

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO**  
**FACULTAD DE INGENIERÍA**  
**ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**



**SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA**  
**IDENTIFICACIÓN DE PERFILES DE LECTORES EN UN SEMANARIO**  
**DIGITAL DE CHICLAYO**

**TESIS PARA OPTAR EL TÍTULO DE**  
**INGENIERO DE SISTEMAS Y COMPUTACIÓN**

**AUTOR**

**CRISTHIAN JESUS DELGADO VILLENA**

**ASESOR**

**MARLON EUGENIO VILCHEZ RIVAS**

<https://orcid.org/0000-0003-2979-0731>

**Chiclayo, 2020**

**SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA  
IDENTIFICACIÓN DE PERFILES DE LECTORES EN UN  
SEMANARIO DIGITAL DE CHICLAYO**

PRESENTADA POR:

**CRISTHIAN JESUS DELGADO VILLENA**

A la Facultad de Ingeniería de la  
Universidad Católica Santo Toribio de Mogrovejo  
para optar el título de

**INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR:

María Ysabel Aranguri García  
PRESIDENTE

Jury Yesenia Aquino Trujillo  
SECRETARIO

Marlon Eugenio Vílchez Rivas  
VOCAL

## **DEDICATORIA**

A mis padres por formarme, enseñarme a soñar y  
que nada es imposible para quien lo intenta,  
por brindarme tanto apoyo y hacer esto posible.

A Dios por iluminarme, ponerme a las personas correctas  
y guiarme por el buen camino en todo momento.

A las personas en general y a los que se desempeñan en  
este ámbito, que les despierte el interés en el tema y  
les permita ver los grandes beneficios que trae consigo.

## **AGRADECIMIENTOS**

A mi familia, a mis padres Ramiro y Aurora a mis hermanos Ramiro y Daniel y a Karla por acompañarme, animarme y apoyarme en todo momento.

A la empresa y a la directora Rosa Chambergo Montejo quien me abrió sus puertas y me permitieron llevar esto a cabo, mi infinito agradecimiento.

A mis amigos y futuros colegas: Carlos, Hugo, Sergio, Max, José, Walter, Daniela, por mencionar algunos, por poner su granito de arena cada uno.

A mi asesor, el Ing. Marlon Vílchez por orientarme, resolver mis dudas y pulir la investigación en el desarrollo de esta.

A todos los profesionales e involucrados que participaron en el desarrollo, Gracias totales.

## **SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA IDENTIFICACIÓN DE PERFILES DE LECTORES EN UN SEMANARIO DIGITAL DE CHICLAYO**

*Delgado, Villena, Cristhian Jesus*

*Universidad Católica Santo Toribio de Mogrovejo, Facultad de Ingeniería*

### **RESUMEN**

La presente investigación tiene como propósito la implementación de una solución de inteligencia de negocios con el fin de identificar perfiles de lectores en un semanario digital de Chiclayo.

Las empresas en la actualidad presentan incertidumbre al no conocer las características de sus consumidores, esto genera que no refuercen o tomen estrategias inadecuadas dirigidas a su público.

La organización propuesta como entidad de estudio no hace un análisis de los datos de los lectores debido a que estos no poseen herramientas o tecnologías que brinden dicho análisis.

Se plantea entonces la hipótesis de que la implementación de esta solución logre la identificación de perfiles de lectores en base a una segmentación por características de estos. Esta solución se desarrollará mediante minería de datos con la técnica de segmentación para la elaboración de perfiles de lectores. Finalmente se plantea un sistema web con los datos y estadísticas obtenidas para ser mostradas hacia los usuarios finales.

**PALABRAS CLAVE:** Inteligencia de negocios (BI), Minería de datos, Análisis de datos, Segmentación, Clustering, Perfiles de consumidores.

## **ABSTRACT**

The purpose of this investigation is the implementation of a business intelligence solution in order to identify reader's profiles on a digital weekly in Chiclayo.

In actuality, companies don't know the characteristics of their consumers; this generate many problems and there are not appropriate strategies aimed at their consumers.

The organization that was proposed as an entity for the study, doesn't make an analysis of data from readers because they don't have resources or technologies that can give this benefice.

The hypothesis of the implementation of this solution is the identification of reader profiles are based on a segmentation of these characteristics. This solution is developed through data mining with the segmentation technique for the preparation of reader profiles. Finally, a report of the data and statistics will be presented in a web system to the end users.

**KEYWORDS:** Business intelligence, Datamining, Data analysis, Segmentation, Clustering, Costumer profiles.

## ÍNDICE

<b>I.</b>	<b>INTRODUCCIÓN .....</b>	<b>8</b>
<b>II.</b>	<b>REVISIÓN DE LITERATURA / MARCO TEÓRICO .....</b>	<b>10</b>
<b>III.</b>	<b>MATERIALES Y MÉTODOS .....</b>	<b>14</b>
<b>IV.</b>	<b>RESULTADOS Y DISCUSIÓN .....</b>	<b>15</b>
<b>V.</b>	<b>CONCLUSIONES.....</b>	<b>18</b>
<b>VI.</b>	<b>RECOMENDACIONES.....</b>	<b>19</b>
<b>VII.</b>	<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>19</b>
<b>VIII.</b>	<b>ANEXOS.....</b>	<b>22</b>
	<b>ANEXO N° 01. CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS .....</b>	<b>22</b>
	<b>ANEXO N° 02. INSTRUMENTOS DE RECOLECCIÓN DE DATOS .....</b>	<b>23</b>
	<b>ANEXO N° 03. ESPINA DE ISHIKAWA .....</b>	<b>24</b>
	<b>ANEXO N° 04. ANÁLISIS DE RIESGOS .....</b>	<b>25</b>

