos.

➤ Pre-Procesamiento de datos y ordenamiento de datos

- corpus = Corpus (VectorSource(tweets\$Tweet))
- length(corpus)
- content (corpus [[1]])
- corpus <- tm_map (corpus, tolower) //Poner los tweets en minúsculas
- corpus <- tm_map (corpus, removePunctuation) // Quitar la puntuación de los tweets

- `corpus <- tm_map (corpus, removeWords, (stopwords("english"), "apple"))`
// remover los enlaces y palabras que no necesitamos
- `corpus <- tm_map (corpus, stemDocument)` // cortar palabras a raíz
- `frequencies <- DocumentTermMatrix(corpus)`
- `frequencies`
- `inspect (frequencies [100:105, 50:65])` // Selección de tweets, comentarios y los términos determinados en un rango
- `findFreqTerms (frequencies, lowfreq = 50)` // términos más frecuentes
- `findFreqTerms (frequencies, lowfreq = 30)` // términos más frecuentes
- `sparse <- removeSparseTerms (frequencies, 0.995)` // Palabras que no nos sirven para el análisis (pocas mencionadas)
- `sparse` // mostrar nuevos valores de la matriz
- `tweetsSparse <- as.data.frame (as.matrix(sparse))` //Retornar la matriz

Entrenar un clasificador (modelo) para clasificar los tweets como positivo o negativo.

➤ Análisis de sentimientos

- `colnames(tweetsSparse) = make.names (colnames(tweetsSparse))` // asignando nombres de cada palabra a las columnas del dataframe
- `tweetsSparse$sentiment <- tweets$Sentiment` // pegar la variable sentimiento al tweets sparse

Modelo de comunicación machine learning (cada tweet se ubica en un espacio dimensional según las palabras que tenga) Busca realizar una pared "división" que divida lo positivo con lo negativo.

Enseñarle al vector machine 80 % de entrenamiento y 20 % de evaluación, para ello necesitamos descargar un paquete CaTools

- `set.seed (12)`//Punto de aleatoriedad y para que nos den los mismos resultados

Creamos la variable split que nos va a decir que observaciones van para el test (evaluación) y para el train (entrenamiento)

- `split <- sample.split (tweetsSparse$sentiment, SplitRatio = 0.8) // 80 % de las observaciones de la base datos se vaya para el conjunto de entrenamiento.`
- `trainSparse = subset (tweetsSparse, split == TRUE) // Observaciones de entrenamiento`
- `testSparse = subset (tweetsSparse, split == FALSE) // Observaciones de evaluación`

Utilizar estadísticas de las etiquetas de los tweets para medir el sentimiento positivo o negativo.

➤ Predicción

Después de tener las dos bases de datos corremos el modelo de support vector machine

Modelo para comparar sentimiento

- `table(testSparse$sentiment) // OBTENER LOS TWEETS POSITIVOS Y NEGATIVOS DE LA VARIABLE TEST`

Llamamos a la variable que deseamos predecir para utilizar todas las palabras que tenemos.

- `SVM <- svm (as.factor (sentiment)~., data=trainSparse)`
- `Summary (SVM) // parámetros del modelo para poder clasificar tweets nuevos`
- `predictSVM <- predict (SVM, newdata = testSparse) // cómo se comporta el modelo haciendo predicciones respecto a los tweets.`
- `confusionMatrix (predictSVM, testSparse$sentiment) // evaluar el desempeño del modelo con el algoritmo respecto de los tweets precisión del 96 %`

3.6 Aspectos Éticos

Para llevar a cabo el presente proyecto de investigación se tendrá en cuenta los siguientes aspectos éticos:

- El respeto por las teorías relacionadas al tema, recurriendo a las referencias propuestas por normas en la documentación.
- Rendimiento a las autoridades o instituciones que han sido sujeto de análisis en lo referente si existiera la divulgación de aspectos confidenciales que afecten su reputación como institución.
- Resguardo a la identificación de los participantes que formen parte del estudio, que pudieran dar su apoyo o ser parte del estudio de forma anónima, o si fuera el caso mencionarlos previa autorización comunicada.

IV. Resultados

Resultados aplicando métodos de extracción y análisis de datos para Facebook y Twitter

Se hace un cruce tablas entre los métodos de análisis de datos obtenidos y el procesamiento aplicado utilizado estos dos métodos como instrumento de evaluación para obtener la cantidad de usuarios que se han interesado por la artesanía en el período establecido.

Tabla 2: Métodos de extracción y análisis de datos para Facebook y Twitter

		Obtención de Información	
		Extracción de Datos de Facebook	Extracción de Datos de Twitter
Procesamiento de la Información	Método Descriptivo	99 usuarios	650 usuarios
	Método Predictivo	97 usuarios	130 usuarios