## I. INTRODUCCIÓN

En la actualidad mundial globalizada se presenta a menudo el desconocimiento de cuáles son los rasgos o características de los consumidores de una organización o a qué sector están dirigidos [1] (público lector en el caso de telecomunicaciones) sumado a esta interrogante el tema se hace más complejo con un mercado tan competitivo y cambiante, como lo es en la realidad debido a esto surgen dificultades al momento de publicar los contenidos, ya que al estar dirigidos a diferentes tipos de públicos y necesidades tan variadas según el “interés” de cada persona no se llega a captar la atención de los lectores esperados [2]. Para resolver dichas interrogantes se realizan complejos estudios de diferentes aspectos, como herramientas de estrategias de marketing, la segmentación, de manera que con esta división se pueda identificar mejor a los diversos grupos de consumidores, ofreciéndoles contenidos que satisfagan de forma única sus intereses [2]; y mediante la tecnología se puede desarrollar un análisis de los datos pertenecientes al mercado [3]. Mediante soluciones de inteligencia de negocios, que transforman los datos del público en información y conocimiento, se pueden agrupar a las personas por características comunes para ofrecer contenidos más

exclusivos y así dar una posible solución al problema, ya que esta puede traer grandes beneficios para las organizaciones, afectando positivamente la rentabilidad y la satisfacción al consumidor. [1]

En Latinoamérica más del 50% de las organizaciones redirigirán sus inversiones a proyectos que mejoren la comprensión del consumidor en base a una personalización de servicios o productos según la consultora Gartner. La personalización de un producto hacia el consumidor influye en la percepción que éste tiene de un producto y puede afectar a su comportamiento y lealtad [4]. En este sentido la digitalización de los servicios o productos ofrecidos es un paso obligado y clave para conocer las características del consumidor y brindarle una mejor experiencia mediante la exclusividad, personalización e interactividad.

Según Capgemini, 2 de cada 3 inversiones de las organizaciones se centran en mejorar la comprensión y el análisis de las características o perfiles de los consumidores y su comportamiento ante los productos o estrategias ofrecidas [5].

En el Perú Según un informe realizado por el Centro Regional para el Fomento del Libro en América Latina y el Caribe (CERLALC), los peruanos son quienes

más leen periódicos en la región con un 71%, [6] y según una investigación de Ipsos Perú manifiesta que un poco más de la tercera parte del Perú son internautas (38%), es aquí donde se abre un mundo de oportunidades para orientar un mercado al consumidor digital.

En complemento, la consultora Deloitte presentó que el 35% de organizaciones considera que la personalización del producto hacia el cliente en base a sus necesidades es el pilar de la estrategia de negocio. De estas, solo el 23% captura la información del consumidor mientras que solo el 12% la analiza y distribuye, teniendo así mucha información potencial por ser explotada [7].

En la región Lambayeque, según la tesis realizada por los autores Ramos Rojas y Hernández Monteza mostró que el uso de los datos de clientes y su seguimiento posterior mejoraron los resultados de una empresa respecto a sus cifras de éxito, número de clientes, fidelización, etc, reafirmando los beneficios de analizar los datos de los consumidores. [8]

La empresa local propuesta como sujeto de estudio de la presente investigación es del rubro de telecomunicaciones y se dedica a la producción y comercialización de un semanario, contando con un público receptor (lectores) y además a la difusión de contenido por plataformas digitales.

Según la entrevista realizada (Ver anexo 3), la compañía presenta incertidumbre al no conocer las características o perfiles de su público lector en su plataforma digital (página web), trayendo como consecuencia el desconocimiento de la priorización de tipos de titulares y contenido, solo lo intuyéndolo a base de su experiencia en el mercado. En adición, esto puede traer como consecuencia el hecho que no refuerce o priorice sus estrategias y no tenga una mayor allegada o impacto en su público, pudiendo aumentar la satisfacción, fidelización o el número de lectores, impactando así negativamente en la rentabilidad e imagen de la empresa.

La presente investigación abordó esta problemática mediante una solución tecnológica que estudió y analizó los datos de los lectores del semanario para brindar un mejor panorama a los directivos de la empresa y así mejorar los ratios descritos en los párrafos anteriores.

Ante esta realidad, la presente tesis denominada “SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA IDENTIFICACIÓN DE PERFILES DE LECTORES EN UN SEMANARIO DIGITAL DE CHICLAYO.”, se inició con el propósito de caracterizar a los perfiles de la

organización propuesta, para esto es importante formular la siguiente pregunta:

¿De qué manera se identificarán los perfiles de lectores con la implementación de una solución de inteligencia de negocios en un semanario digital de Chiclayo?

Frente a esta pregunta y la necesidad de profundizar el problema, se realizó la investigación del tipo tecnológica aplicada cuya población fue toda la data generada de los lectores. Para ello, se desarrolló una solución de inteligencia de negocios para identificar los perfiles de lectores en un semanario digital de Chiclayo, considerando las siguientes fases de la metodología CRISP-DM: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado y evaluación. Se consideró también el uso de las últimas fases de RUP: diseño, implementación y pruebas porque la metodología de minería de datos no contempla el desarrollo de un sistema que muestre los resultados; para ello se determinaron las variables o características de los perfiles de lectores en un semanario digital de Chiclayo, se determinó qué técnica de BI fue la apropiada para desarrollar la solución, se construyó una solución en base a la técnica propuesta y finalmente se validó la solución.

## **II. REVISIÓN DE LITERATURA / MARCO TEÓRICO**

### **Segmentación de clientes**

La segmentación es una herramienta que busca agrupar personas, datos, etc en diferentes grupos o segmentos, donde cada característica o datos pertenecientes a un segmento son parecidas entre sí y diferentes a los demás clústeres. Se define cómo el proceso de identificar y caracterizar, mediante técnicas estadísticas, subgrupos de consumidores en un mismo mercado que tienen distintas preferencias o intereses, como la identificación de quienes pueden ser parte de una estrategia comercial diferente adaptada a los mismos, con el fin de incrementar la allegada o impacto de los productos y rentabilidad de la empresa [2].

### **Perfil de cliente**

Un perfil de cliente, representa un grupo o segmento del público que tienen intereses, preferencias o comportamientos parecidos. Al identificar diferentes intereses o preferencias, se puede conocer qué es de mayor importancia para ellos a nivel de gustos como persona. Los perfiles les brindan información necesaria a los directivos o estrategias para hacer que la organización se enfoque más en las preferencias de sus consumidores.