Figura 1: Resultados de extracción de datos de Twitter

Screen Name	Full Name	Tweet Text
@davidmingorance	David Mingorance	(3/3) Desarrollo del concepto #proyectoOximoron, se trabajó e
@mycomx	Sistema Myco	La muerte de hace muy viva a través de la artesanía. (Cortes
@Fundesarte	Fundesarte	Ya tenemos finalistas de los Premios Nacionales de Artesanía
@InesPoveda	Ines Poveda	Creo que lo mejor que te ofrece cualquier tipo de artesanía he
@Lellana4	Lellana	La ciudad de Holguín ultima los preparativos como sede de la
@FARCAMA	FARCAMA	En el Espacio de Venta agrupado de Farcama, mostramos uni
@Radiocamarapy	Radio Cámara Py	Dip. Juan Carlos Glaverna "Agradecer a los colegas por parti
@Gloria_A41	Gloria	Socarrat valenciano #Valencia #arte #cultura #artesanía
@ProvincialPuno	Municipalidad Provincial de Puno	La Municipalidad de Puno a través de la Subgerencia de Prom
@wamnews_Spanish	WAM News / Spanish	El Príncipe Andrew elogia el Fondo Khalifa por apoyar el empr
@24clm	clm24.es	Luz verde a las obras para la puesta en marcha del Centro de
@promocionFCBC	FCBC	¡Feliz miércoles! Estamos en modo #Iberoarte2019 Próximama
@SdeAudiovisuais	Servizo de Audiovisuais	📺📺📺 Queres mercar artesanía galega ? A mosámoche algu
@DipuCuenca	Diputación de Cuenca	📺 Hoy se celebra el día de Cuenca en FARCAMA, la Feria de A
@Bienvenido_QL	Bienvenido_QL	FIESTA GASTRONÓMICA LATINOAMERICANA, ARTESANÍA Y
@paulrobsoncs	School Admin	¡ÚNASE A NOSOTROS PARA UNA NOCHE DE COMIDA, ARTE
@Corbatur_Vzla	#CORBATUR	#EIArteDeLasManos En Barinas, podrás encontrar la extens
@murciaeconomia	MurciaEconomía.com	La murciana Tictactoy, finalista en los Premios Nacionales de
@BenitoMosaico	mosaico hidraulico benito	Aquí os dejo nuestro canal de YouTube, donde podréis ver la
@cristilucenac	Cristina Lucena C	Para mi las #chivas o #escaleras eran un símbolo de #Colomb
@repeusate	Repelúsate	Y con este precioso par de pendientes doy por finalizada nue
@JavieGonal	Javier González	Estoy de acuerdo. Apreciemos su esfuerzo para ganase la vi
@rtp_bolivia	RTP Bolivia	Taller de artesanía en totora en la comunidad Sicuani, a 15 mi

Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Twitter Archiver - extracción de tweets

Figura 2: Resultados de extracción de comentarios de Facebook

1	name	Profile ID	Date	Comment
2	Natura Zafiro Dalila Ferreyra	ID: 185775691955	27-4-19 21:44	Precio
3	Lichi Ac	ID: 100007120784	27-4-19 21:45	Liza Espinoza
4	Verito Ramirez Tuesta	ID: 100002758838	27-4-19 21:46	Precio
5	Kris Giles	ID: 100000377841	27-4-19 21:47	Precio
6	Lesly Corcuera Asmat	ID: 1651200491	27-4-19 21:48	Precio
7	Shuriita Chavez	ID: 100013032754	27-4-19 21:49	Precio???
8	Katty Navarrete	ID: 100000778460	27-4-19 21:51	Precio
9	Doris Eca Panta	ID: 100004169112	27-4-19 21:53	Precio
10	Ruby Balcazar	ID: 100007774516	27-4-19 21:54	Precio
11	Cynthia Vega Ramos	ID: 1525345626	27-4-19 21:54	Precio
12	Elizabeth Nancy López Torres	ID: 100001573391	27-4-19 21:55	Precio?
13	Mayra Farfán Marquez	ID: 100001008522	27-4-19 21:57	hacen letras 3D
14	Kristel Eyzaguirre	ID: 100001941248	27-4-19 21:57	Precio por mayor?
15	Jahita Minchán	ID: 1181652617	27-4-19 21:57	Precio
16	Luisa Castro Farfan	ID: 100003997597	27-4-19 21:59	Precio
17	Amae Esteban	ID: 100003308426	27-4-19 22:01	Precio.x mayor.x favor
18	Liliana Noemi	ID: 100003194884	27-4-19 22:08	Precio
19	Karina Uribe Contreras	ID: 1279208720	27-4-19 22:08	precio x mayor
20	Kathy Vargas Torres	ID: 1598179522	27-4-19 22:10	Precio
21	Karen Jannet Landa Salgado	ID: 100001496728	27-4-19 22:15	Precio por fv
22	Madeleine Avalos Saavedra	ID: 100002319335	27-4-19 22:17	Precio???
23	Brigitte Sandoval Aranda	ID: 1379059814	27-4-19 22:21	Precio

Autor: Carrasco Cruz Luis Germán – 2019

Fuente: EXPORTSCOMENTS.COM – extraer comentarios de Facebook

Resultados de los Objetivos

- Identificar los métodos de extracción de datos: Público objetivo que consume productos de Artesanía a través de Big Data.

Tabla 3: Resultados de primer objetivo

Método de extracción de datos de Facebook		
Definición	Ventajas	Desventajas
Este método consiste únicamente en exportar los comentarios de Facebook de alguna publicación en específica que sea de nuestro interés para su posterior análisis.	<ul style="list-style-type: none"> -Acceso a ID y nombre de usuario -Acceso a fecha y hora de publicación de comentario -Exportación a una hoja de cálculo 	<ul style="list-style-type: none"> -Hay que buscar cada publicación que se necesita para extraer los comentarios que se van a analizar y no hay una búsqueda automática.

Tabla 4: Resultados de primer objetivo

Método de extracción de datos de Twitter		
Definición	Ventajas	Desventajas
Este método es básicamente un disparador que es lanzado hacia Twitter con algunas palabras de referencia, cuando encuentre dichas palabras o frases en un tweet, este será almacenado de manera automática en una hoja de cálculo en la nube.	<ul style="list-style-type: none"> -Acceso a nombre de usuario -El disparador hace una búsqueda automática sobre la palabra o frase que se haya configurado previamente para luego analizar cada tweet. -En algunos tweets se guarda hasta la ubicación del usuario 	<ul style="list-style-type: none"> -No hay una clasificación de cada usuario por sexo o por edad.

- Describir el nivel de interés de artesanía con el modelo usado en cuanto al análisis de datos

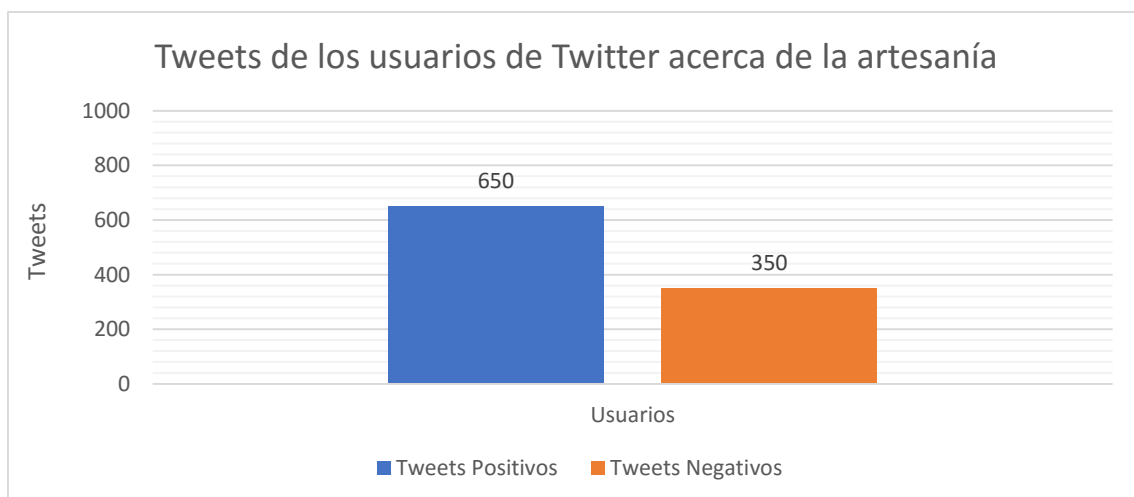
Nivel de interés de los usuarios de Twitter

En este método de extracción de datos de Twitter se tuvo una suma de 1000 tweets válidos, con un análisis de clasificación de 61% los cuáles eran usuarios que hablaban positivamente de la artesanía desde el 1 al 10 de octubre que fueron los días que se tomaron en cuenta.

Mediante las siguientes líneas de código en RStudio hizo la clasificación de los tweets positivos y negativos.

- `tweets <- read.csv ("tweets3.csv", sep=",") //llamar a la base datos de nombre "tweets3.csv"`
- `table(tweets$Sentiment) //Mostrar los tweets positivos y negativos en una tabla`
- `trainSparse = subset (tweetsSparse, split == TRUE) // Observaciones de entrenamiento`
- `positive <- subset (tweetsSparse, tweetsSparse$sentiment==1) // Observación de los tweets positivos`
- `positivas <- as.data.frame (colSums(positive)) // Generar un DataFrame de los tweets positivos`

Figura 3: Grafico de los tweets usuarios



Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Clasificación de la ejecución de análisis de sentimientos a los datos recogidos.

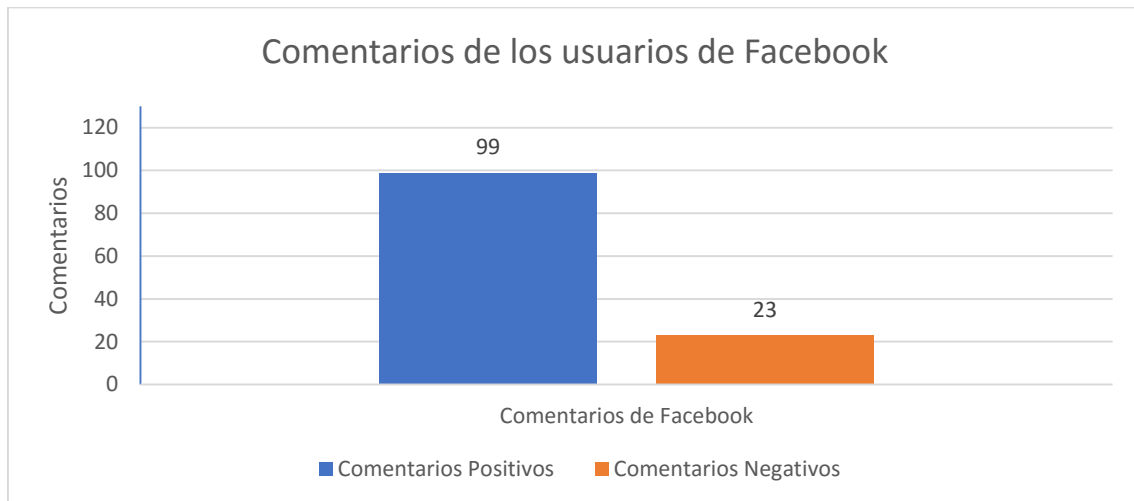
Nivel de interés de los usuarios de Facebook

Para este método se obtuvieron 122 comentarios de Facebook sobre publicaciones de artesanía de los cuales 99 fueron positivos.