Brindan comportamientos, tendencias y que es lo que necesitan o prefieren. Esto se considera en un punto clave al momento de tomar decisiones sobre las decisiones estratégicas de la organización [2].

### **Inteligencia de negocios**

La inteligencia de negocios es la serie de herramientas que permitirán a los usuarios lograr el análisis de datos para el hallazgo de nuevo conocimiento para que el usuario pueda emplearlo en la toma de mejores decisiones para su organización.

### **Análisis de datos**

El análisis de datos es una serie de procedimientos de extracción, limpieza y transformación de datos con la finalidad de conseguir información antes desconocida.

Se define el análisis de datos como:

Técnicas o proceso para analizar datos, de forma que se pueda interpretar los resultados de dicho proceso, planeación de la recolección de datos para lograr un análisis más sencillo, más preciso o más exacto [3].

### **Proceso de extracción de conocimiento (KDD)**

El proceso KDD es el proceso de descubrir de conocimiento a partir de

data histórica almacenada en algún repositorio de datos. Es un proceso iterativo que tiene fundamento en encontrar relaciones en la data existente [9].

Sus etapas son:

#### **Identificación de los objetivos**

Etapas en la que se definen las metas desde la perspectiva del usuario y el conocimiento requerido.

#### **Selección de datos**

Etapas en la que se determinan las fuentes de datos y la clase de información a usar.

#### **Pre procesamiento**

Etapas en la que se prepara y limpia los datos seleccionados desde las distintas fuentes de datos

#### **Transformación**

Etapas en la que modela la data para que encaje en los intentos de análisis y algoritmos. Se ordenan y se dejan listos los datos de una manera necesaria para la próxima etapa.

#### **Minería de datos**

Etapas en la que se realizan métodos para la extracción de patrones y tendencias antes desconocidos, válidos o nuevos que eran desconocidos en la data.

#### **Interpretación**

Etapas en la que se interpretan los patrones obtenidos relevantes basándose en métricas y se realiza una evaluación del resultado.

### **Integración al negocio**

Creación de una interfaz adecuada para utilizar el conocimiento adquirido para resolver los problemas inicialmente planteados.

### **Minería de datos**

La minería de datos proporciona conocimientos acerca de los datos estratégicos, para hallar comportamientos, patrones, segmentos y relaciones no vistas superficialmente en grandes repositorios de datos e infiriendo reglas partiendo de ellos para describir y predecir una posible conducta futura. Las tendencias, reglas y conductas se usan para dar soporte a la toma de decisiones ejecutivas y describir o predecir el impacto de dichas decisiones. Los diferentes tipos de información que permite alcanzar haciendo uso de la minería de datos presentan: secuencias, agrupaciones, pronósticos asociaciones y clasificaciones [10], es una etapa del proceso de descubrimiento de conocimiento (KDD). Tiene como objetivo la generación de información de una serie de datos para poder ser presentada de una manera más entendible.

### **Técnicas de minería de datos**

Existen dos tipos de técnicas de minería de datos: descriptiva o predictiva. Estas

técnicas poseen propias tareas que se pueden lograr a partir del uso de cualquiera de estas [10].

#### **Descriptiva**

Se identifican comportamientos que representan a los datos analizados, estas se pueden dar mediante agrupamiento o reglas de asociación. Esta técnica es útil si se quiere explorar particularidades en los datos, se hace uso del aprendizaje no supervisado, es decir, adquiere conocimiento a partir de los datos ingresados, sin necesidad de la intervención de un operador externo que guíe el comportamiento esperado al sistema [10].

#### **Predictiva**

Se estiman posibles valores en el futuro de una variable en específica. Se trata de un proceso en el que se usa la data histórica para predecir el comportamiento futuro de una variable. Esto se logra a partir del uso de clasificaciones, categorizaciones o regresiones.

A diferencia del modelo descriptivo, los modelos predictivos hacen uso del aprendizaje supervisado pues se logra el aprendizaje mediante la intervención de un externo que indique al sistema el resultado esperado. El atributo a predecir se conoce como variable dependiente u objetivo, mientras que los atributos utilizados para realizar la predicción se

llaman variables independientes o de exploración [10].

### **Algoritmos para tareas de minería de datos**

Por cada técnica de minería de datos existen tareas, por cada tarea que se quiere realizar de minería de datos existen algoritmos para poder lograrlas.

#### **Algoritmos de clasificación**

Se usan para predecir muchas variables discretas basándose en los demás atributos del conjunto de datos [11].

#### **Algoritmos de regresión**

Se usan para predecir muchas variables numéricas continuas, como pérdidas o ganancias, basándose en otros atributos del conjunto de datos [11].

#### **Algoritmos de segmentación**

Se usan para dividir los datos clústeres o en grupos de elementos que presentan características parecidas [11].

#### **Algoritmos de asociación**

Se usan para buscar correlaciones entre distintas propiedades de un grupo de datos. La aplicación más frecuente de este tipo de algoritmo es la generación de reglas de asociación, que pueden servir en el análisis de un carro de compras [11].

#### **Algoritmos de análisis de secuencias**

Se usan para resumir los pasos más frecuentes o episodios en los datos, como

un conjunto de clics en una página web o un conjunto de registros que le antecedan al mantenimiento del equipo [11].

### **Analítica web**

La analítica web es el análisis, medición, recopilación y reporte de data web con el objetivo de mejorar el uso de la data web. Las herramientas de análisis web ayudan a las organizaciones a medir los indicadores y las tendencias de las campañas de atracción de usuarios que es útil para la investigación y descubrimiento del mercado [12]. El análisis web brinda conocimiento sobre los visitantes de una página web y el número de visitas.

En los últimos tiempos, el Internet ha experimentado un crecimiento muy elevado, ampliando los mercados al sector digital o virtual. Esto ha generado nuevas conductas por parte de los usuarios y a un mercado más cambiante e incierto, caracterizado por una gran variedad de datos que deben ser transformados en conocimiento útil para las organizaciones. Aquí es cuando la analítica web desempeña un rol crítico como herramienta para disminuir las dudas de la organización y poder tomar decisiones más precisas basadas en la información relevante [13].