Mediante las siguientes líneas de código en R se hizo la clasificación:

- `coments1 <- read.csv("Fb_bd.csv", sep = ",")`
- `table(coments1$sentiment)`

Figura 4: Grafico de comentarios de facebook



Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Clasificación de la ejecución de análisis de sentimientos a los datos recogidos.

- Describir que tan útil es el modelo de minería de datos para obtener los resultados requeridos

Métodos de Minería de Datos

Método Descriptivo

Es el agrupamiento no supervisado que se hace de manera global sobre los datos que se tienen. Ayuda a tener una idea general de los resultados obtenidos.

-Para este método se hizo una Clusterización de los datos, que nos ayudó a clasificar de una manera general los tweets o comentarios que fueron positivos y saber que usuarios estarían interesados en artesanía.

- `tweets <- read.csv("tweets3.csv", sep=",") // Llamar la base de datos de los tweets`

- `table(tweets$Sentiment) //Mostrar los tweets positivos y negativos`

**Resultados de Método
descriptivo aplicado en los
comentarios de Facebook**

Tabla 5: Resultados de
método descriptivo en
Facebook

Centros de clústeres finales

	Clúster	
	1	2
sentimientos	1	-1

Tabla 6: Resultados en cada clúster

**Número de casos en
cada clúster**

Clúster 1	99,000
2	23,000
Válidos	122,000
Perdidos	,000

Resultados de Método descriptivo aplicado en los Tweets de Twitter

Tabla 7: Resultados de método descriptivo en Twitter

Centros de clústeres finales

	Clúster	
	1	2
sentimiento	1	-1

Tabla 8: Resultados en cada clúster

Número de casos en cada clúster

Clúster	1	650,000
	2	350,000
Válidos		999,000
Perdidos		,000

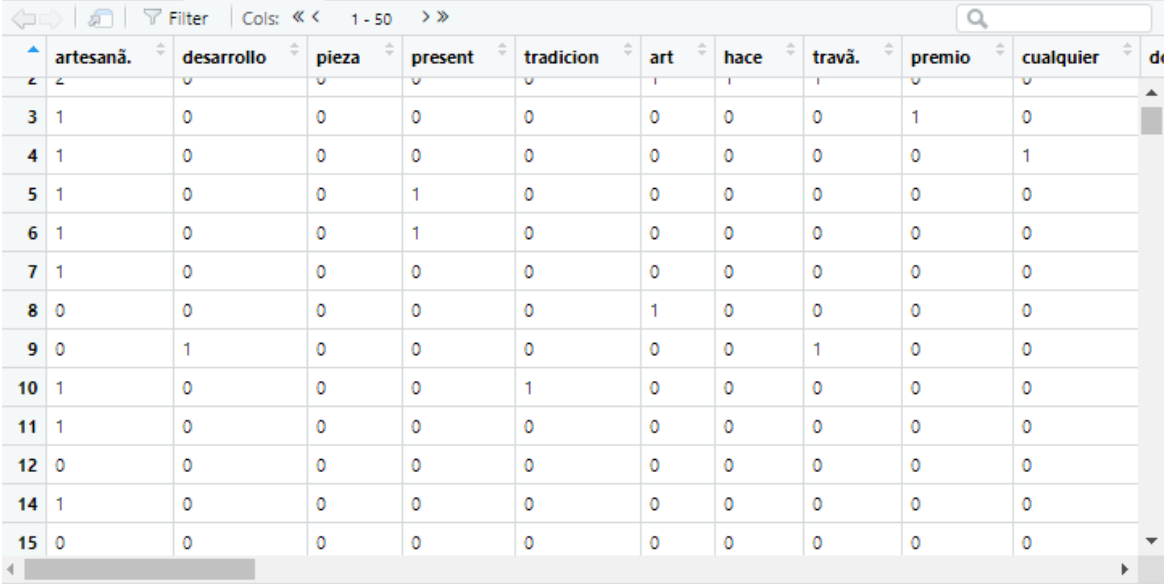
Método Predictivo

Este método hace un análisis sobre los resultados del método descriptivo para hacer un aprendizaje automático y predecir valores de otras variables o nuevos datos.

-En este método primero se hizo una clasificación de las palabras que más se repiten en una matriz de datos para luego hacer una predicción de cuantos tweets serían positivos.