### Google analytics

Google Analytics es un instrumento de analítica web de Google, que brinda información de los usuarios que llegan a las páginas web según los visitantes y la conducta que se desarrolla en la página. Se puede generar conocimiento como el seguimiento de los visitantes privilegiados, los resultados de las estrategias de marketing online, las fuentes de tráfico, duración de las sesiones, contenidos visitados, etc. Además, Proporciona datos sobre el origen del visitante, el sistema operativo utilizado, esto revela más datos demográficos, o brinda información sobre cuántos usuarios accedieron al sitio con dispositivos móviles. De este modo, se puede determinar el valor de página de diferentes páginas [14].

### Metodología CRISP-DM

Es una metodología para soluciones de minería de datos (Cross-Industry Standard Process for Data Mining) [26]. Siendo la metodología que más usan los expertos en este ámbito. Es conocida por brindar la aplicabilidad de resultados al entorno de la organización escogida y permitir interactuar por las fases de desarrollo hacia adelante y atrás (dinámica entra estas).

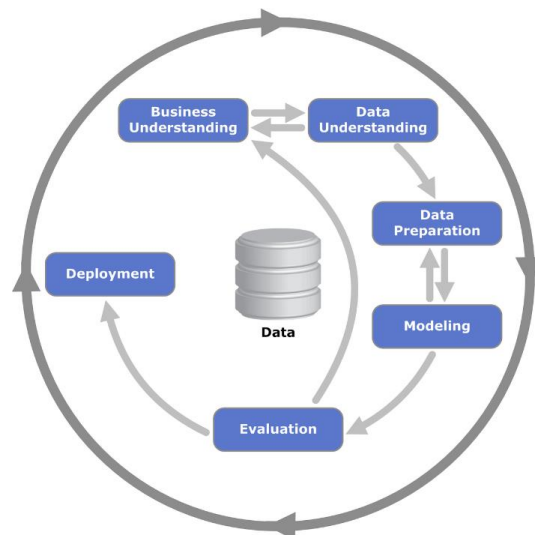


Fig. 1. Fases de CRISP-DM [26]

### III. MATERIALES Y MÉTODOS

El tipo de la investigación fue tecnológica aplicada y el nivel post-experimental porque se realizó una prueba después de aplicar la solución al grupo definido [16]. De acuerdo al desarrollo de la investigación, el diseño de contrastación que se utilizó fue post experimental. El diseño se diagrama como sigue

**G X O**

Donde:

**G** = Data de los lectores.

**X** = Solución de Inteligencia de negocios.

**O** = Medición post experimental.

La población objeto del estudio estuvo constituida por todos los datos

demográficos de los lectores de la organización propuesta.

La muestra estuvo constituida por la población siendo todos los datos demográficos de lectores, ya que mi solución está dirigida a este público. La técnica de muestreo que se aplicó es No Probabilístico, porque no se hizo un cálculo matemático ya que se eligieron todos los datos existentes. Como criterio de selección se tomaron los datos de lectores ya que es lo que mi solución usó, además de contar con información variada respecto a estos.

#### **Variables**

Las variables que se han utilizado como elementos básicos en el desarrollo de la hipótesis están identificadas como variable independiente los perfiles de lectores y variable dependiente la solución de inteligencia de negocios.

#### **IV. RESULTADOS Y DISCUSIÓN**

##### **Determinación de las variables o características de los perfiles de lectores**

Se hizo una búsqueda exhaustiva de cómo se segmenta el mercado apoyándose de los diferentes proyectos de investigación mencionados, se concluyó que para lograr una adecuada segmentación se hace uso de los datos demográficos de los clientes y para esto

se usaron todos los datos demográficos obtenidos, siendo los siguientes:

El semanario cuenta con una página web de donde se obtuvieron los datos mediante google analytics para la realización de este proyecto.

#### **Lectores**

Esta tabla contiene información de los lectores del semanario, que en la presente investigación son las variables que se usaron para determinar los perfiles de lectores. Los campos de esta tabla son los siguientes:

**Users:** Es un campo de número tipo entero. Es el número de lectores del semanario.

**Age:** Es un campo de tipo numérico que representa la edad de cada lector.

**Gender:** Es un campo de texto que representa el sexo del alumno.

**Affinity Category:** Es un campo de texto donde indica el tipo de noticia que prefiere leer el usuario.

**City:** Es un campo de texto que tiene la ubicación de cada lector

**Country:** Es un campo de texto que tiene la ubicación de cada lector

**Region:** Es un campo de texto que tiene la ubicación de cada lector

##### **Determinación de la técnica para la solución propuesta**

Se listaron las técnicas existentes de la minería de datos y se compararon, como en el siguiente cuadro.

TABLA I  
Cuadro de comparación de técnicas

Técnicas / Características	Técnica clasificación	Técnica regresión	Técnica segmentación o clustering	Técnica asociación	Técnica análisis de secuencias
Descriptiva			X	X	
Predictiva	X	X			X
Variablenuméricas		X	X		X
Variablen cualitativas	X		X	X	X
Segmentación	X		X		

Se determinó la técnica más apropiada para la solución la cual permite describir las variables discretas, numéricas y segmentarlas, en este caso fue clustering, además con la revisión de los proyectos de investigación mencionados anteriormente se pudo reafirmar lo escogido.

### Construcción de la solución en base a la técnica propuesta

Luego de que se determinó la técnica propuesta, se procedió a la construcción de la solución de acuerdo a la

metodología escogida (en este caso CRISP -DM).

### Construcción del modelo

Se comenzó con la construcción del modelo con la carga de datos que servirán de parámetros para el modelo de minería de datos.

Luego se determinó el número óptimo de clústeres según el método elbow.

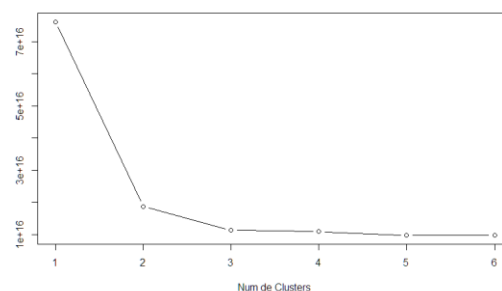


Fig. 2. Método de elbow para obtener el número óptimo de clústeres

Se observó la formación de 6 segmentos. Se estableció entonces que el número óptimo de clústeres es 6.

Después, se usó la técnica clustering para modelar la solución, los argumentos que recibe esta técnica son una matriz numérica de datos, o un objeto (como un vector numérico o un marco de datos con todas las columnas numéricas). El número máximo de iteraciones y los centros que son el número de centros o agrupaciones que se realizarán, en este caso fueron 6, ya que es el número óptimo que estableció el método elbow.

### **Evaluación del modelo**

Para la evaluación del modelo se usaron algoritmos de validación de los cuales se obtuvo un 82.60% de exactitud.

### **Determinación de los siguientes pasos**

Se desarrolló un sistema web para plasmar la información y la interactividad del usuario y los reportes.

### **Validación de la solución propuesta**

Para la validación de la solución, en la fase de construcción del modelo se aplicaron algoritmos, los cuales determinaron el número óptimo de perfiles y evaluaron el modelo arrojando un alto porcentaje de exactitud (82.6%). Posteriormente se hicieron pruebas de funcionalidad o usabilidad (caja negra y blanca) con los usuarios finales, a los cuales está dirigido el sistema.

### **Discusión**

En la presente tesis hemos realizado una solución de inteligencia de negocios, con un enfoque a la segmentación de los lectores en base a características comunes de la organización propuesta.

Durante el desarrollo de la tesis se determinó que la forma de segmentar a los lectores es mediante los datos demográficos, ya que describen la información del lector como tal, como lo manifiesta Morelos [20] y De la cruz [15], en sus investigaciones sostienen que la segmentación es en base a los

datos demográficos para describir al consumidor.

La forma de llegar a la segmentación para este proyecto basado en el análisis de datos fue usando la técnica de clustering la cual logró segmentar a los lectores. Asimismo, Morelos [20] manifiesta que para una segmentación descriptiva se hace uso de técnicas de clustering, ya que es la más adecuada para llevar consigo dicho proceso.

Para la construcción de la solución se hizo uso de las técnicas de análisis de datos (minería) y los resultados se plasmaron en un sistema web por la interconectividad que tienen con las herramientas usadas, los autores Quepuy [19], De la cruz [15] y Morelos [20] nos dicen que, para la explotación de datos se usan técnicas de análisis de datos y los resultados pueden ser mostrados de diversas formas, como en sistemas web o de escritorio.

La implementación de la solución basada en la segmentación de lectores contribuye a establecer mejores estrategias para conseguir más allegada y satisfacción de los lectores, los autores De la cruz [15], Quepuy [19] y Morelos [20], comprobaron que al aplicar análisis de datos se contribuyó a mejorar el proceso al que estuvo enfocado la solución.

Finalmente, los resultados encontrados permitieron identificar los perfiles de lectores haciendo uso de la metodología CRISP-DM, teniendo como entregable final un sistema web.

## V. CONCLUSIONES

1. Se logró identificar las características de los perfiles de lectores con respecto a la data existente (en este caso de investigación aproximadamente 65 mil datos de lectores y en aumento) y la forma de hacerlo es mediante los datos demográficos de las personas (se usaron 4 variables pertenecientes a los lectores como edad, genero, interese, etc.). Los cuales permiten caracterizarlos de una forma en la que se puedan distinguir en distintos tipos de perfiles

2. Se determinó la técnica a ser usada para la construcción de la solución mediante una comparación la cual evaluó a 5 técnicas (clustering, regresión, clasificación, etc.) y como resultado se obtuvo a la técnica clustering. Se concluye que, para la segmentación y descripción de perfiles con variables descriptivas y cuantitativas en la minería de datos la técnica más adecuada es clustering, permitiendo agrupar mediante características comunes a las personas.

3. Se construyó la solución en base a la técnica propuesta clustering, en la cual se elaboró un modelo estadístico

haciendo uso de algoritmos (elbow, kmeans, etc), el cual determino 6 perfiles evaluando las características comunes de los lectores y tiene como entregable final un sistema web donde se muestran todos los datos de los perfiles resultantes, siendo la mejor manera de poder visualizar este tipo de datos, visto que en su mayoría usan conexiones con otras herramientas en tiempo real (Google analytics, PowerBI y Rstudio), las cuales hacen uso de internet.

4. Se validó el modelo con la aplicación de algoritmos (como Hartigan Wong, Lloyd, etc.) los cuales arrojaron un alto porcentaje de exactitud (82.6%) estando dentro del rango de exactitud óptimo (70% - 90%) indicando que la solución es exacta, luego se validó la solución mediante 2 pruebas funcionales (3 de caja negra y 3 de caja blanca) para cada caso de uso o interfaz del sistema, las cuales arrojaron parámetros correctos de reportes y de la gestión de usuarios, teniendo así la aprobación y satisfacción de los usuarios finales.

5. La implementación de la solución basada en la segmentación de los perfiles de lectores contribuye a establecer, enfocar y reforzar mejor las estrategias de la organización propuesta con el fin de conseguir más allegada (número y alcance, del cual se ha identificado múltiples lugares del mundo

teniendo como zona de acción principal Lambayeque y la Libertad con un 89%) y también mejorar la satisfacción del lector ofreciéndole un contenido más personalizado.

## VI. RECOMENDACIONES

1. Se sugiere aplicar otros métodos, tipos de tecnología o enfoques disciplinarios (Estadística, marketing, etc.) y comparar los resultados con esta investigación.
2. Se recomienda la generación de nuevos reportes y funcionalidad para el sistema, con respecto a las nuevas necesidades de la organización que puedan surgir en un futuro.
3. Se aconseja hacer más validaciones u otro tipo de validaciones que permitan mejorar la solución.

## VII. REFERENCIAS

### BIBLIOGRÁFICAS

- [1] P. Armstrong. G. Camara Ibáñez. D. y Cruz Roche. I. Kotler, Marketing (10a ed.), Madrid: Pearson Education, 2004.
- [2] M. Bernal, Procedimiento para desarrollar estudios de mercados en organizaciones., 2005. [Online]. Available: [https://degerencia.com/articulo/procedimiento\\_para\\_desarrollar\\_estudios\\_mercados\\_en\\_organizaciones/](https://degerencia.com/articulo/procedimiento_para_desarrollar_estudios_mercados_en_organizaciones/).
- [3] M. S. Lewis-Beck, Data analysis: an introduction, London: Sage Publications Inc., 1995.
- [4] C. Gartner, Gartner data and Analytics Summit, 2018. [Online]. Available: <https://www.gartner.com/en/newsroom/press-releases/2018-01-17-gartner-announces-data-analytics-summit-2018>.
- [5] Capgemini, «World Quality Report» 2018. [Online]. Available: <https://www.capgemini.com/service/world-quality-report-2018-19/>.
- [6] (CERLALC), Fomento del Libro en América Latina y el Caribe, «El libro en cifras,» 2017. [Online]. Available: [http://cerlalc.org/wpcontent/uploads/2017/02/PUBLICACIONES\\_OLB\\_Libro-en-cifras-10\\_vf\\_311216.pdf](http://cerlalc.org/wpcontent/uploads/2017/02/PUBLICACIONES_OLB_Libro-en-cifras-10_vf_311216.pdf).
- [7] Deloitte, “¿Cómo “respiran cliente” las empresas peruanas?” 2017. [Online]. Available: [https://www2.deloitte.com/content/dam/Deloitte/pe/Documents/finance/CEX\\_Per%C3%BA.pdf](https://www2.deloitte.com/content/dam/Deloitte/pe/Documents/finance/CEX_Per%C3%BA.pdf).
- [8] H. Hernández, G. Monteza y. A. Ramos., “Estrategia de CRM y su relación con la fidelización de los clientes en una empresa comercial en insulas, olmos, Lambayeque,.” Tesis de pregrado, Lambayeque, 2018.
- [9] J. García y R. Reyes, “El proceso de descubrimiento de conocimiento en bases de datos” Repositorio académico

digital, 2005. [Online]. Available: <http://eprints.uanl.mx/10160/>

[10] M. Kamber y J. Han, *Data mining: Concepts and Techniques.*, USA: The Morgan Kaufmann Series in Data Management Systems. 2a Ed., 2006.

[11] O. O. Duncan & S. C. Owen, *Algoritmos de minería de datos (Analysis Services: Minería de datos).*

[Online]. Available: <https://opdhsblobprod04.blob.core.windows.net/contents/9503556bbe040c896b8a287c97a7a0a/e464e0c31006947837214e56572054a4?s v=2018-03-28&sr=b&si=ReadPolicy&sig=F2dgPg9PqCooxLDcpp3e5W6POcN%2B3WTp%2FoUjjDAITHo%3D&st=2020-12-03T19%3A11%3A04Z&se=2020-12-04T19%3A21%3A04Z>.

[12] G.S. Linoff y M. J. A. Berry, *Mining the Web: Transforming customer data into customer value*, John Wiley & Son, 1a Ed, 2001.

[13] S. Maldonado, “Analítica Web: Medir para triunfar”, MV Consultoría, España, 2009.

[14] B. Plaza, “Monitoring web traffic source effectiveness with Google Analytics: An experiment with time series.” *Aslib Proceedings*, vol. 61, n° 5, 2009.

[15] K. D. D. Gutiérrez, “Segmentación de clientes con Inteligencia Analítica para personalizar

las ventas de los servicios de las agencias turísticas”, Tesis Mg, Lima - Perú: Universidad Peruana Unión, 2017.

[16] C. A. Bernal, “Proceso de la investigación científica,” en *Metodología de la investigación científica: administración, economía, humanidades y ciencias sociales*, 3ra. ed. Bogotá D.C., Colombia: Pearson Educación, 2010.

[17] G. A. N. Morales Gonzáles, “Perfiles de clientes para proponer estrategias de segmentación según modelos de vehículos KIA”, Tesis Mg, Trujillo: upao, 2017.

[18] Z. E. M. Rodríguez, “Aplicación de la minería de datos distribuida usando algoritmo de clustering k-means para mejorar la calidad de servicios de las organizaciones modernas”, Tesis Mg, Lima: UNMSM, 2015.

[19] C. E. Q. Perleche, “Sistema de gestión y soporte de toma de decisiones basado en algoritmos de Bayes y clúster para mejorar los procesos analíticos del área comercial de una empresa educativa”, Tesis de pregrado, Chiclayo: USAT, 2016.

[20] K. A. M. Tapias, “Sistema para caracterización de perfiles de clientes de la empresa zona t”, Tesis de pregrado, Cartagena - Colombia: Universidad de Cartagena, 2014.

[21] S. F. C. Jiménez, “Minería de datos para segmentación de clientes en la empresa tecnológica Master pc”, Tesis de pregrado, Loja - Ecuador: Universidad Nacional de Loja, 2015.

[22] V. G. Cotrina, “Aplicación de la metodología Crisp-DM a un proyecto de minería de datos en el entorno universitario”, Tesis de pregrado, Universidad Carlos III, Madrid - España, 2015.

[23] J. Burby y S. Atchison, “Actionable Web Analytics: Using Data to Make Smart Business Decisions”, 1a Ed, Sybex, 2007.

[24] J. B. Bustíos Polar y A. I. Díaz Limo, “Análisis de segmentación del mercado de clientes de microfinanzas en Bagua Grande” Tesis Mg, Universidad Nacional Pedro Ruiz Gallo, Lambayeque - Perú, 2017.

[25] S. E. G. Alejandro, “Aplicación Web Utilizando Estrategias De Segmentación con apoyo del marketing 1x1 para mejorar la satisfacción de clientes en Batan Grande Restaurante” Tesis de pregrado, Universidad católica santo toribio de Mogrovejo, Chiclayo-Perú, 2016.

[26] IBM, Manual CRISP-DM, 2012. [Online]. Available:<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwikz47X0L>

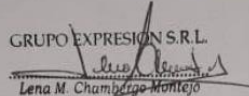
LtAhXpYt8KHY4mCCcQFjABegQIAxAC&url=ftp%3A%2F%2Fftp.software.ibm.com%2Fsoftware%2Fanalytics%2Fspss%2Fdocumentation%2Fmodeler%2F15.0%2Fes%2FCRISPDM.pdf&usg=AOvVaw15ETZyrbXCOpKRTrVaWMGX.

**VIII. ANEXOS****ANEXO N° 01. CONSTANCIA DE APROBACIÓN DEL PRODUCTO  
ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS**

Se constata que el sistema “SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA IDENTIFICACIÓN DE PERFILES DE LECTORES EN UN SEMANARIO DIGITAL DE CHICLAYO” desarrollado por “CRISTHIAN JESUS DELGADO VILLENNA” con DNI N° 73263061 funciona y muestra los datos correctamente teniendo en cuenta los parámetros y criterios establecidos anteriormente en las fichas de revisión, cumpliendo con las necesidades existentes esta herramienta sirve de ayuda para la identificación y segmentación de perfiles en el semanario del grupo expresión.

Chiclayo, 26 de noviembre del 2019

GRUPO EXPRESIÓN S.R.L.

  
Lena M. Chamblago-Montejo  
GERENTE

## **ANEXO N° 02. INSTRUMENTOS DE RECOLECCIÓN DE DATOS**

### Guía para la entrevista.

**1. ¿Tienen algún problema en su entorno?**

Pues hemos percibido que cuando lanzamos contenido a través de las plataformas digitales a veces los lectores no llegan a darle la acogida que se espera teniendo cifras de 1000 a 3000 vistas cuando se esperan cifras de 10000 a 30000, como es el caso de algunos titulares en los que se invierte para publicitarlos y ni aun así se llega a tener la acogida esperada. No se tiene claro el porqué de la situación siendo así una interrogante para nosotros la falta de interés por parte de nuestros lectores.

**2. ¿Cómo se relacionan los problemas con la información que manejan?**

Bueno creemos que la causa del problema podría ser que desconozcamos los intereses de nuestro público y al no conocer esta información no podemos enfocar nuestras estrategias como la exclusividad del contenido brindado.

**3. ¿Considera que es importante el conocimiento de los intereses de sus lectores?**

Sí, pensamos que es una de las prioridades de la empresa, ya que si conociéramos sus características, gustos y preferencias podríamos personalizar nuestro contenido con el fin de impactar en su interés.

**4. ¿Considera que es importante el uso de la tecnología?**

Pensamos que en la actualidad la tecnología es una herramienta muy poderosa y que sin ella las organizaciones no saldrían adelante, es crítica para competir ante otras empresas y además simplifica las actividades de la empresa.

**5. ¿Manejan algún tipo de sistema en la empresa?**

Sí, contamos con una web la cual brinda contenido hacia los lectores, además de eso ofrecemos contenido variado mediante redes sociales y YouTube.

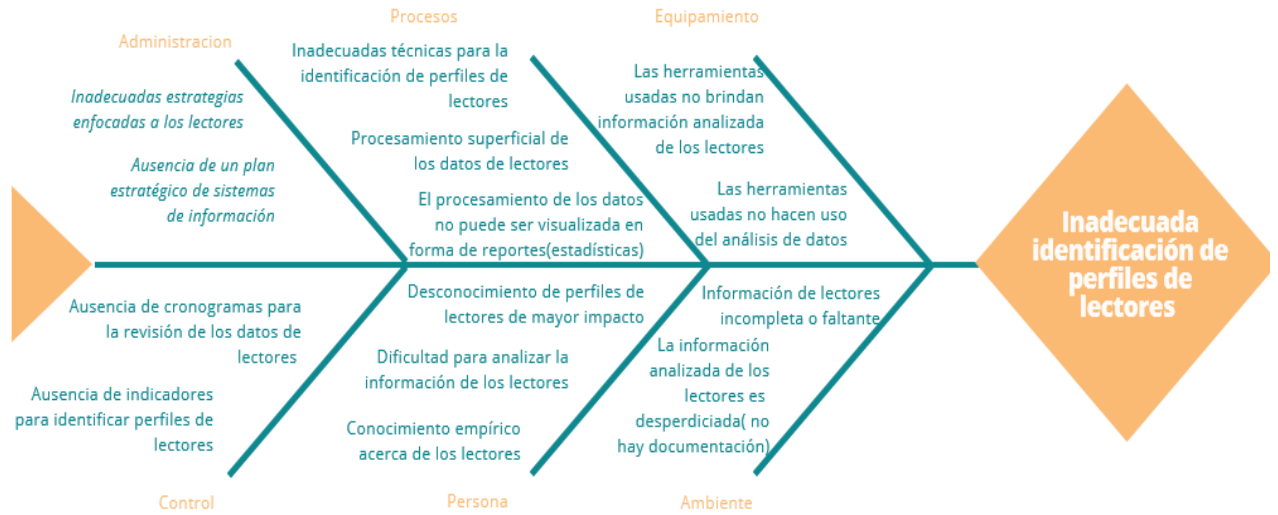
**6. ¿Dónde almacenan su información?**

La información se almacena en un repositorio tercerizado por otra empresa la cual brinda el servicio de gestión del sitio web, además de los repositorios propios de las plataformas digitales.

**7. ¿Qué espera de una solución tecnológica?**

Esperamos que nos ayude a mejorar nuestra personalización de contenido en base a una determinación de características e información de los lectores.]

### ANEXO N° 03. ESPINA DE ISHIKAWA



## ANEXO N° 04. ANÁLISIS DE RIESGOS

### 1. Datos generales

- **Tesista** : Cristhian Jesus Delgado Villena
- **Fecha inicial** : Julio del 2019
- **Fecha final** : Julio del 2020

### 2. Alcance del proyecto

Se desarrolló la solución de inteligencia de negocios para la organización elegida, con la finalidad de identificar los perfiles de lectores haciendo uso de la data y de las herramientas adecuadas propuestas.

El sistema implementado permitió identificar los perfiles de lectores en un semanario digital de Chiclayo.

La información se presentó mediante reportes de datos y estadísticas, para ello es necesario el uso de las distintas herramientas propuestas.

### 3. Interesados (Stakeholders)

Durante el desarrollo de la presente tesis se identificó a los siguientes interesados:

- **Internos**

TABLA II  
INTERESADOS INTERNOS

Interesado	Participación
Directivos	Brindarán el acceso y la información necesaria además de que usarán la solución propuesta.
Trabajadores	Brindan información necesaria del negocio.

- **Externos**

TABLA III  
INTERESADOS EXTERNOS

Interesado	Participación
Lectores	Se usará la información de los lectores para desarrollar la solución.
Asesor	Guiará al investigador en el transcurso del desarrollo de la investigación

#### **4. Beneficios**

Los beneficios que se obtuvieron con el producto que se ha desarrollado son:

- Se identifican las características de los lectores.
- Se segmenta a los lectores para brindar una mayor exclusividad y personalización de contenido.
- Se brinda reportes de datos y estadísticas con respecto a los datos analizados de los lectores.
- Se fortalecen o priorizan las estrategias con respecto a los lectores.
- Se mejora la allegada o recibimiento del contenido difundido por parte de los lectores.

## 5. Etapas de desarrollo

Para el desarrollo del producto de la presente tesis se ha realizado considerando las etapas de la Metodología CRISP-DM:

### – Matriz de riesgos

Entre los riesgos identificados se mencionan:

TABLA IIIII  
MATRIZ DE RIESGOS

Código del riesgo	Descripción del riesgo	Fase afectada	Causa raíz	Entregables afectados	Estimación probabilidad	Objetivo afectado	Estimación Impacto	Probabilidad por impacto	Nivel de riesgo
RE - 1	Cierre de la empresa o se produzca alguna alteración crucial para el desarrollo de la solución	Comprensión del negocio	Por algún motivo la empresa cierre o se produzca algo anormal dentro de la empresa para el desarrollo de la solución	Todo el proyecto	1	Alcance	5	5	BAJO
						Tiempo	5	5	
						Costo	2	2	
						Calidad	5	5	
						TOTAL PROBABILIDAD POR IMPACTO		17	
RE - 2	Los datos recolectados no sean suficientes para desarrollar la solución	Comprensión de los datos	No se obtuvieron los datos suficientes para desarrollar la solución	Solución e informe propuesto	3	Alcance	5	15	MEDIO
						Tiempo	4	12	
						Costo	2	6	
						Calidad	5	15	
						TOTAL PROBABILIDAD POR IMPACTO		48	
RE - 3	Los datos no se puedan limpiar, tratar e integrar	Preparación de los datos	Los datos están sucios o existen datos faltantes difíciles de intuir	Solución e informe propuesto	3	Alcance	4	12	MEDIO
						Tiempo	4	12	
						Costo	2	6	
						Calidad	4	12	
						TOTAL PROBABILIDAD POR IMPACTO		42	
RE - 4	Que la selección de la técnica sea inadecuada y se construya un modelo erróneo	Modelado	La técnica seleccionada no sea la adecuada y se construya una solución no adecuada.	Solución e informe propuesto	2	Alcance	5	10	MEDIO
						Tiempo	5	10	
						Costo	5	10	
						Calidad	5	10	
						TOTAL PROBABILIDAD POR IMPACTO		40	
RE - 5	La solución no sea funcional.	Evaluación	Por algún motivo la solución generada no haga lo que debe hacer.	Solución e informe propuesto	2	Alcance	5	10	MEDIO
						Tiempo	5	10	
						Costo	5	10	
						Calidad	5	10	
						TOTAL PROBABILIDAD POR IMPACTO		40	

						TOTAL PROBABILIDAD POR IMPACTO		40	
RE - 6	No se concluya con el informe final.	Despliegue e implementación	Por falta de tiempo o algún agente externo o tercero no se pueda concluir con el informe.	Informe propuesto	3	Alcance	3	15	MEDIO
						Tiempo	5	15	
						Costo	4	12	
						Calidad	4	12	

– **Matriz salvaguarda de riesgos**

Entre los planes de mitigación para superar riesgos identificados se mencionan:

TABLA IV  
MATRIZ DE SALVAGUARDA DE RIESGOS

Código del riesgo	Amenaza / Oportunidad	Descripción del riesgo	Fase	Nivel de riesgo	Tipo de respuesta	Responsable	Plan de mitigación
RE – 1	Amenaza	Cierre de la empresa o se produzca alguna alteración crucial para el desarrollo de la solución	Comprensión del negocio	BAJO	Salvaguarda	Tesista	<ul style="list-style-type: none"> <li>✓ Coordinación con la parte directiva.</li> </ul>
RE – 2	Amenaza	Los datos recolectados no sean suficientes para desarrollar la solución	Comprensión de los datos	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> <li>✓ Búsqueda de métodos para recolectar o extraer datos.</li> <li>✓ Extracción de datos de calidad.</li> <li>✓ Generación de datos mediante alguna herramienta.</li> </ul>
RE – 3	Amenaza	Los datos no se puedan limpiar, tratar e integrar	Preparación de los datos	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> <li>✓ Búsqueda tipos de datos compatibles.</li> <li>✓ Búsqueda de datos sin necesidad de limpiar o tratar.</li> </ul>
RE – 4	Amenaza	Que la selección de la técnica sea inadecuada y se construya un modelo erróneo	Modelado	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> <li>✓ Búsqueda de técnicas adecuadas.</li> <li>✓ Evaluación del modelo construido.</li> </ul>
RE – 5	Amenaza	La solución no sea funcional.	Evaluación	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> <li>✓ Desarrollo de pruebas funcionales.</li> <li>✓ Orientación por parte de los usuarios finales (directivos).</li> </ul>
RE – 6	Amenaza	No se concluya con el informe final.	Despliegue o Distribución	MEDIO	Salvaguarda	Tesista	<ul style="list-style-type: none"> <li>✓ Generación de un cronograma de avances</li> <li>✓ Desarrollo de avances de acuerdo a lo planificado.</li> </ul>