- `predictSVM <- predict (SVM, newdata = testSparse) // cómo se comporta el modelo haciendo predicciones`
- `confusionMatrix (predictSVM, testSparse$sentiment) // evaluar el desempeño del modelo con el algoritmo respecto de los tweets precisión del 96 %`
- `positive <- subset(tweetsSparse, tweetsSparse$sentiment==1)`
- `positive$sentiment <- NULL// eliminar la variable sentiment porque ya tenemos solo los tweets positivos`

Figura 5: Clasificación de Tweets



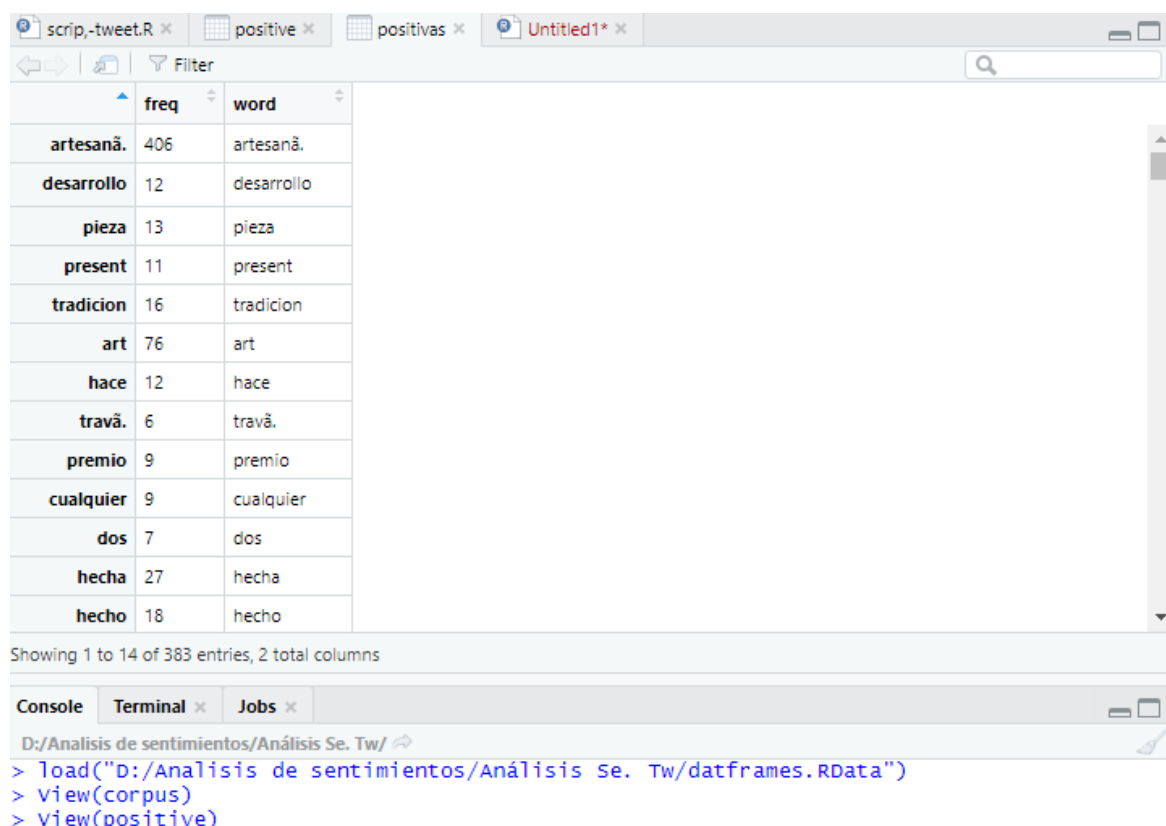
	artesanã.	desarrollo	pieza	present	tradicion	art	hace	travã.	premio	cualquier	dc
3	1	0	0	0	0	0	0	0	1	0	
4	1	0	0	0	0	0	0	0	0	1	
5	1	0	0	1	0	0	0	0	0	0	
6	1	0	0	1	0	0	0	0	0	0	
7	1	0	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	1	0	0	0	0	
9	0	1	0	0	0	0	0	1	0	0	
10	1	0	0	0	1	0	0	0	0	0	
11	1	0	0	0	0	0	0	0	0	0	
12	0	0	0	0	0	0	0	0	0	0	
14	1	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	

Showing 2 to 16 of 650 entries. 383 total columns

Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Clasificación de palabras positivas y negativas de tweets en R.

Figura 6: Palabras positivas de los tweets



	freq	word
artesanã.	406	artesanã.
desarrollo	12	desarrollo
pieza	13	pieza
present	11	present
tradicion	16	tradicion
art	76	art
hace	12	hace
travã.	6	travã.
premio	9	premio
cualquier	9	cualquier
dos	7	dos
hecha	27	hecha
hecho	18	hecho

Showing 1 to 14 of 383 entries, 2 total columns

```
D:/Análisis de sentimientos/Análisis Se. Tw/
> load("D:/Análisis de sentimientos/Análisis Se. Tw/datframes.RData")
> view(corpus)
> view(positive)
```

Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Matriz de palabras positivas que más se repiten en los tweets usando R.

Figura 7: Clasificación de comentarios de Facebook

	ahora	ayacucho	belleza	bonito	buena	bueno	caballo	chiclayo	costo	cual	cuanto
42	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	1	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0
50	0	1	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0	0	0
54	0	0	0	0	0	0	0	0	0	0	0

Showing 40 to 51 of 99 entries, 81 total columns

```

D:/Analisis de sentimientos/Fb/
> view(positivas)
> view(positive)
> |
    
```

Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Clasificación de palabras positivas y negativas de comentarios de Facebook en R.

Figura 8: Palabras positivas de comentarios de Facebook

	freq	word
envi	0	envi
envio	0	envio
espinoza	0	espinoza
está.	1	está.
fabor	1	fabor
favo	1	favo
favor	17	favor
ficha	0	ficha
gracias	1	gracias
hacen	0	hacen

Showing 21 to 32 of 81 entries, 2 total columns

```

D:/Analisis de sentimientos/Fb/
> View(positivas)
    
```

Autor: Carrasco Cruz Luis Germán – 2019

Fuente: Matriz de palabras positivas que más se repiten en los comentarios de Facebook usando R.

V. DISCUSIÓN

El propósito de la investigación es identificar métodos en las redes sociales que analicen un público objetivo que esté interesado en consumir o conocer productos de artesanía, haciendo uso del Big Data y la información recolectada sea analizada a través de un análisis de sentimientos.

Se hizo una recolección de datos de las redes sociales, Facebook y Twitter donde millones de usuarios tienen interacción diaria.

Así mismo Katya Garvich (2017) en Lima, su investigación tuvo como título "Propuesta de análisis de datos no estructurados para generar decisiones oportunas en la empresa GMD" y su objetivo era determinar el impacto en tiempos y costos a la hora de aplicar un análisis en los datos no estructurados que tenía la empresa, coincide con la presente investigación en cuanto al análisis o minería de datos que se va a realizar para una futura toma de decisiones con respecto a las predicciones obtenidas, pero se diferencian en que la investigación de Garvich ya cuenta con los datos que va a analizar sin embargo para reunir los datos de las redes sociales hay que esperar un cierto tiempo para obtener una cantidad considerable y luego poder analizarlos para tener una predicción acerca de lo que los usuarios opinan en este caso sobre la artesanía.

En la investigación de Montesinos García 2014 quien hizo un análisis de sentimientos para saber cuál sería el posible ganador de las elecciones de su país tomando en cuenta los tweets de los usuarios clasificándoles en positivos y negativos, coincide con la presente investigación pues se hizo un análisis similar sobre los comentarios de Facebook y los tweets de Twitter a usuarios que hablen sobre artesanía del 1 al 10 de octubre, para saber que usuarios estarían interesados en comprar o conocer más acerca de la artesanía.

Este análisis de sentimientos permite tener una predicción bastante certera que ayudaría a tener un conocimiento sobre lo que podría pasar o tener en cuenta cuál es la realidad actual de alguna situación, como lo explica (Montesinos García 2014)

Se identifica los métodos para público objetivo que consume artesanía se hace una clasificación de los datos recogidos para saber que usuarios son los interesados o hablan sobre la artesanía, Miguel Grández (2017) en su investigación donde aplica

minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales, obtuvo como conclusión que el producto que tiene más demanda es el que contiene proteínas y sería comprado por clientes que tengan entre 25 y 32 años.

Esta conclusión se maneja de forma específica acerca de cada cliente y en la presente investigación se hace un análisis de clasificación primero descriptiva sobre un aprendizaje no supervisado y posteriormente con esos resultados se hace una clasificación predictiva para los nuevos datos que se podrían generar con el objetivo de conocer a los usuarios interesados que deseen adquirir o conocer más sobre artesanía.

Por otra parte, la información sobre cómo funciona el análisis de sentimientos de Javier Gamboa (2017) fue muy útil para esta investigación por qué fueron tomados en cuenta los pasos para la elaboración de un buen análisis de sentimientos acerca del texto que ya se había obtenido, se tiene que hacer un pre procesamiento y limpieza de los datos para aplicar el análisis y luego hacer las predicciones, también habla sobre los beneficios que tiene consigo este análisis, se puede aplicar en diferentes realidades, en este caso se aplicaría en el Marketing de una empresa para conocer su realidad en cuanto a sus clientes y saber sus opiniones o sus preferencias para mejorar la relación o tener un mejor acercamiento.

VI. CONCLUSIONES

-Para identificar los métodos de extracción de datos del público objetivo que esté interesado en artesanía se concluyó que se puede realizar con una API que extraiga los datos ya a una base de datos para su posterior procesamiento y limpieza de datos, también se puede lanzar un disparador que reúna en determinado tiempo los mismos datos pero se tendría que luego seleccionar solo los campos interesados, estos métodos de extracción se aplicaron a Facebook y Twitter las cuales al momento de su aplicación hay algunas ventajas y desventajas en cuanto a la recolección y cuando se tienen todos los datos debidamente estructurados se procede a hacer el análisis de sentimientos para determinar qué cantidad de usuarios estaría interesado en conocer o adquirir artesanía.

-Acerca del nivel de interés sobre la artesanía utilizando el modelo de análisis de datos se concluyó que en el método de extracción aplicado en Facebook hay ciertas desventajas por qué se tiene que buscar de forma manual cada publicación que sea de nuestro interés para extraer los comentarios y analizarlos sin embargo para la aplicación de este método en Twitter se lanza un disparador que de forma automática reúnen los tweets que se necesitan y se almacenan en una hoja de cálculo. Para el caso de los datos obtenidos en Facebook se obtuvo 122 comentarios para analizar y en Twitter más de 1000 tweets, en cuanto al nivel de interés por parte de los usuarios los comentarios positivos superaron el 80% y los tweets el 61%.

-En cuanto a la utilidad del modelo de minería de datos se llegó a la conclusión que con el método descriptivo se hace una clasificación de forma general y no supervisada, son los resultados en primera instancia, tanto para los datos de Facebook y Twitter se obtuvo un resultado positivo mayor al 80% y 60% respectivamente y al utilizar el método predictivo, primero se crea una matriz sobre las palabras que más se repiten en los textos y se genera un aprendizaje automático para los nuevos datos que puedan ingresar sobre la artesanía.

VII. RECOMENDACIONES

- Para una futura investigación donde se aplique análisis de datos para realizar Big Data se recomienda tomar en cuenta los diferentes métodos de análisis, puede ser de Asociación para que se encuentre una relación entre las diferentes variables o la Agrupación para encontrar similitudes o algunas características que prefieran los usuarios sobre la artesanía, ya sea una clasificación por el tipo artesanía que es más concurrida.
- Se recomienda abordar otro tipo de variables que no fueron tomadas en cuenta en esta investigación, por ejemplo, el análisis de datos acerca de la artesanía sería mejor implementado con más profundidad a la comercialización, hoy en día es necesario conocer cuál es la realidad del mercado al que queremos llegar para tener un aumento en las ventas y sobre todo conocer el interés del cliente para tener una mejor relación.
- La aplicación de un análisis de datos en alguna realidad se puede incluir también al ya famoso Marketing Digital por ello se recomienda aplicar técnicas para la extracción de datos de la web para su posterior procesamiento y hacer llegar algún tipo de publicidad a los usuarios acerca de la artesanía piurana y comenzar a hacer realidad una vitrina digital que muestre todo lo que se produce en Piura.

REFERENCIAS

- AGUILAR, L.J., 2016. *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. S.l.: Alfaomega Grupo Editor. ISBN 978-607-707-757-2.
- GRANDEZ MÁRQUEZ, [2017]. Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales. , pp. 91.
- HEREDIA, N., 2018. Diseño de un plan de exportación de artesanías mexicanas de arte huichol de México a Europa. ,
- LAUDON, K.C. y TRAVER, C.G., 2016. *E-commerce: business, technology, society*. S.l.: s.n.
- LÓPEZ, M., MATAMOROS, J., OLEAS, O. y LAYEDRA, N., 2010. Proyecto de inversión para la implementación de un negocio de exportación de artesanías de tipo textil. ,
- MONTESINOS GARCÍA, L., 2014. Análisis de sentimientos y predicción de eventos en twitter. ,
- YDONES, D., KATTIA, D., VÁSQUEZ, V. y SABINA, B., [sin fecha]. ESTUDIO EXPLORATORIO DE LA OFERTA DE COMERCIO ELECTRÓNICO EN UN CONJUNTO DE MICRO Y PEQUEÑAS EMPRESAS (MYPES) LOCALIZADAS EN DIVERSOS DISTRITOS DE LIMA METROPOLITANA. , pp. 164.
- DATA MINING TECHNIQUES: A SOURCE FOR CONSUMER BEHAVIOR ANALYSIS. [En línea]
- DIANA LLANOS RODRÍGUEZ, FIORELLA DEL PILAR TIRADO TELLO. 2015. ANÁLISIS DE LAS EXPORTACIONES. IQUITOS – PERÚ. : s.n., 2015.**
- Fandiño, Víctor López. 2005. Aplicación de técnicas de minería de datos para la fidelización y retención de clientes. 2005.**

- Garvich, Katya Elizabeth. 2017.** *PROPUESTA DE ANÁLISIS DE DATOS NO ESTRUCTURADOS PARA GENERAR DECISIONES OPORTUNAS EN LA EMPRESA GMD.* Lima - Perú : s.n., 2017.
- Guevara, Iván Mendoza. 2012.** *Informe Artesanía Chulucanas.* 2012.
- Javier, Gamboa Cruzado. 2017.** *Text Mining - Análisis de Sentimientos Para la Toma de Decisiones.* Lima : s.n., 2017.
- Liu, Bing. 2015.** *Sentiment Analysis - Mining Opinions, Sentiments, and Emotions.* 2015.
- Mariño, José Camilo Barriga. 2015.** *Técnicas de Análisis de Datos en Big Data.* Colombia : s.n., 2015.
- Mitchell, I., Locke, M. & Wilson, M. 2012.** *The white book of Big Data.* United Kingdom: Fujitsu Services Ltd. : s.n., 2012.
- Paltoglou, Georgios. 2012.** *Twitter, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media.* 2012.
- Paul, Zikopoulos, Dirk Deroos, Krishnan Parasuraman. 2013.** *Harness the power of Big Data.* United States of America: The McGraw-Hill Companies. : s.n., 2013.
- PROMPERU. 2014.** *Catalogo exportador de la Región de Piura.* 2014.
- Twitter.** Reglas y Políticas de Twitter. [En línea] <https://dev.twitter.com/rest/>.

ANEXOS

Tabla: Operacionalización de Variables

Variables	Dimensiones	Definición Conceptual	Definición Operacional	Indicadores	Escala de Medición
Método de Análisis (Variable Dependiente)	Minería de Datos Descriptivo	Según Katya Garvich, la minería de datos se define como el proceso de examinar la situación de alguna organización con el propósito de mejorar con la ayuda de métodos y procedimientos adecuados.	Extraer y almacenar la información de la base datos y seleccionar herramientas y técnicas adecuadas para la rápida agrupación de los datos	-Comprensión de los datos	Escala
	Minería de Datos Predictivo			-Nivel de procesamiento de datos	Escala
Big Data (Variable Independiente)	- Datos Estructurados	Según (Garvich, 2017) Big Data es la información que no se puede procesar o analizar con el uso de las herramientas o procesos tradicionales	Selección de datos útiles, distribución y almacenamiento según la variedad del origen de los datos.	- Nivel de predicción	Escala
	- Datos Semiestructurados			- Nivel de reducción y transformación de datos	Escala
	- Datos no estructurados			- Ordenamiento de los datos	Escala
				-Conocimiento aportado	Escala
				- Ordenamiento de datos	Escala
				-Conocimiento aportado	Escala
				- Ordenamiento de datos	Escala

Técnicas e Instrumentos de recolección de datos

Técnica: Análisis Documental

Instrumento: Ficha de Registro de Datos

Ficha de Análisis 01

Objetivo del Instrumento: Este instrumento me permitirá obtener conocimiento sobre el método de extracción de datos de las redes sociales, como lo son: Facebook y Twitter. Además, tiene la finalidad de medir indicadores como: Comprensión de Datos, Nivel de procesamiento de datos, Nivel de predicción.

Obtención de Información

Extracción de datos en Facebook: Se selecciona en primer lugar la publicación que se requiere o se necesita analizar, luego se extrae su URL para hacer un análisis de sentimientos sobre los comentarios de la publicación.

Se exporta los comentarios, la fecha, ID y nombre de usuario en un archivo CSV que puede ser descargado para luego procesar sus datos.

Extracción de datos en Twitter: Es una aplicación web que lanza un disparador en Twitter, donde previamente se configura que palabra o Hashtag se desea analizar para que el disparador haga una búsqueda exhaustiva en cada publicación que es utilizada.

Cada vez que la palabra o Hashtag es utilizado por los usuarios, el tweet es grabado automáticamente en Google Drive en una hoja de cálculo, con el usuario y la fecha respectiva.

MÉTODOS	Datos Comprensibles	Búsqueda automática de los datos	Exportación y procesamiento automático
Extracción de datos de Facebook			
Extracción de datos de Twitter			

Ficha de Análisis 02

Objetivo del Instrumento: Este instrumento me permitirá obtener conocimiento sobre los métodos de procesamiento de datos obtenidos. Además, tiene la finalidad de medir indicadores como: Ordenamiento de datos, Conocimiento aportado, procesamiento de los datos.

Procesamiento de la Información

Método Descriptivo: Al hacer un análisis descriptivo de los datos obtenidos, se necesita hacer un análisis no supervisado o se realiza un análisis clúster, que es una técnica de clasificación automática de datos, donde se clasifica los datos en positivos o negativos para un próximo análisis predictivo.

Método Predictivo: Para hacer un análisis predictivo previamente se aplica un aprendizaje automático de datos para que en un futuro los nuevos datos que ingresan estén clasificados automáticamente en una o más variables.

Procesamiento de la información	Fácil procesamiento de los datos	Nivel de conocimiento Aportado	Alto nivel de predicción de datos
Método Descriptivo			
Método Predictivo			

EVALUACION POR JUICIO DE EXPERTOS

Respetado Juez: Usted ha sido seleccionado para evaluar el instrumento "Ficha de Registro de extracción y análisis de Datos de Redes Sociales". La evaluación del instrumento es de gran relevancia para lograr que sea válido y que los resultados obtenidos a partir de éste sean utilizados eficientemente; aportando tanto al área investigativa. Agradecemos su valiosa colaboración.

I. DATOS GENERALES DEL JUEZ

Nombre del Juez : Winner Agurto Marchan
Grado profesional : Maestría
Doctor ()
Área de Formación Académica :
Clínica () Social ()
Educativa () Organizacional ()
Área de Experiencia Profesional : Catedrático Universitario
Institución donde labora : Universidad Cesar Vallejo - Piura
Tiempo de experiencia profesional en el área:
2 a 4 años
Más de 5 años ()

Mg. Ing. Winner Agurto Marchan
COORDINADOR PROGRAMAS POSTGR
UCV - PIURA

MGTR. WINNER AGURTO MARCHAN

DNI:

40673760

MATRIZ DE VALIDACIÓN DE INSTRUMENTO

NOMBRE DEL INSTRUMENTO : Ficha de Registro de extracción y análisis de Datos de Redes Sociales

OBJETIVO : Este instrumento me permitirá obtener conocimiento sobre el método de extracción y análisis de datos de las redes sociales, como lo son: Facebook y Twitter.

APELLIDOS Y NOMBRE DEL EXPERTO : Winner Agurto Marchan

DNI : 40673760

GRADO ACADÉMICO DEL EVALUADOR : Magister

VALORACIÓN:

MUY DEFICIENTE	DEFICIENTE	ACEPTABLE	BUENA	EXCELENTE
			✓	

Mg. Ing. Winner Agurto Marchan
COORDINADOR PROGRAMAS POSGRUOS
UCV - PIURA

MGTR. Winner Agurto Marchan
DNI: