



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería Mecánica Industrial

**DISEÑO DE INVESTIGACIÓN PARA LA PROPUESTA DE UN PROCESO DE
SEGMENTACIÓN PARA LA VENTA DE PRODUCTOS DIGITALES A CLIENTES
EXISTENTES EN UNA INSTITUCIÓN BANCARIA DE LA CIUDAD DE GUATEMALA**

Mynor Moshé Mendoza Chon

Asesorado por el Msc. Ing. Bryan Josué Bernal Corey

Guatemala, febrero de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE INVESTIGACIÓN PARA LA PROPUESTA DE UN PROCESO DE
SEGMENTACIÓN PARA LA VENTA DE PRODUCTOS DIGITALES A CLIENTES
EXISTENTES EN UNA INSTITUCIÓN BANCARIA DE LA CIUDAD DE GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

MYNOR MOSHÉ MENDOZA CHON

ASESORADO POR EL MSC. ING. BRYAN JOSUÉ BERNAL COREY

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO INDUSTRIAL

GUATEMALA, FEBRERO DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANA	Inga. Aurelia Anabela Cordova Estrada
EXAMINADOR	Ing. Aldo Ozaeta Santiago
EXAMINADOR	Ing. Erwin Danilo González Trejo
EXAMINADOR	Ing. Jaime Humberto Batten Esquivel
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**DISEÑO DE INVESTIGACIÓN PARA LA PROPUESTA DE UN PROCESO DE
SEGMENTACIÓN PARA LA VENTA DE PRODUCTOS DIGITALES A CLIENTES
EXISTENTES EN UNA INSTITUCIÓN BANCARIA DE LA CIUDAD DE GUATEMALA**

Tema que me fuera asignado por la Dirección de Escuela de Estudios de Postgrado con fecha 17 de agosto de 2021.

Mynor Moshé Mendoza Chon

Ref. EEPFI-1142-2021
Guatemala, 17 de agosto de 2021

Director
César Ernesto Urquizú Rodas
Escuela de Ingeniería Mecánica Industrial
Presente.

Estimado Ing. Urquizú:

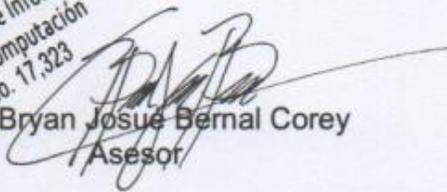
Reciba un cordial saludo de la Escuela de Estudios de Postgrado. El propósito de la presente es para informarle que se ha revisado y aprobado el **DISEÑO DE INVESTIGACIÓN: PROPUESTA DE UN PROCESO DE SEGMENTACIÓN PARA LA VENTA DE PRODUCTOS DIGITALES A CLIENTES EXISTENTES EN UNA INSTITUCIÓN BANCARIA DE LA CIUDAD DE GUATEMALA**, presentado por el estudiante **Mynor Moshé Mendoza Chon** carné número **200216473**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la Maestría en Artes en Gestión Industrial.

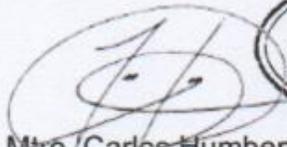
Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

Atentamente,

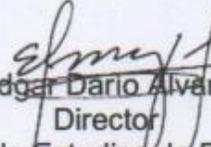
"Id y Enseñad a Todos"

Bryan Josué Bernal Corey
Ingeniero en Sistemas de Información
y Ciencias de la Computación
Colegio No. 17,323


Bryan Josué Bernal Corey
Asesor


Mtro. Carlos Humberto Aroche
Coordinador de Maestría
Gestión Industrial – Fin de Semana




Mtro. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería





EEP-EIMI-003-2022

El Director de la Escuela de Ingeniería Mecánica Industrial de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **PROPUESTA DE UN PROCESO DE SEGMENTACIÓN PARA LA VENTA DE PRODUCTOS DIGITALES A CLIENTES EXISTENTES EN UNA INSTITUCIÓN BANCARIA DE LA CIUDAD DE GUATEMALA**, presentado por el estudiante universitario **Mynor Moshé Mendoza Chon**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería en esta modalidad.

ID Y ENSEÑAD A TODOS



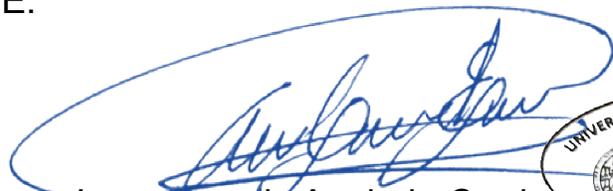
Ing. César Ernesto Urquizú Rodas
Director

Escuela de Ingeniería Mecánica Industrial

Guatemala, enero de 2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería Mecánica Industrial, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN PARA LA PROPUESTA DE UN PROCESO DE SEGMENTACIÓN PARA LA VENTA DE PRODUCTOS DIGITALES A CLIENTES EXISTENTES EN UNA INSTITUCIÓN BANCARIA DE LA CIUDAD DE GUATEMALA**, presentado por: **Mynor Moshé Mendoza Chon**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



ingra. Aurelia Anabela Cordova Estrada

Decana

Guatemala, febrero de 2022

AACE/gaoc

ACTO QUE DEDICO A:

Dios	Por haberme permitido realizar una más de mis metas.
Mis padres	Mynor Mendoza y Vilma Chon de Mendoza. Por haberme traído al mundo y guiado a través de él.
Mis hermanos	Kemly, Javier y Heber Mendoza, por su apoyo.
Mi esposa	Estri López de Mendoza por sus sabias enseñanzas y consejos.
Mis hijos	Sharon y Hector Mendoza. Por ser la inspiración en mi vida.

AGRADECIMIENTOS A:

**Universidad de San
Carlos de Guatemala**

Por ser el alma *mater* que me permitió nutrirme de conocimientos.

Facultad de Ingeniería

Por proporcionarme los conocimientos que me han permitido realizar este trabajo de graduación.

**Familia y amigos en
general**

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	V
SÍMBOLOS	VII
GLOSARIO	IX
RESUMEN.....	XI
1. INTRODUCCIÓN	1
2. ANTECEDENTES	5
3. PLANTEAMIENTO DEL PROBLEMA	9
3.1. Descripción del problema	9
3.2. Pregunta central	10
3.3. Preguntas de investigación.....	10
3.4. Delimitación del problema	10
3.5. Viabilidad.....	10
3.6. Consecuencias de la investigación.....	11
4. JUSTIFICACIÓN	13
5. OBJETIVOS	15
5.1. General.....	15
5.2. Específicos	15
6. NECESIDADES A CUBRIR Y ESQUEMA DE SOLUCIÓN.....	17

7.	MARCO TEÓRICO	19
7.1.	Capítulo I: institución bancaria privada.....	19
7.1.1.	Definición de institución bancaria	19
7.1.2.	Tipos de operaciones y servicios	20
7.1.2.1.	Operaciones y servicios bancarios tradicionales	20
7.1.2.2.	Banca virtual.....	22
7.1.3.	Clases de bancos	23
7.1.3.1.	Banca especializada	24
7.1.3.2.	Banca universal o multibanca.....	24
7.2.	Capitulo II: productos digitales	24
7.2.1.	Pago digital – pago con el móvil	25
7.2.2.	Tarjetas de crédito digitales	26
7.2.3.	Pagos de servicios a terceros	26
7.2.4.	Transferencias internacionales	26
7.2.5.	Doble factor de autenticación	27
7.3.	Capitulo III: segmentación.....	27
7.3.1.	Crear modelo de aprendizaje supervisado para ofrecer productos a clientes existentes	28
7.4.	Análisis factorial	30
7.4.1.	Tipos de análisis factorial	31
7.4.2.	Aplicaciones	32
7.5.	Correlación.....	32
7.6.	Test de hipótesis de r	36
7.7.	Intervalo de confianza del coeficiente de correlación	37
7.8.	Presentación de la correlación	39
7.9.	Interpretación de la correlación	40
7.10.	Coeficiente de correlación de los rangos de Spearman	42
7.11.	Cálculo de correlaciones	45

7.12.	K-means	47
7.13.	Teorema igualdad de Fisher	52
7.14.	Objetivo del método K-means	52
7.15.	Método del codo de Jambú	54
8.	PROPUESTA DE ÍNDICE DE CONTENIDOS	57
9.	METODOLOGÍA.....	59
9.1.	Enfoque	59
9.2.	Diseño de investigación.....	59
9.3.	Tipo	59
9.4.	Alcance.....	60
9.5.	Variables e indicadores	60
9.5.1.	Variables independientes	60
9.5.2.	Variables dependientes	60
9.6.	Fases de investigación	61
9.6.1.	Fase I. Revisión	61
9.6.2.	Fase II. Diagnóstico	62
9.6.3.	Fase III. Determinar variables de importancia	62
9.6.4.	Fase IV. Elaborar sistema de medición	62
9.6.5.	Fase V. Propuesta de la metodología de segmentación	62
9.7.	Instrumentos.....	63
9.7.1.	Técnicas	63
9.7.2.	Instrumentos	63
9.7.3.	Población y muestra	63
9.7.3.1.	Población.....	64
9.8.	Resultados esperados.....	64

10.	TÉCNICAS DE ANÁLISIS DE INFORMACIÓN	65
11.	CRONOGRAMA	67
12.	RECURSOS NECESARIOS Y FACTIBILIDAD DEL ESTUDIO	69
13.	REFERENCIAS	71
14.	APÉNDICES	75

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Correlación entre talla y peso.....	40
2.	Representación gráfica de matriz de correlación	47
3.	Segmentos con centroides escogidos aleatoriamente	48
4.	Cálculo de centroide	49
5.	Nuevos centroides calculados.....	50
6.	Evolución en la segmentación en cada iteración	53
7.	Muestra de segmentos vs. codo de Jambú.....	55
8.	Cronograma de actividades	67

TABLAS

I.	Cálculo del coeficiente de correlación de Pearson entre las variables talla y peso en 20 niños varones.....	35
II.	icotina existente en sangre vs. cigarrillo.....	43
III.	Edad vs. rango	44
IV.	perativización de variables.....	61
V.	Recursos.....	70

SÍMBOLOS

Símbolo	Significado
cm	Centímetro
kg	Kilogramo
%	Porcentaje
Q	Quetzales

GLOSARIO

Algoritmo	Es una serie ordenada de instrucciones, pasos o procesos que llevan a la solución de un determinado problema.
CCV	Código que trae la parte de atrás de la tarjeta. Se utiliza para validar pagos en los cuales no hay lectura de banca magnética o chop, principalmente telefónicos o por internet.
Clustering	Es una tarea que consiste en agrupar un conjunto de objetos (no etiquetados), en subconjuntos de objetos llamados <i>Clústeres</i> o segmentos.
Coefficiente	Número o parámetro que se escribe a la izquierda de una variable o incógnita y que indica el número de veces que este debe multiplicarse.
Correlación	Correspondencia o relación recíproca entre dos o más acciones o fenómenos.
Interpolación	Consiste en hallar un dato dentro de un intervalo en el que se conocen los valores en los extremos.

Matriz	Arreglo bidimensional de números consistente en cantidades abstractas que pueden sumarse y multiplicarse entre sí.
NFC	Es un campo electromagnético de corto alcance 5 o 10 cm que se utiliza para comunicar principalmente a <i>wearables</i> o pagos sin contacto.
Polígono	En geometría, un polígono es una figura geométrica plana compuesta por una secuencia finita de segmentos rectos consecutivos que encierran una región en el plano. Estos segmentos son llamados lados, y los puntos en que se interceptan se llaman vértices.
Segmentación	La segmentación de mercado divide un mercado en segmentos más pequeños de compradores que tienen diferentes necesidades, características y comportamientos que requieren estrategias o mezclas de <i>marketing</i> diferenciadas.

RESUMEN

El propósito de la investigación realizada es el de poder realizar y proponer una segmentación a clientes actuales de la institución bancaria ya que en el proceso de ventas, la necesidad de ofrecer a los clientes los productos que se adecúan a cada uno de ellos es esencial principalmente porque la competencia también incursiona en la personalización de la oferta que realiza, al realizar una revisión del proceso de ofrecimiento de productos digitales a clientes existentes se encontró que no existe una segmentación efectiva y que esto ocasiona el gasto en anuncios a clientes que tienen una muy baja probabilidad de aceptar los productos objetivos.

Esta es de enfoque mixto, diseño no experimental, de alcance descriptivo porque se realizará recolección de datos para realizar una propuesta; tipo transversal con un tiempo determinado de 6 meses para la realización del estudio con toda la población que son los clientes de los cuales se cuenta con información para realizar un proceso de segmentación.

Al personalizar la oferta no sólo disminuyen los costos de cada campaña dado que digitalmente se paga por cada alcance, es decir cada anuncio que se le muestra a un cliente que no aplica es dinero perdido, sino que se maximiza la cantidad de clientes a los cuales se puede llegar. Para esto históricamente se utilizan datos demográficos sin embargo al ir avanzando la tecnología se han desarrollado procesos de segmentación en los cuales se puede llegar a encontrar grupos con gustos y necesidades similares, lo que hace posible el poder predecir cuando alguien pueda ser un objetivo ideal para cada campaña comercial.

1. INTRODUCCIÓN

En el proceso de ventas de cualquier institución uno de los elementos más importantes, independientemente de cuál sea el producto o servicio que se desea comercializar, es la capacidad que se tiene para determinar cuál es el mejor perfil y producto que se adecue a las necesidades del cliente; ya que principalmente una institución de servicios, se basa en la premisa de que el cliente debe de percibir una alta satisfacción por los servicios prestados, por la facilidad que tiene de buscar otros servicios iguales en instituciones distintas.

Esta investigación se propone como un trabajo de sistematización, con la finalidad de segmentar a distintos grupos de clientes que compartan características similares y que por lo tanto puedan ser tratados bajo premisas aprendidas de estos mismos grupos, principalmente para poder realizar una oferta adecuada a las características y variables que puedan ser estudiadas y que puedan indicar cuales son las necesidades de estos clientes.

Teniendo en cuenta estos conceptos se vuelve necesario realizar la evaluación de las técnicas que se tienen para segmentar a estos clientes y proporcionar propuestas fundamentadas que permitan realizar una mejor segmentación y que por consecuencia de esta segmentación la aceptación de los productos ofrecidos y el nivel de satisfacción del cliente crezca.

La importancia de diseñar un sistema de segmentación adecuado es consecuencia de la problemática que se tiene en la organización que proviene de no conocer adecuadamente al cliente y esto resulta en ofrecimientos

excesivos de opciones que no se adecuan a las necesidades del cliente, esto repercute en consecuencias graves a la gestión.

El principal inconveniente es que se tienen recursos limitados, con los cuales operar mensualmente y el ofrecer un producto que no se adecue al cliente es un desperdicio de recursos; en primer lugar ya que en la mayoría de casos, estos ofrecimientos se realizan por medio de medios digitales especializados quienes cobran una cantidad exacta por pautar al público objetivo que se le indica, esto puede ser por características demográficas, pero también puede ser dirigido exactamente a la persona que uno desea por medio de una variable que permita identificarlo y contactarlo.

El segundo inconveniente es que si un cliente recibe publicidad que no se adecua o no es de interés, para el con el pasar del tiempo puede llegar a insensibilizarse a la publicidad, lo que causará que en el futuro pueda ser la oferta adecuada a la persona; sin embargo, al estar insensibilizado a la publicidad producto de una mala gestión de acercamiento, este no acepte dicho producto.

La presente investigación tiene una metodología con un enfoque mixto ya que tiene variables cuantitativas y cualitativas. Además, es un diseño no experimental, un tipo y alcance descriptivo, porque se hará recolección de datos y solamente es una propuesta.

La necesidad identificada es aumentar el contacto correcto y eliminar el contacto incorrecto, para eficientizar el proceso y mejorar los niveles de alcance con los mismos recursos invertidos.

El diseño de investigación en esta área es viable, ya que es la encargada de utilizar y coordinar la pauta y ofrecimiento, por lo que sería responsabilidad del área el lograr un mejor proceso.

Lo resultados que se esperan es el identificar las variables demográficas y otras variables que puedan ser analizadas como patrones de uso de los productos, para poder crear un perfil de cliente al cual se le generé un portafolio de productos y ofrecimiento.

El beneficio que se espera en este caso es lograr una mayor tasa de aceptación, lo que repercutirá en que con los mismos recursos monetarios en pauta se logre una mayor base de resultados, independientemente de cuál sea el producto por ofrecer.

A continuación, se presenta el esquema de solución:

Se propone trabajar en 4 fases que consisten en teoría existente, análisis de la situación, diseño de solución según resultados arrojados y creación de un nuevo proceso de segmentación.

El primer capítulo se enfocará en el marco teórico donde se concentrará en la información relevante del trabajo. Como también el desarrollo del planteamiento del problema como los objetivos y el alcance de la investigación, con ayuda de la técnica K-means y la herramienta R-Studio que por su capacidad de procesamiento permite evaluar el universo completo en pocos días, antes de aplicar dicha técnica se estarán revisando todas las variables de las cuales se tiene acceso y que ayudará a segmentar al cliente.

En el segundo capítulo se hará la presentación de resultados.

En el tercer capítulo se hará la discusión de resultados.

2. ANTECEDENTES

Steven Fortune (1986) según indica que se realizó un estudio acerca de los polígonos de Thiessen, los cuales toman este nombre en honor al meteorólogo estadounidense Alfred Thiessen, dicho estudio habla sobre el uso de estos polígonos para poder construir sobre el plano euclídeo una pequeña partición a partir de una construcción geométrica. Anteriormente también fueron estudiados en 1907 por el matemático ruso Guerogiuo Voronói y es gracias a estos estudios que también se le conocen como Teselación de Voronoi o Diagramas de Voronoi, aunque también se tiene al matemático alemán Gustav Lejeune Dirichlet quién más de 50 años antes aproximadamente en 1850 hizo estudios sobre estos objetos y toman por nombre Teselación de Dirichlet.

Los diagramas de Voronoi son actualmente considerados como la forma más simple de realizar una interpolación, y se basan en la distancia euclidiana, esta toma una gran ventaja y es apropiada cuando lo que se está trabajando es información o datos cualitativos. para poder crear los diagramas se deben de unir los puntos entre sí, trazando las mediatrices de los segmentos de Unión.

En un espacio bidimensional cuando se quiere determinar los polígonos lo que se debe de utilizar son las intersecciones de estas mediatrices alrededor de un conjunto de puntos de control, con esto se está logrando que los polígonos que se generaron sean equidistantes a los puntos contiguos y así se logra designar el área de influencia que tendrán, con esta área de influencia segmentada asociada a clientes de la institución bancaria se puede generar un modelo efectivo en el cual los conjuntos similares tendrán un comportamiento similar.

Estos diagramas generados podrán ser utilizados para realizar regiones de un mismo plano de clientes y de esta manera para efectos de estudio será más fácil estudiar las variables que los diferencian, también agruparlos en grupos o clasificarlos por sus preferencias.

Moreno (2004), indica en su publicación sobre la clasificación del vecino más cercano indicando que es una de las técnicas que más se utilizan y esta consiste tomando como base un conjunto de objetos que han sido etiquetados, asignar una etiqueta correspondiente al objeto que se desconoce la del vecino más cercano en el conjunto que se estudió, tomando en cuenta la distancia que tengan.

Actualmente y gracias a las computadoras de alto rendimiento se pueden utilizar gran variedad de algoritmos, para encontrar el vecino más cercano de una manera eficiente y, tomando en cuenta que cada día se logran encontrar mejores resultados de clasificación al utilizar los k vecinos, que se encuentran más cercanos cuando se toma en cuenta k mayor a la unidad, dichos algoritmos se pueden ir extendiendo cada vez más para encontrar estos k vecinos, pero habrá que tomar en cuenta que siempre que se aumente y el valor de k el tiempo que se tomará en clasificar estar aumentando de manera proporcional, dependiendo de cuantas variables se evalúan y de cuantos objetos se están ingresando en el algoritmo.

Al crear los diagramas anteriormente mencionados y también creando un algoritmo de clasificación, con la intención de que el algoritmo realice un aprendizaje supervisado, sobre el total de clientes existentes se podrán predecir los nuevos clientes a manera de clasificarlos y saber cuál será el segmento a asignarles.

Tapis (2001) en la tesis de su autoría indica que la segmentación de mercado, o como también se le conoce y qué es una segmentación de clientes o una segmentación de audiencias, habla sobre un método por el cual tomando todos los clientes se pueden dividir los potenciales en varios grupos, sabiendo a cuál grupo pertenece cada uno se pueden tomar acciones específicas para tratarlos de manera especial y personalizada dependiendo del segmento en el cual sean asignados.

Según dicha segmentación las categorías principales siempre se podrán llegar a dividir en categorías mejor segmentadas y estas pueden estar definidas por cualquier tipo de variable tomando en cuenta su ubicación, su tiempo en la empresa, la edad o cualquier otra característica que se considere importante según el modelo efectuado cuando se logra realizar de manera personalizada una campaña la efectividad de esta irá aumentando, ya que se puede escoger el mensaje específico para la audiencia correcta tomando en cuenta que se irá adaptando a cada uno de los segmentos.

3. PLANTEAMIENTO DEL PROBLEMA

La inexistencia de un procedimiento adecuado, para el ofrecimiento de soluciones digitales a clientes acrecentando los costos de obtención de clientes.

3.1. Descripción del problema

En la industria bancaria internacional la búsqueda por optimizar los recursos y generar un mayor valor agregado al servicio de los clientes, se han implementado soluciones digitales las cuales pueden ser ofrecidas también por canales digitales; sin embargo, es un reto al día de hoy el poder dirigir eficientemente estas campañas hacia el grupo objetivo y de esta manera tener un proceso adecuado de segmentación, para poder ofrecer los productos apropiados a las personas apropiadas y de esta manera lograr un rendimiento mayor por los gastos de mercadeo y a la misma vez evitar el estar importunando a clientes con anuncios que no aplican.

En un Banco del sistema guatemalteco, se ofrecen soluciones a clientes existentes para su uso en plataformas digitales, sin embargo, no existe una metodología de ofrecimiento para saber que clientes puedan aceptar el ofrecimiento, lo que ocasiona gastos excesivos en publicidad dirigida y poca eficiencia en el dinero invertido en publicidad, lo que no permite llegar a los crecimientos estándar del mercado.

3.2. Pregunta central

¿Cómo un proceso de segmentación mejorará la venta de productos digitales para clientes existentes en una institución bancaria?

3.3. Preguntas de investigación

- ¿Cómo se hace la segmentación de ventas a los clientes existente en la institución bancaria?
- ¿Qué factores críticos tiene la segmentación en ventas al no contar con recursos digitales?
- ¿Qué ventajas tiene utilizar soluciones digitales para la modernización de atención en ventas de los clientes existentes en la institución?

3.4. Delimitación del problema

El estudio se realizará en la gerencia de transformación digital de un banco del sistema. Durante un lapso de octubre-noviembre 2020. Tomando de referencia los niveles de aceptación y alcance del año 2020.

3.5. Viabilidad

Los recursos para la realización del proyecto serán otorgados por la empresa

3.6. Consecuencias de la investigación

Entre las causas que general el problema, está la falta de un procedimiento para saber que producto ofrecer al cliente, lo que causa 2 grandes inconvenientes, que todos los clientes reciban información y esto deriva en que el cliente pierde interés en los anuncios de la empresa y segundo es que el presupuesto es limitado para los anuncios, lo que también causa que se gasten y desperdicien recursos en clientes que no aplican para el producto pero que se gaste en ofrecérseles.

Con la aplicación de una metodología de segmentación y ofrecimiento se logrará el no ofrecer productos a sectores que no aplican y adicional se obtendrá una mayor ratio de aceptación de los productos al ofrecerles únicamente a los clientes que muestran una alta tasa de probabilidad de aceptación lo que logrará atraer a muchos más clientes con la misma cantidad de recursos, sin invadir al cliente con publicidad.

Si no se realiza esta mejora, no solo no se tendrán resultados esperados de conversión de clientes, sino que el cliente podrá verse cansado de recibir tanta publicidad y dejará simplemente de prestar atención, lo que ocasionará que en un futuro sea aún más difícil lograr una tasa de conversión a niveles que se tienen hoy en día.

4. JUSTIFICACIÓN

El siguiente trabajo de investigación, se enfoca dentro de la línea de investigación de inteligencia de negocios de la Maestría en Gestión Industrial. El curso relacionado con la línea de investigación es ingeniería de Sistemas de planificación industrial.

La importancia de las soluciones digitales se deba a la economía de dichos servicios frente a servicios tradicionales por lo que las tendencias internacionales dirigen a adoptar estas nuevas formas de negocios, lo cual permite mayor capacidad de respuesta a los clientes y atención en horarios mayores a los tradicionales.

La necesidad que se tiene de realizar este trabajo de investigación es desarrollar un procedimiento adecuado para segmentar y ofrecer efectivamente dichos productos a clientes existentes y de esta manera lograr mejores rendimientos con los presupuestos que se tienen.

La motivación del investigador que suscribe el presente trabajo se basa en el interés de implementar el conocimiento adquirido, respecto a inteligencia de negocios en el área de trabajo, y contribuir a la mejora continua del departamento de transformación digital, para facilitar el trabajo a los empleados, a fin de mejorar la productividad del área, para aportar al desarrollo y crecimiento del departamento.

Los beneficios que se obtienen al implementar un proceso efectivo de segmentación y ofrecimiento es principalmente la mejora en las tasas de

conversión por anuncio, así como evitar el ofrecimiento innecesario a clientes por medio de anuncios y ofertas de productos a los cuales no aplica, generando mayores ingresos con menor inversión.

Los principales beneficiarios de esta investigación son los clientes al tener un producto que se adecue a sus necesidades y que puedan tener desde la comodidad de su hogar con un horario de atención mayor al de los canales tradicionales. La empresa al incrementar la tasa de aceptación de productos reduciendo los costos en publicidad y los empleados, al adquirir conocimientos sobre la metodología y un proceso eficiente para la realización de sus tareas en el área de trabajo.

5. OBJETIVOS

5.1. General

Proponer un proceso de segmentación para la venta de productos digitales a clientes existentes en una institución bancaria de la ciudad de Guatemala.

5.2. Específicos

- Identificar la forma en que la empresa hace la segmentación de clientes existentes.
- Analizar los factores críticos para la solución de la segmentación en ventas al no contar con recursos digitales.
- Determinar las ventajas que tiene la utilización de soluciones digitales en la segmentación de ventas en una empresa bancaria para los clientes existentes.

6. NECESIDADES A CUBRIR Y ESQUEMA DE SOLUCIÓN

La principal necesidad es encontrar un proceso de segmentación, para la venta de productos digitales, a clientes existentes en una institución bancaria de la ciudad de Guatemala. La falta de una segmentación real ocasiona que se desperdicien recursos en presentar anuncios a clientes, a los cuales no aplica dicha publicidad lo cual se traduce en un malgasto de recursos.

El estudio de investigación propone un proceso de segmentación para la venta de productos digitales a cliente existentes, herramienta con la cual se tendrá una mayor tasa de conversión y disminución de gastos de publicidad.

Esquema de solución

- Revisión documental.
- Diagnóstico de los resultados del año 2021.
- Conocer el proceso que se ejecuta.
- Diseñar nuevo modelo y proceso.
- Evaluar nuevo proceso.
- Proponer segmentación y anuncios personalizados por tipo de cliente.

7. MARCO TEÓRICO

7.1. Capítulo I: institución bancaria privada

Consiste en una gestión altamente profesionalizada y global del patrimonio de un cliente. Se trata de satisfacer las necesidades de inversión, planificación patrimonial, financiera y fiscal de personas o grupos familiares con un elevado patrimonio.

7.1.1. Definición de institución bancaria

La definición de una institución bancaria principio por ser una empresa mercantil, la cual está constituida y bajo los estatutos de la legislación general de la República y también leyes que regulan específicamente el sector financiero, la función de estas instituciones es:

La intermediación financiera bancaria, consistente en la realización habitual, en forma pública o privada, de actividades que consistan en la captación de dinero, o cualquier instrumento representativo del mismo, del público, tales como la recepción de depósitos, colocación de bonos, títulos u otras obligaciones, destinándolo al financiamiento de cualquier naturaleza, sin importar la forma jurídica que adopten dichas captaciones y financiamientos. (Ley de bancos y grupos financieros, 2002, p.3)

Las actividades de una institución bancaria actualmente no se suscriben únicamente a su función fundamental, sino que también tienen actividades distintas a la intermediación financiera pueden actuar como asesores,

intermediar valores, arrendamiento financiero, entre otros, lo que las convierte en grupos financieros.

7.1.2. Tipos de operaciones y servicios

La principal función de un banco es la intermediación financiera, esta consiste en actividades que se realizan normalmente en forma pública o privada tales como: colocación de bonos, títulos, recepción de depósitos o cualquier otra obligación la cual se destina al financiamiento de bienes o servicios que pudieran ser de cualquier naturaleza licita. Las operaciones que puedan realizar variarán y pueden ser físicas como virtuales, las cuales se detallan a continuación.

7.1.2.1. Operaciones y servicios bancarios tradicionales

Tomando en cuenta las operaciones más importantes y generales que se realizan en los bancos se mencionan a continuación las siguientes:

- Operaciones activas

El propósito de estas operaciones es canalizar recursos financieros, pero también se puede otro tipo de servicios, también están las que funcionan como el derecho por parte del Banco de ejercer contra terceros, sin tomar en cuenta la forma jurídica al momento de formalizar o su registro contable, un ejemplo puede ser el otorgar créditos; inversión de valores los cuales pueden variar entre el corto mediano y largo plazo; se toma cooperación activa también la constitución de depósitos en bancos extranjeros; arrendamiento factoraje, tarjeta de crédito, incluyendo la emisión, la operación y otras.

- Operaciones pasivas

Se refiere a cuyo propósito es captar recursos financieros y que a su vez el banco garantiza la seguridad y recuperación de los mismos, para con el cliente final sin tomar en cuenta su forma jurídica de formalización o de su registro contable, ejemplo: depósitos monetarios, a plazo que a su vez pueden ser de corto mediano o largo plazo, emisión de deuda u obligación por medio de títulos; refinanciamiento por medio de otros bancos que puedan ser nacionales o extranjeros; y otras.

- Operaciones de confianza

Se refieren a todas las que son realizadas por las instituciones bancarias y que actúan como mandatarias, cuyo propósito es el de prestar servicios que no necesariamente implican una intermediación financiera, éstas no originan derecho y tampoco crean compromiso alguno; sin embargo, esto siempre debe quedar debidamente consignado cuando se realizan los convenios o se firman contratos entre las partes contratantes. Las principales operaciones que se toman como de confianza son: cobros y pagos a terceros o por cuenta ajena; acciones; administración de fideicomisos, títulos de crédito y otras.

- Pasivos contingentes

La naturaleza de estas operaciones no crea una obligación inmediatamente, sin embargo, la falta de cumplimiento que tenga cualquiera de las partes involucradas en estas transacciones, podría derivar en problemas graves para la institución, como, por ejemplo: otorgar garantías o fianzas; emitir cartas de crédito de garantía.

Dado que el registro de estas operaciones se realiza, en cuentas llamadas de orden y son por su naturaleza transacciones expuestas a riesgos que están fuera de balances, se requiere un mayor cuidado y una constante supervisión por parte del banco ya que son administrados.

- **Servicios**

Los servicios son productos no tangibles que las instituciones prestan y que son complementos de sus operaciones diarias hacia los clientes. Estos servicios deben de registrarse en cuentas de orden y realizarse jurídicamente utilizando contratos donde se estipula la prestación de los servicios, que pueden ser por comisión, por mediación u otros.

Entre estos servicios se pueden mencionar: cobranza, intermediación para transferencias cablegráficas; cambio de moneda extranjera ya sea en documentos o en efectivo, arrendamiento de Casillas de seguridad y otras.

7.1.2.2. Banca virtual

Esta se ha popularizado mucho principalmente a partir del año 2020, tomando en cuenta los tiempos de pandemia, puesto que permite una mayor disponibilidad para los clientes sin tomar en cuenta las restricciones de movilidad y horario, está se ha ido entregando al cliente por 2 etapas principalmente:

La primera generación, su principal componente que eran componentes electrónicos capaces de automatizar servicios e introduce hacia el cliente los conceptos de autoservicio, ya que la principal actividad es realizada por el cliente mismo, sus principales herramientas fueron los cajeros automáticos y la

banca telefónica, atendida por respuestas programables las cuales van guiando al cliente en su interacción para con el banco y respondiendo a sus necesidades en base a las respuestas que el mismo cliente va dando.

La segunda se refiere a una digitalización completa, utilizando principalmente la computadora, en el inicio por medio de conexiones a internet y datos, segundo el teléfono móvil con acceso a internet, lo cual terminaría de automatizar el autoservicio y en el autogestionamiento por parte del cliente minimizando en gran parte la interacción de un empleado bancario hacia el cliente, relegándolo a un papel de asesoramiento únicamente más no de servicio como tal.

La banca virtual permite minimizar los costos puesto que no es necesario espacio físico para atender al cliente y minimiza la cantidad de operarios llamados receptores o secretarias de atención al cliente, relegándolos a un papel de asesoramiento mucho más especializado y automatizando todas las gestiones o transacciones de baja complejidad masificando la capacidad de atención sin necesidad de crecer físicamente de manera proporcional, además permite al usuario final mayores tiempos de atención que puede llegar hasta las 24 horas sin depender de un asociado o eliminando la necesidad de llegar físicamente a un lugar, exponiéndose a las dificultades que ésta plaza física pueda tener en cuestión de distancia, horario o tráfico.

7.1.3. Clases de bancos

La banca es un negocio universal y es por esto que se puede especializar en varios segmentos, por el tipo de actividad o por el tipo de clientes, que se les está prestando un servicio, y dentro de estos segmentos se pueden encontrar:

7.1.3.1. Banca especializada

En este segmento se encuentran todas las bancas que atienden a segmentos con necesidades muy específicas y principalmente las que son excluyentes en muchos casos del mercado financiero. en estos casos se realizan operaciones de intermediación muy específicas definidas y autorizadas por la ley financiera y se enmarcan únicamente en su giro específico.

7.1.3.2. Banca universal o multibanca

Actualmente este es el tipo de banca que más se encuentra en el mercado actual ya que en estos casos pueden dedicarse, no sólo a la intermediación financiera, si no que pueden englobar muchos otros servicios de valor agregado; englobando así todas las actividades financieras que tradicionalmente no pertenece a la banca, lo que permite que éstas puedan ofrecer no sólo en su oficina principal, sino en todas las agencias que puedan llegar a tener en el territorio operaciones en valores, almacenaje, seguros, fideicomiso y muchos otros servicios asociados a la intermediación financiera, pero que no tradicionalmente son productos bancarios.

7.2. Capítulo II: productos digitales

Actualmente los productos digitales han ganado en muchos casos y principalmente en las bancas jóvenes, mayor cuota de mercado que los productos tradicionales, tomando en cuenta la penetración de teléfonos inteligentes en la población y la masificación de acceso a internet, permite que los crecimientos actuales se concentren en este segmento.

Aunado al creciente acceso a internet y teléfonos inteligentes, también se encuentra una creciente tendencia global de empresas no financieras pero que prestan servicios comúnmente llamadas Fintech, han permitido evolucionar rápidamente en productos más acercados a las tendencias de los clientes o usuarios finales, acortando las distancias físicas que antes no permitían bancarizar a dichos sectores, es por eso que muchos bancos han decidido implementar fuertes estrategias de digitalización y volcar sus productos a una transformación digital, tomando en cuenta que estos son mucho menos costosos económicamente que sus pares tradicionales, entre estos se encuentran:

7.2.1. Pago digital – pago con el móvil

Esta es la principal atracción en términos de servicios que un banco puede llegar a prestar; puesto que no sólo mejora sustancialmente la experiencia del usuario, sino que aporta altas tasas de seguridad y conveniencia al proceso de realizar un pago por medio del teléfono o cualquier dispositivo vestible, eliminando los fraudes por clonación de banda magnética o por hurto físico de tarjetas, ya que en la mayoría de bancos desarrollados, para poder realizar un pago por medio de un dispositivo móvil, tendrá que validar con la huella digital o algún otro dispositivo de seguridad que el mismo teléfono contenga, esto quiere decir que aunque el teléfono móvil se encuentra expuesto a un hurto el mismo no podrá ser utilizado para realizar pagos ya que la huella digital no podrá ser duplicada.

También hay otros tipos de pagos que dependen de la tecnología NFC (*Near Field Communication*), como lo son dispositivos vestibles que se refieren a relojes, anillos o cualquier dispositivo que se pueda adjuntar una etiqueta NFC,

pero que sin embargo no cuentan con la popularidad o penetración que al día de hoy cuentan el teléfono móvil.

7.2.2. Tarjetas de crédito digitales

Es la que puede ser obtenida por medio 100 % digital sin necesidad de acudir a una sucursal bancaria, sino que el 100 % del proceso se realiza de manera digital, procesos que van desde identificación efectiva del cliente para evitar suplantación de identidad, hasta la emisión de un número de tarjeta que no necesita ser troquelado en un plástico de seguridad, ni contar con elementos como chip o banda magnética puesto, que se asocia directamente a un teléfono móvil o una sucursal digital, para realizar pagos tanto por medio de tecnología NFC o por medio de corroboración de numeración, más fecha de vencimiento más código CCV (*Card Verification Value*).

7.2.3. Pagos de servicios a terceros

Se refiere a todos los pagos que son de una empresa distinta a la de la institución financiera, pero que sin embargo se pueden realizar por medio de las sucursales físicas o digitales que el banco tiene para atender a sus propios clientes, en estos casos no hay distinción del tipo de servicio, del cual pueda originarse dicho cobro, ya que la institución únicamente presta el servicio de obtención de fondos y liquidación de pagos independientemente de cuáles hayan originado dichos cargos.

7.2.4. Transferencias internacionales

Son transferencias cablegráficas o "wire transfers", y son dirigidas al extranjero, sin importar el monto estas pueden ser dirigidas a todos los países

que cuenten con autorización para utilizar bancos de Estados Unidos como intermediarios, ya que en casi todo el mundo se utilizan bancos intermediarios para hacer llegar a un tercer banco el dinero.

7.2.5. Doble factor de autenticación

Este es un dispositivo que permite al cliente identificarse digitalmente por medio de las plataformas del Banco y es distinto al factor principal que es un usuario y contraseña. Este puede ser físico el cual es denominado Token y es un dispositivo que usualmente otorga un número generado por un algoritmo, usualmente de 6 dígitos, que permite confirmar al banco que la persona que está digitando es aquella a la que se le otorgó el dispositivo, este generalmente está programado para vencer en un tiempo de 5 años, principalmente por la vida de la batería pero también porque entre más tiempo se tiene registro de los números, más fácil sería para una persona descifrar el algoritmo que los genera, también los hay en forma digital los cuales pueden ser una aplicación instalable en un dispositivo celular, el cual no vence nunca, ya que el algoritmo puede ser cambiado remotamente y también puede ser un mensaje de texto o correo electrónico denominado OTP (*One Time Password*) por sus siglas en inglés, enviado a los registros que el cliente haya otorgado al iniciar la relación con el Banco, este número suele vencer en un tiempo de 1 a 5 minutos.

7.3. Capítulo III: segmentación

Es un método que pretende limitar los clientes objetivos posibles a los que dirigirse.

7.3.1. Crear modelo de aprendizaje supervisado para ofrecer productos a clientes existentes

Para crear modelos de aprendizaje supervisado se tendrá que revisar la teoría que fundamenta las bases de la presente investigación. Empezando por la normalización de la información y el trato que se debe de otorgar a las variables independientemente, cuál es su tipo para obtener patrones reales de búsqueda de los clientes, pueden ser de varios tipos.

Cualitativamente se refieren a cualidades que no se pueden tratar como números, pero que sin embargo otorga cierta diferenciación que servirá al modelo para establecer grupos específicos o segmentos, que luego serán analizados para entender sus especificaciones, cuando se tiene una variable cualitativa, ésta a su vez también puede ser nominal que se refiere a que al ser nominal pueden haber varios tipos; pero que sin embargo, ningún tipo puede llegar a tener un valor más que el otro ya que no tienen un orden específico por ejemplo es el estatus civil del cliente este puede ser viudo, casado, soltero, unido, divorciado.

Todos los clientes pueden ser asignados a cualquiera de las 5 categorías anteriormente mencionadas, sin embargo el hecho de que un cliente sea viudo o que otro cliente sea soltero no necesariamente significa que se puede ordenar numéricamente dando mayor peso a una o a otra, simplemente son distinciones dentro del grupo estudiado, cuando se ve este tipo de variables también se tienen las que son nominales puesto que no se muestran como un número, pero que sí pueden ser ordenadas de mayor a menor aunque se refieran a cualidades esas cualidades pueden ser evaluadas y asignadas un valor, en estos casos.

Uno de los ejemplos más fáciles de reconocer es el grado académico que pueda llegar a tener un cliente, un grado de licenciatura será mayor que un grado de nivel medio o nivel básico, entonces, aunque se refiera a una variable cualitativa, una cualidad esta cualidad sí puede ser ordenada de mayor a menor o de menor a mayor.

Para continuar y teniendo entendimiento sobre las variables que se van a tener y que pueden llegar a ayudar a definir los segmentos, así como su representación, el siguiente paso es realizar una minería de datos, se le llama minería de datos porque ésta se refiere a una cantidad muy grande de información puesto que se está hablando de cientos de miles de datos y cada dato a su vez puede llegar a tener una cantidad infinita de variables que se deseen estudiar.

El principal objetivo de la minería de datos es el poder descubrir patrones que puedan llegar a ser de interés y que puedan ser utilizados en el entendimiento del grupo que se va a estudiar, esos patrones en general se convertirán en reglas de asociación o clasificación.

Para esto es indispensable el poder obtener el método idóneo que únicamente puede ser realizado por medio de algoritmos utilizados por la minería de datos.

Primero se realiza un *clustering* para definir un grupo de clientes, ya sea jurídicos o personas que tengan características similares por medio de un modelo no supervisado, es decir se analizarán los registros con etiquetas de clase o categorías que puedan tener, esto para poder buscar generalidades y/o patrones en estos pequeños clúster de clientes para poder agruparlos en categorías más comunes y ya en base a cada clúster generar un modelo de

recomendación, con un modelo supervisado que pueda analizar los con variables cuantitativas tanto discretas como continuas.

El objetivo de la clasificación es encontrar un modelo (una función o algoritmo) para predecir la clase a la que pertenecería cada registro, esta asignación de una clase se debe hacer con la mayor precisión posible.

En ese caso como lo que se busca como primer paso, es lograr segmentar a todos los clientes en un clúster donde tenga características similares con otros clientes por lo que se debe de utilizar el método de las K-medias (*k. means*).

Clustering o clusterización: esta es una clasificación que se realiza sin supervisión de otras bases o también se puede definir como aprendizaje no supervisado, es bastante parecido a la discriminación o clasificación, con una pequeña variación de que estos grupos no están predefinidos. El objetivo principal es partir o segmentar los individuos o conjunto de datos en estudio y éstos a su vez pueden ser disjuntos o no. Estos conjuntos están formados en las igualdades que puedan llegar a tener en las variables estudiadas sin partir previamente de una segmentación predefinida, estos como no se tienen una segmentación al estudiarlo se deberán de dar sus principales características para entender o interpretar las razones por las que pudieron llegar a ser conformados en un mismo grupo o conjunto.

7.4. Análisis factorial

El análisis factorial es una técnica que se desprende de la estadística y que permite el poder reducir los datos estudiados y que se utilizan para poder interpretar o explicar las principales correlaciones que hay dentro de todas las

variables sujetas al análisis sin necesidad de utilizar todo el conjunto de variables.

Es decir, se estarán generando nuevas variables llamadas factores que explicarán en su conjunto todas las variables sin necesidad de introducir al modelo toda la data que puede suponer cada una de las variables estudiadas; el análisis factorial tiene sus orígenes en la psicometría, y ha sido ampliamente utilizado en ciencias que estudian el comportamiento como lo es la mercadotecnia, Ciencias Sociales, demanda de productos, investigaciones operativas y otras que para su estudio requieren el análisis de una gran cantidad de datos que sin las tecnologías actuales de computación serían prácticamente imposibles de realizar.

7.4.1. Tipos de análisis factorial

Uno de los tipos de análisis factorial es el exploratorio o AFE, y éste es utilizado para entender y descubrir la estructura de una gran cantidad de variables, bajo el concepto de qué existen factores que se pueden asociar a ciertos grupos de variables. el peso de cada uno de los factores es utilizado para describir las relaciones que guardan las variables entre sí. por su facilidad este análisis factorial es el que más se utiliza.

También existe otro tipo de análisis factorial que es el confirmatorio o AFC, y éste consiste en tratar de entender o determinar si los factores que se obtuvieron del análisis y las cargas de cada uno sobre las variables son los esperados tomando en cuenta teorías previamente definidas. Lo que busca este análisis es confirmar que tengamos factores preestablecidos que cada uno este asociado a un conjunto de variables específicos. es decir, este análisis lo que otorga es un indicador de confianza que permite aceptar o negar la hipótesis

previamente establecido. en este análisis las variables son consideradas dos medidas que constantemente pueden ser cuantificados

7.4.2. Aplicaciones

La aplicación de estos factores principalmente se centra en identificar o explicar características en conjunto que puedan llevar a entender resultados de distintas pruebas.

Un caso de uso ampliamente estudiado es aquel que define que una persona que obtiene buenas notas, en habilidad verbal obtendrá altas notas en la que las asignaturas que requieran un alto desarrollo de la habilidad verbal. esto se puede explicar con el uso de los análisis factoriales que aíslan un factor llamado inteligencia cristalizada, que describe como una persona puede llegar a ser capaz de poder resolver un problema utilizando su habilidad verbal.

En psicología el análisis factorial se utiliza frecuentemente para investigar la inteligencia, la personalidad, las actitudes o creencias. también está asociado a la ciencia de la psicometría, ya que está evalúa la validez que tiene un instrumento al establecer si el instrumento mide correctamente los factores que están postulados.

7.5. Correlación

La correlación es aquella relación que pueda llegar a tener una variable o numérica respecto de otra, y está puede llegar a ser cuantificado utilizando el método de Pearson para medir cuál es su coeficiente de correlación. Este arroja un resultado que puede llegar a identificarse entre la recta empezando desde menos uno hasta más uno.

Siendo -1 una correlación inversamente proporcional y +1 una correlación proporcional a medida que una crece la otra también crece en cierta proporción, cuando se tiene una correlación de cero esta indica que ninguna de las 2 variables guarda similitud en su comportamiento linealmente hablando

También existen relaciones no lineales y es en estos casos que ayuda bastante realizar una representación gráfica y entender visualmente si existiera algún tipo de correlación.

A continuación, se detallan algunas características del coeficiente de correlación:

- El valor que toma el coeficiente de correlación y las unidades que se utilizan en la medición de las variables son totalmente independientes.
- Al analizar el coeficiente de correlación puede llegar a ser alterado de manera importante si hay una alta variación en la medición de las variables tal y como también ocurre cuando se analiza la desviación típica. Es altamente recomendable transformar los datos para que tengan una misma escala de variación y no se vea afectada únicamente por una variable.
- La relación de un coeficiente de correlación es únicamente medida con una línea recta. Es por esto que se recomienda primero representarlas gráficamente, para entender las relaciones que pueda haber entre 2 variables y después proceder a calcular el coeficiente.
- Cuando se extrapolan los datos utilizando un coeficiente de correlación, no se deben de realizar en datos que se dan los límites estudiados entre

X y Y, esto porque la relación arrojada en el coeficiente será únicamente la que hay dentro del rango llegando a variar considerablemente fuera del mismo.

- El hecho de que 2 variables tengan correlación no significa que una sea causalidad de la otra.

La validez del coeficiente de correlación será válida sí, las 2 muestras que se están estudiando fueron tomados aleatoriamente dentro del grupo estudiado, y que al menos 1 grupo de los que se está estudiando tenga una distribución normal.

En la tabla I se observa el peso y la talla de 20 personas de género masculino.

Para calcular la covarianza, se multiplica el peso (kg) y la talla (cm) cuya dimensión se elimina para generar un coeficiente por medio de la división de la desviación típica de la talla y la desviación típica del peso.

El resultado es 0.885 como se vio anteriormente una relación de 1 significaba que las dos variables de estudio guardan una correlación perfecta, por lo que un valor de 0.885 indica una alta relación entre el peso de la persona y la talla de la misma. Como se mencionó anteriormente una alta correlación no significa una alta causalidad.

Al elevar al cuadrado 0.885 (coeficiente de correlación) el resultado que se obtiene el coeficiente de determinación ($r^2=0.783$), el cual está indicando el peso está explicado por la talla en un 78.3 %. Es decir, si bien hay otras variables que pueden llegar a modificar o explicar la variabilidad del peso en un

22.7 % la talla es un excelente indicador para intuir el peso. Sin embargo, para entender el 22.7 % que explica el peso, se deberán de introducir otras variables para explicar al 100 % el peso.

Tabla I. **Cálculo del coeficiente de correlación de Pearson entre las variables talla y peso en 20 niños varones**

Y Peso (Kg)	X Talla (cm)	$X-\bar{X}$	$y-\bar{y}$	$(X-\bar{X})*(y-\bar{y})$
9	72	5.65	1.4	7.91
10	76	9.65	2.4	23.16
6	59	-7.35	-1.6	11.76
8	68	1.65	0.4	0.66
10	60	-6.35	2.4	-15.24
5	58	-8.35	-2.6	21.71
8	70	3.65	0.4	1.46
7	65	-1.35	-0.6	0.81
4	54	-12.35	-3.6	44.46
11	83	16.65	3.4	56.61
7	64	-2.35	-0.6	1.41
7	66	-0.35	-0.6	0.21
6	61	-5.35	-1.6	8.56
8	66	-0.35	0.4	-0.14
5	57	-9.35	-2.6	24.31
11	81	14.65	3.4	49.81
5	59	-7.35	-2.6	19.11
9	71	4.65	1.4	6.51
6	62	-4.35	-1.6	6.96
10	75	8.65	2.4	20.76

Fuente: elaboración propia.

$$\bar{x} = \frac{\sum x}{n} = 66.35$$

$$\bar{y} = \frac{\sum y}{n} = 7.6$$

$$\text{Covarianza} = \frac{\sum(\bar{x} - x) * (\bar{y} - y)}{n - 1} = \frac{290.8}{19} = 15.30$$

$$r = \frac{\text{Covarianza}}{s_x * s_y} = \frac{15.30}{8.087 * 2.137} = 0.885$$

$$s_x = \text{Desviación típica } x = 8.087$$

$$s_y = \text{Desviación típica } y = 2.137$$

7.6. Test de hipótesis de r

Luego de que se tiene la correlación de Pearson, hay que validar si este cálculo es distinto de cero estadísticamente. Se utiliza un test el cual está basado en una distribución explicada por la t de student.

$$\text{Error estandar de } r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Para afirmar que la correlación es significativa, en este caso $r = 0.885$ este debe de ser mayor al valor p (Error estándar de r) cuando se multiplica por la t de student.

Tomando el ejemplo anterior donde se evaluaron 20 personas de género masculino queda que los $n - 2$ es igual a 18, siendo los grados de libertad establecido y el valor de este resultado en la tabla de la t de student cuando se busca una seguridad mayor al 97.5 % esta es de 2.10 y si lo que se busca es una seguridad mayor al 99.5 % el valor resultante en la tabla es de 2.8784 .

$$\text{Error estandar de } r = \sqrt{\frac{1 - 0.885^2}{20 - 2}} = 0.109$$

Al aplicar la fórmula el resultado es 0.109

Al evaluar $2.10 * 0.109 = 0.2289$ según lo que se observa anteriormente, el coeficiente de correlación puede tomarse como significativo puesto que el resultado de 0.2289 es menor al resultado obtenido al calcular la correlación (0.885).

Realizando el mismo análisis, pero ahora buscando una seguridad mayor al 99.5 % en la tabla 2 se observa 2.8784 y reemplazándolo en la fórmula $2.8784 * 0.109 = 0.3137456$ el resultado obtenido sigue siendo menor a la correlación calculada anteriormente por lo que se puede asegurar que el coeficiente continúa siendo significativo para una confianza mayor al 99.5 %. Este proceso puede ser aplicado indiferentemente del tamaño de la muestra al revisar la tabla 2 se tienen los valores cuando la muestra tiende a infinito para una seguridad mayor al 99.5 % el valor será de 2.576.

7.7. Intervalo de confianza del coeficiente de correlación

Para estimar correctamente los intervalos de confianza para el coeficiente de correlación de Pearson, se puede utilizar la transformación Z esto se debe a que cuando se tiene la distribución del coeficiente de correlación de Pearson esta puede no estar centrada, lo que imposibilita el poder estimar correctamente los límites de manera directa. Es por este motivo que se utiliza una transformación de Fisher del coeficiente de correlación (Z de Fisher), esta transformación se asume de distribución normal y con una desviación típica

$$\frac{1}{\sqrt{n-3}}$$

La transformación es:

$$z = 1/2L_n \frac{1+r}{1-r}$$

Ln representa el logaritmo neperiano en la base e

$$\text{El error standar de } z \text{ es } = \frac{1}{\sqrt{n-3}}$$

donde n se refiere a el tamaño muestral.

El intervalo de confianza de z se puede calcular de la siguiente forma para un 95 %:

$$z_1(\text{limite inferior}) = z - 1.96/\sqrt{n-3}$$

$$z_2(\text{limite superior}) = z + 1.96/\sqrt{n-3}$$

Luego de que se calcularon los intervalos de confianza utilizando el valor z, se debe de volver a calcular inversamente para poder calcular correctamente los intervalos de confianza del coeficiente r

$$\frac{e^{2z_1} - 1}{e^{2z_1} + 1} \quad \alpha \quad \frac{e^{2z_2} - 1}{e^{2z_2} + 1}$$

Utilizando el ejemplo de que se está trabajando, se obtiene r = 0.885

$$z = 1/2L_n \frac{1+0.885}{1-0.885} = 1.398$$

Con un 95 % de intervalo de confianza para z

$$z_1 = 1.398 - 1.96 / \sqrt{20 - 3} = 0.922$$

$$z_2 = 1.398 + 1.96 / \sqrt{20 - 3} = 1.873$$

Nuevamente para calcular los intervalos de confianza en z, se realiza un cálculo inverso que darán nuevamente los intervalos del coeficiente r que es el originalmente buscado previo a la transformación logarítmica realizada.

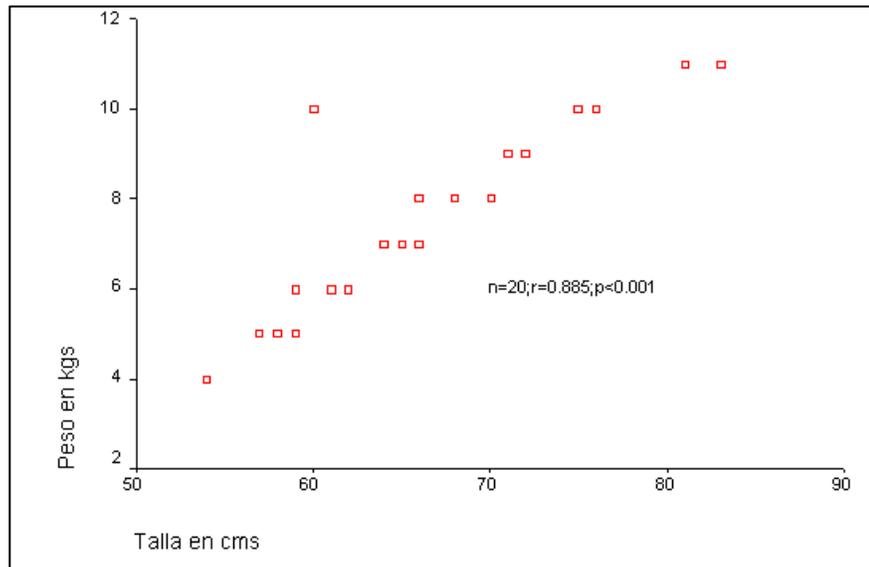
$$\frac{e^{2 \cdot 0.922} - 1}{e^{2 \cdot 0.922} + 1} \quad \alpha \quad \frac{e^{2 \cdot 1.873} - 1}{e^{2 \cdot 1.873} + 1}$$

0.73 a 0.95 serán los intervalos de confianza a 95 % de r.

7.8. Presentación de la correlación

Tomando en cuenta lo anteriormente mencionado es recomendable en todos los casos mostrar gráficamente las 2 variables que se están estudiando y visualizar la correlación que pueda llegar a existir (figura 1). Se muestra el valor de la correlación bajo la letra r a una exactitud de 3 decimales y a su derecha el valor del test de hipótesis demostrando que R es distinto de cero estadísticamente hablando. Por último, se muestran el tamaño de la muestra bajo la letra n.

Figura 1. **Correlación entre talla y peso**



Fuente: elaboración propia, realizado con Rstudio.

7.9. Interpretación de la correlación

Para interpretar correctamente el coeficiente de correlación se debe tomar en cuenta que éste puede ser entre -1 y +1, sin embargo, el valor 0 es el que indica que no existe una correlación lineal entre las variables que se están estudiando.

En el caso de tener un coeficiente cercano a cero no es indicativo de la inexistencia de una correlación, puede ser que las variables tengan una relación que no necesariamente sea lineal es decir mientras una puede variar bajo una recta de tendencia la otra puede llegar a variar exponencialmente, cuando se tiene un caso de esta manera la correlación r subestima la asociación que las variables tienen al estar las midiendo de manera lineal, en estos casos será necesario utilizar métodos que no son paramétricos para poder entender

correctamente si las variables crecen conjuntamente o puedan estar moviéndose en direcciones distintas.

Cuando se utilizan las correlaciones es importante también estudiar la significancia de este coeficiente ya que ayudarán a evaluar si estas son significativas o no, es decir para coeficientes de 0.7 si la muestra es realmente pequeña el intervalo de confianza demostrará que su significancia es alta tomando en cuenta que en estos casos este tiende a ser más amplio.

El coeficiente de determinación que se representa mediante r^2 no es más que un porcentaje que está mostrando cuál es la variabilidad que pueden tener los datos que se están explicando por la asociación de las variables que se están evaluando.

Se debe tener en cuenta que como se ha mencionado anteriormente una correlación cercana a uno, aunque ésta sea significativa no necesariamente se debe a una causalidad, puede que estas 2 variables estén causadas por el mismo fenómeno por lo cual mantengan una alta correlación pero que no sean causantes la una del otro, un ejemplo muy antiguo es el de mencionar la correlación que hay entre las personas que tienen un caballo y su expectativa de vida, al realizar el estudio se muestra que las personas que tienen una alta expectativa de vida, muchas comparten la misma característica de que tienen un caballo sin embargo no es la variable de tener un caballo la que determina su alta expectativa de vida, sino que ésta se debe al poder adquisitivo que pueda llegar a tener una persona por la cual le hace más fácil acceder a tratamientos y cuidados de salud, así como a tener pasatiempos que demandan un alto gasto para obtenerlos es entonces donde se observa que tanto la variable de caballo como la variable de expectativa de vida están

asociadas a una tercera variable que es ingresos económicos por lo cual al subir una sube la otra, pero no necesariamente una causa la otra.

Cuando se intenta medir el mismo evento mediante 2 métodos distintos, no es conveniente utilizar el coeficiente de correlación, ya que en esencia los 2 métodos están enfocados al mismo evento, por ejemplo la medición de temperaturas mediante un termómetro de mercurio versus la medición por parte de un termómetro digital, al medir la correlación que hay entre ambas medidas aunque se refieran a 2 instrumentos que utilizan tecnologías distintas, el evento evaluado es el mismo por lo que pudiera llegar a haber una correlación muy cercana a uno, sin embargo el nivel de concordancia será nulo ya que lo que mide el coeficiente de correlación es la asociación que existe entre 2 cantidades, pero no es capaz de medir el nivel de concordancia.

7.10. Coeficiente de correlación de los rangos de Spearman

Es la medición de la asociación o interdependencia lineal que es utilizada por los rangos. Para medir correctamente esta asociación o para calcular dicho coeficiente existen 2 métodos el primero estudiado por Kendal y el segundo por Spearman, cuando se habla del método de Spearman se denota bajo ρ (rho) y este es más fácil de calcular si se compara contra el método de Kendall. El ρ (rho) es idéntico al coeficiente de correlación de Pearson cuando éste se calcula sobre el rango observado. Cuando se presentan valores externos se recomienda utilizar el método ρ (rho), ya que estos valores pueden llegar afectar mucho el de Pearson, también cuando se tienen distribuciones que no son normales.

La forma de cálculo de dicho coeficiente se muestra bajo la fórmula:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

En donde $d_i = r_{xi} - r_{yi}$ se refiere a la diferencia de los rangos existente en X y Y.

Al colocar los valores de estos rangos deberán de llevar el orden numérico que existen entre los datos de la variable.

Los valores de estos rangos deberán de colocarse según el orden numérico de los datos de la variable.

Caso I: se busca determinar la correlación que existe entre la nicotina presente en la sangre de un grupo de estudio y la cantidad de nicotina que hay que en una conocida marca de cigarrillos.

Los rangos se representan dentro de paréntesis

Tabla II. **Nicotina existente en sangre vs. cigarrillo**

X	Y
Nicotina existente en sangre (nmol/litro)	Nicotina existente en cigarrillo (mg)
185.7 (2)	1.51 (8)
197.3 (5)	0.96 (3)
204.2 (8)	1.21 (6)
199.9 (7)	1.66 (10)
199.1 (6)	1.11 (4)
192.8 (6)	0.84 (2)
207.4 (9)	1.14 (5)
183.0 (1)	1.28 (7)
234.1 (10)	1.53 (9)
196.5 (4)	0.76 (1)

Fuente: elaboración propia.

En el caso de que existan valores que sean coincidentes, se debe colocar el promedio de los rangos que hubieran sido asignados, de no existir coincidencias, tómesese el siguiente párrafo como ejemplo:

Tabla III. **Edad vs. rango**

X (edad)	Rangos (Establecido)
23	1.5
23	1.5
27	3.5
27	3.5
39	5
41	6
45	7
...	...

Fuente: elaboración propia.

Escribiendo según la fórmula el cálculo estimado para la tabla 3 los valores quedarían de la siguiente forma:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} = 1 - \frac{6[(2-8)^2 + (5-3)^2 + (8-6)^2 + \dots + (4-1)^2]}{10(10^2 - 1)} = 1 - \frac{6(120)}{10(99)} = 0.27$$

Como se mencionó anteriormente, al utilizar la fórmula de cálculo para el coeficiente de correlación de Pearson, si obtiene el mismo resultado que con la fórmula anterior, en este caso 0.27 tomando en cuenta los rangos.

$$r_s = \frac{n \sum r_x r_y - \sum r_x \sum r_y}{\sqrt{[n \sum r_x^2 - (\sum r_x)^2][n \sum r_y^2 - (\sum r_y)^2]}}$$

$$\sum r_x = \sum r_y = 55 \quad \sum r_x^2 = \sum r_y^2 = 385$$

$$\sum r_x r_y = 2(8) + 5(3) + 8(6) + \dots + 4(1) = 325$$

$$r_s = \frac{10(325) - 55(55)}{\sqrt{[10(385) - 55^2][10(385) - 55^2]}} = 0.27$$

Al igual que se interpreta el coeficiente de correlación de Pearson el de Spearman r_s también varía entre -1 y +1, exactamente igual que un Pearson un valor cercano a uno significa que existe una correlación fuerte y que ésta a su vez es positiva, es decir si una crece la otra crecerá de igual manera; sí el coeficiente fuera muy cercano a cero y esto nuevamente significa que la variación de una variable no está ligada o correlacionada a la otra variable mientras que al tener coeficientes cercanos a -1 esto querrá decir que las variables variarían de manera significativa una de la otra pero en sentido contrario es decir cuando crece una la otra r_s^2 decrece, también se puede estimar que el r_s tendrá exactamente el mismo significado que con el r^2 .

Por último, las distribuciones existentes entre r y r_s similares por lo que, para realizar el cálculo de los intervalos de confianza en ambos casos, pueden realizarse de la misma manera que como se mostró anteriormente para el de correlación de Pearson.

7.11. Cálculo de correlaciones

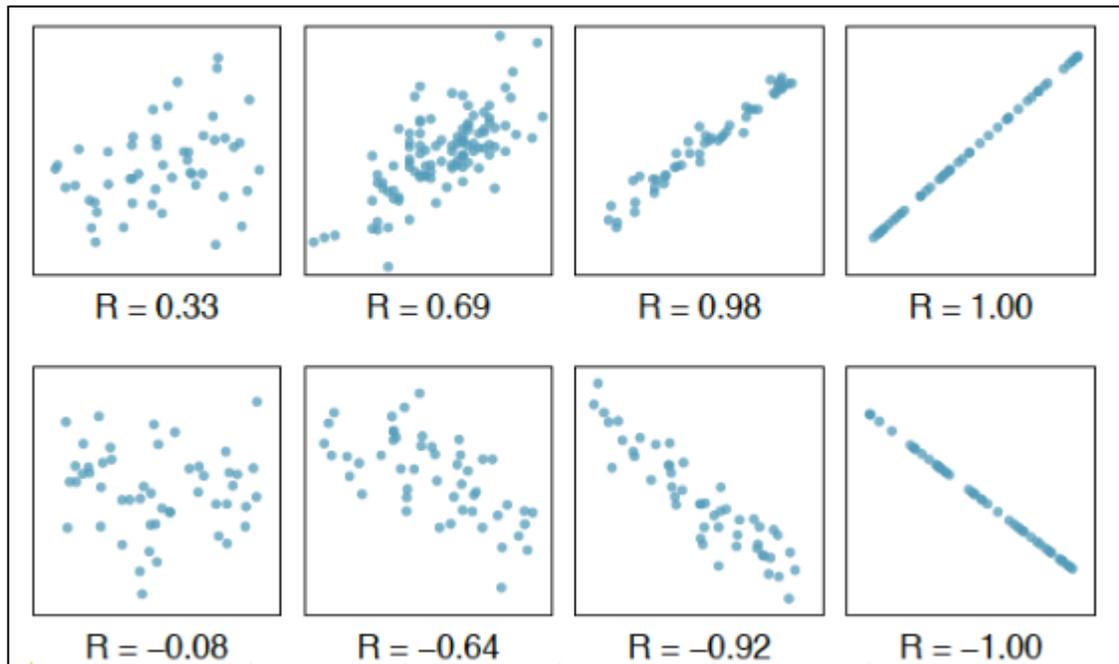
Para calcular las correlaciones como ya se ha documentado es el grado en el cual variarían 2 variables una de la otra, un ejemplo fácil de estudio puede ser la relación que hay entre la variable de salario comparada con la variable de

experiencia laboral siendo la tendencia que al tener mayor experiencia se tendrá un mayor.

Esta se mide con un coeficiente que va de -1 a 1.

- $r = 1$, la relación es positiva perfecta (al crecer una la otra crece).
- $0 < r < 1$ la relación es positiva (al crecer una la otra crece en menor proporción).
- $r = 0$ no hay relación lineal (son totalmente independientes).
- $-1 < r < 0$ la relación es negativa (al crecer una la otra decrece en menor proporción).
- $r = -1$ la relación es negativa perfecta (al crecer una la otra decrece).

Figura 2. **Representación gráfica de matriz de correlación**



Fuente: elaboración propia, realizado con Rstudio.

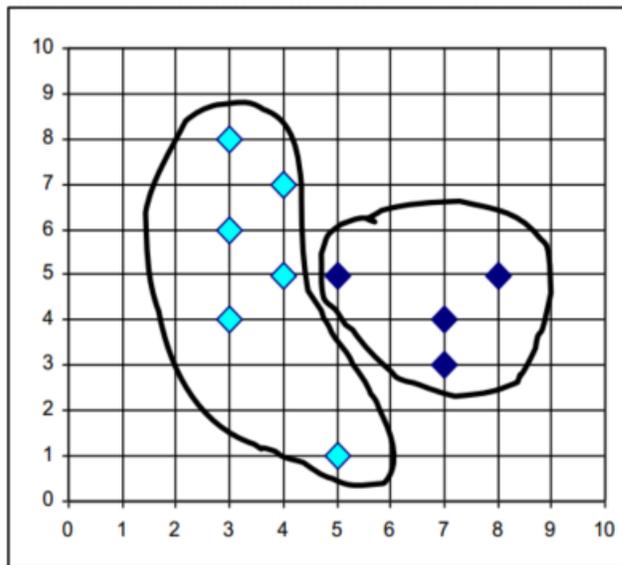
7.12. K-means

Cuando se refiere a este algoritmo es del tipo de clasificación no supervisado o clusterización donde se van agrupando los objetos en una k cantidad de grupos tomando como base las características de las variables estudiadas. Dicha clasificación se logra mediante el proceso de minimizar la suma de las diferencias que existen entre cada objeto y el centroide propuesto, en general se pueden utilizar varias distancias desde la euclidiana hasta la.

Dicho algoritmo se puede realizar en los siguientes pasos los cuales son 3:

Inicio: cuando ya se tiene la cantidad de grupos que se desea, se buscan los centroides esto en un principio pueden ser escogidos aleatoriamente.

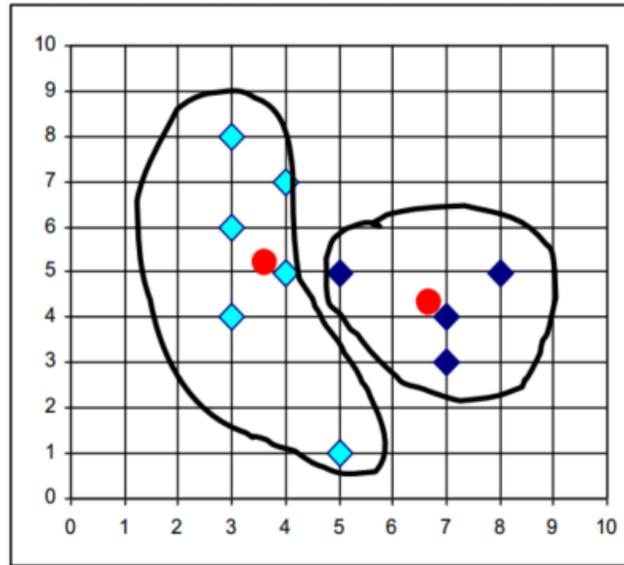
Figura 3. **Segmentos con centroides escogidos aleatoriamente**



Fuente: elaboración propia, realizado con Rstudio.

Cuando ya se tiene el centroide establecido tomando en cuenta la cantidad de segmentos que se desean realizar, el siguiente paso es evaluar todos los puntos y calcular las distancias entre el punto y los centroides y asignarlo al centroide que quede más cercano mediante el cálculo de la distancia establecido para el ejercicio.

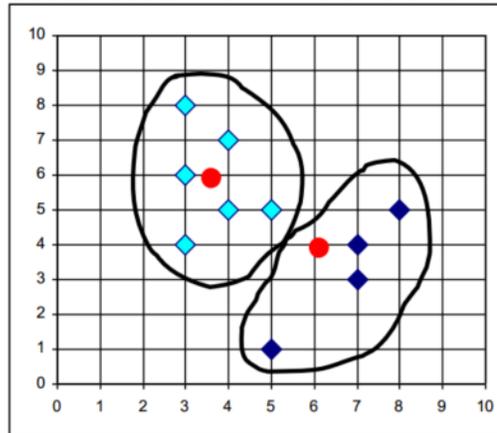
Figura 4. **Cálculo de centroide**



Fuente: elaboración propia, realizado con Rstudio.

Actualización de los centroides: luego de que ya se tienen los 2 segmentos en el caso de este ejemplo, se procederá a calcular el centroide sin tomar en cuenta el centroide que se estableció aleatoriamente al principio del ejercicio y esto nos dará los nuevos centroides.

Figura 5. **Nuevos centroides calculados**



Fuente: elaboración propia, realizado con Rstudio.

Luego de que ya se tienen los nuevos grupos de centroide se procederán a realizar los pasos 2 que es el cálculo y asignación de cada valor a un segmento específico y como paso 3 calcular nuevamente el centroide de estos grupos, al continuar iterando se observará que por cada iteración el centroide empezará a disminuir su variación respecto de la iteración anterior por lo que se podrá detener cuando se logre una variación mínima establecida para cada paso.

Este algoritmo lo que busca es resolver un problema de optimización, que es el total de las distancias sumadas del grupo de objetos en un cluster y el centroide de este.

Definiciones utilizadas en el método de k-means

Inercia total: Es la suma de la inercia de todos los puntos evaluados

$$V = \frac{1}{I} \sum_{s=1}^I \|x^s - \mathbf{g}\|_2$$

Inercia inter-clases: se utiliza para calcular la inercia que existe entre los centros de gravedad (en este caso estudiado 2 centros) y el centro de gravedad que genera el 100 % de los puntos:

$$B(P) = \sum_{k=1}^K \frac{|C_k|}{n} \|\mathbf{g}_k - \mathbf{g}\|^2$$

Inercia intra-clases: esto se refiere a la inercia existente dentro de cada uno de los segmentos establecidos

$$W(P) = \sum_{k=1}^K I(C_k) = \frac{1}{n} \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mathbf{g}_k\|^2$$

Una de las principales ventajas que se tienen con este método es la sencillez que representa y dependiendo de la cantidad de datos este puede ser bastante rápido tomando en cuenta las capacidades de computación existentes en la actualidad, sin embargo una de las desventajas que existen es que se debe de escoger primeramente la cantidad de segmentos que se desean buscar y el resultado final pueda llegar a variar significativamente dependiendo en qué lugar del plano se colocan los primeros centroides ya que como se recordara se tomó de manera aleatoria, es decir el resultado final variará siempre y en algunos casos significativamente dependiendo donde se coloquen los centrales.

7.13. Teorema igualdad de Fisher

I. total = i. inter-clases + i. intra-clases

$$I = B(P) + W(P)$$

Objetivo: el objetivo principal es que la inercia que existe entre cada una de las clases sea mínima, mientras que la inercia que existen entre las clases evaluadas sea máxima, es decir matemáticamente que los clustering sean lo más separado posible uno del otro y que cada clúster esté lo más concentrado posible dentro del clúster mismo.

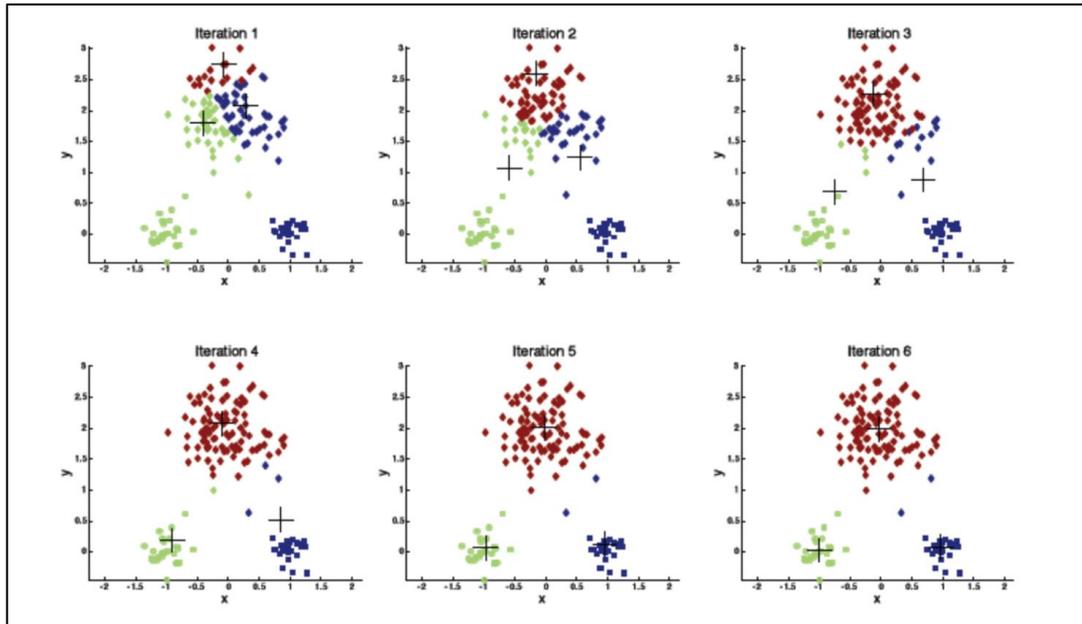
La inercia total no cambia dado que los puntos siempre serán los mismos, por lo que a lograr maximizar la inercia interclases automáticamente se logrará minimizar la inercia intraclase por lo que con enfocarse en minimizar la inercia intraclases se logrará automáticamente maximizar la inercia interclases logrando un grupo más homogéneo dentro de cada clúster.

Se deben tomar en cuenta que si lo que se busca es una segmentación de k número de clases en una nube de n individuos, por definición no se obtendrá ningún beneficio el tener una partición k clases = n individuos.

7.14. Objetivo del método K-means

Como se mencionó el principal objetivo de este método es encontrar una partición que logre que $W(P)$ sea lo más mínima posible

Figura 6. Evolución en la segmentación en cada iteración



Fuente: elaboración propia, realizado con Rstudio.

Tomando en cuenta la figura anterior la principal decisión a tomar es cuántas segmentaciones o clases se realizarán, ya que de esto parte la homogeneidad que se pueda lograr en los grupos que se quieren establecer, en este caso se utilizarán la técnica gráfica a la cual se le conoce como “codo de Jambú”

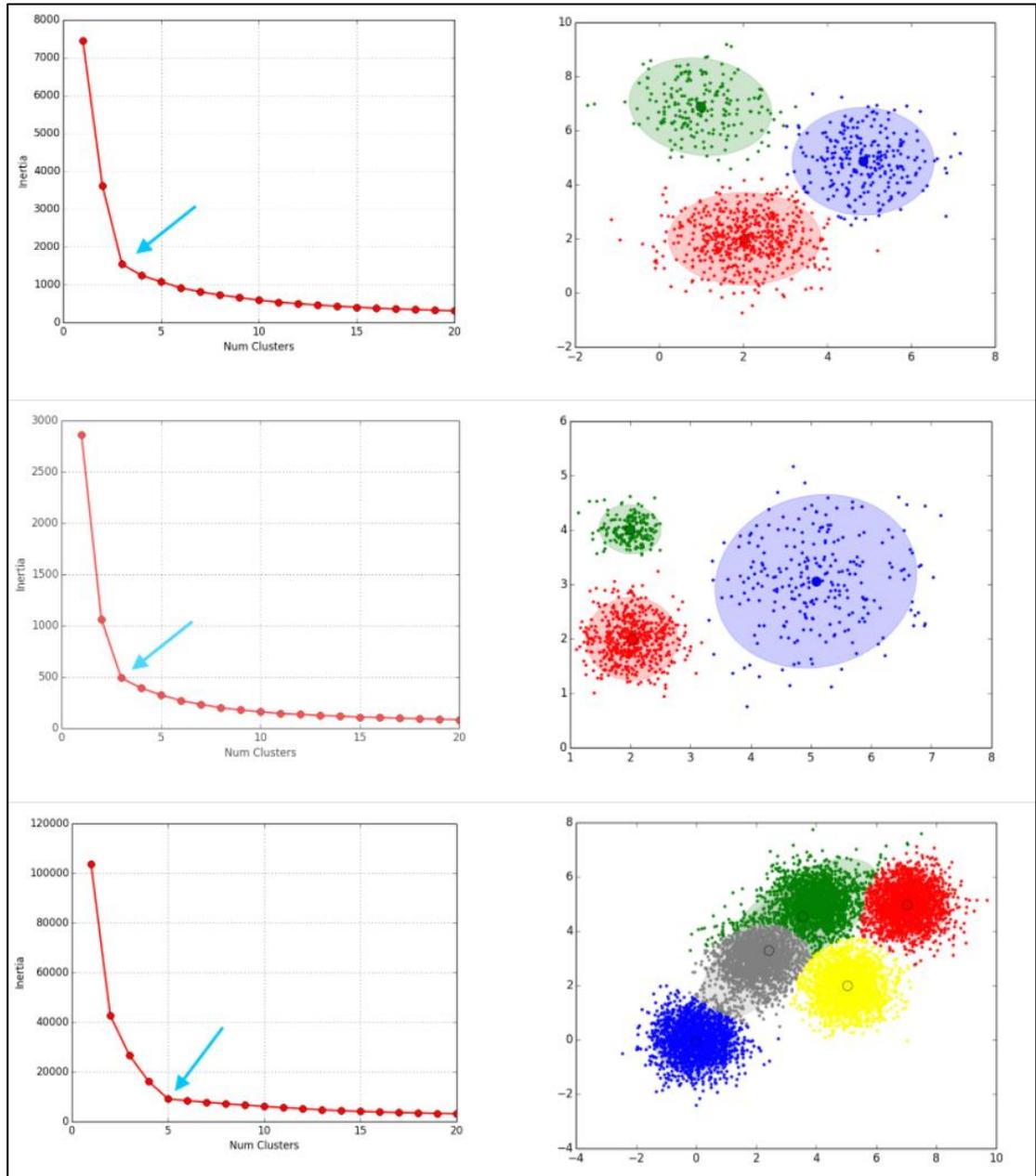
Observando el método se puede notar que todo parte de cuantos clústeres debemos de generar para lograr una cantidad óptima, para esto se puede utilizar una técnica gráfica que se le conoce comúnmente como “codo de Jambú”

7.15. Método del codo de Jambú

Para utilizar el método lo primero que se realiza es el cálculo de la inercia si k fuera igual a uno, luego él se calcula si k fuera igual a 2 y se hace crecer k hasta un número n de clúster eso dará un valor de inercia dependiendo de la cantidad de clúster que se vayan generando al graficar esta inercia en un plano cartesiano siendo x la cantidad de clúster y llegue la inercia total generada se podrá ir observando una gráfica que va disminuyendo exponencialmente a medida que la cantidad de clúster va creciendo, siendo en los últimos valores o cuando la cantidad de clúster tiende al infinito la variación en la inercia será muy bajo por lo que la recomendación es buscar el último valor que hace de crecer considerablemente la inercia y tomar es como la cantidad de clúster que se van a trabajar puesto que es lo que se entiende es que si a partir de esos *clustering* se sigue creciendo se aumentará considerablemente la necesidad de cálculos siendo el beneficio que se obtiene de esa clusterización despreciable.

En la siguiente figura se puede observar visualmente a lo que se refiere este método, del lado izquierdo se tiene el codo mientras del lado derecho se muestra ya la clusterización segmentada por colores lo que permite validar de forma gráfica que dicha cantidad de segmentos tiene coherencia con la inercia total generada por los mismos.

Figura 7. Muestra de segmentos vs. codo de Jambú



Fuente: elaboración propia, realizado con Rstudio.

8. PROPUESTA DE ÍNDICE DE CONTENIDOS

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTEAMIENTO DEL PROBLEMA DE PREGUNTAS ORIENTADORAS

OBJETIVOS

RESUMEN DEL MARCO METODOLÓGICO

INTRODUCCIÓN

1. MARCO TEÓRICO

1.1. Capítulo I: institución bancaria privada

1.1.1. Definición de institución bancaria

1.1.2. Tipos de operaciones y servicios

1.1.2.1. Operaciones y servicios bancarios tradicionales

1.1.2.2. Banca virtual

1.1.3. Clases de bancos

1.1.3.1. Banca especializada

1.1.3.2. Banca universal o multibanca

1.2. Capítulo II: Productos Digitales

1.2.1. Pago digital – pago con el móvil

1.2.2. Tarjetas de créditos digitales

1.2.3. Pagos de servicios a terceros

1.2.4. Transferencias internacionales

1.2.5. Doble factor de autenticación

- 1.3. Capítulo III: Segmentación
 - 1.3.1. Crear modelo de aprendizaje supervisado para ofrecer productos a clientes existentes
 - 1.4. Análisis factorial
 - 1.4.1. Tipos de análisis factorial
 - 1.4.2. Aplicaciones
 - 1.5. Correlación
 - 1.6. Test de hipótesis de r
 - 1.7. Intervalo de confianza del coeficiente de correlación
 - 1.8. Presentación de la correlación
 - 1.9. Interpretación de la correlación
 - 1.10. Coeficiente de correlación de los rangos de Spearman
 - 1.11. Cálculo de correlaciones
 - 1.12. K-means
 - 1.13. Teorema igualdad de Fisher
 - 1.14. Objetivo del método K-means
 - 1.15. Método del codo de Jambú
2. DESARROLLO DE LA INVESTIGACIÓN
3. PRESENTACIÓN DE RESULTADOS
4. DISCUSIÓN DE RESULTADOS
- CONCLUSIONES
- RECOMENDACIONES
- REFERENCIAS
- APÉNDICES

9. METODOLOGÍA

9.1. Enfoque

El estudio propuesto es con enfoque mixto, ya que sé que adquirirá información cuantitativa como por ejemplo la cantidad de clientes que aceptan el producto y también información cualitativa como las causas por las cuales no han aceptado la oferta ofrecida, ya que sean descripciones y sucesos y por lo tanto estos no son cuantificables ni numéricos. El estudio es de carácter transversal al tener una fecha propuesta de inicio y fin.

9.2. Diseño de investigación

El diseño de investigación es no experimental, porque no se intervienen con las variables, será únicamente la recolección de datos del fenómeno a estudiar.

Los datos analizar en esta investigación en particular, tanto como los datos recolectados, serán en un tiempo determinado, es decir que tiene un tiempo de inicio y un tiempo final, sin recurrir a ningún tipo de manipulación y estos datos posteriormente pasarán a evaluarse.

9.3. Tipo

Se ha seleccionado un tipo de estudio descriptivo. Se van a recolectar datos los cuales ayudaran a responder a las preguntas de investigación.

9.4. Alcance

El alcance será descriptivo, ya que se refiere a una propuesta.

Se le proporcionará a la empresa todo el material para poder implementar al sistema de segmentación a clientes y que pueda ser aplicado si así lo consideran necesario.

9.5. Variables e indicadores

A continuación, se presentan las definiciones de las variables de esta investigación.

9.5.1. Variables independientes

Para el diagnóstico de los procesos que se tienen con respecto a la segmentación se evaluarán los ofrecimientos y el estatus del cliente antes y después que realizará el ofrecimiento, hasta 15 después del envío de la comunicación. Estos datos serán de tipo cuantitativos.

9.5.2. Variables dependientes

Para las ventas reconocidas como reales se revisarán los productos a los cuales fue sometido el entrenamiento del modelo. Estas variables se pueden medir a través de indicadores y poder así medir el tiempo de respuesta para que el cliente obtenga el producto.

Tabla IV. **Operativización de variables**

Objetivos	VARIABLES	Tipo de Variable	Indicador	Instrumento
Identificar la forma en que la empresa hace la segmentación de clientes existentes	- Edad - Estado Civil - Salario - Saldo en cuentas - Categoría crédito	Cualitativa y cuantitativas	Porcentaje de ventas efectivas	Observación directa por medio de procesador de datos R Studio
Analizar los factores críticos para la solución de la segmentación de ventas de forma digital de los clientes existentes.	- Productos aceptado - Productos ofrecidos	Cualitativas nominales Y cualitativas ordinales	Cantidad de productos por cliente	Observación directa por medio de procesador de datos R Studio

Fuente: elaboración propia.

9.6. Fases de investigación

A continuación, se presenta la descripción de las fases de la investigación.

9.6.1. Fase I. Revisión

Consiste en la revisión documental de la teoría existente para fundamentar cada uno de los planteamientos a realizar.

9.6.2. Fase II. Diagnóstico

Diagnostico en el año 2020 para la evaluación del proceso de segmentación, cuáles son las ventajas y aprendizajes que se han tenido. Evaluar datos históricos para comparar si hay una mejora en el aprendizaje y correr el modelo para establecer un punto de partida.

9.6.3. Fase III. Determinar variables de importancia

Revisión de resultados de proceso, establecer cuáles de las variables tienen correlación con la aceptación de un producto y cuáles son las mejores para establecer un modelo de segmentación que se adecue a los requerimientos y niveles de aceptación buscados.

9.6.4. Fase IV. Elaborar sistema de medición

En esta fase, se apoyará con los resultados obtenidos de la fase III, para establecer un sistema idóneo que nos permitirá tener un control efectivo en el proceso de segmentación tomando en cuenta los niveles de respuesta en cada uno de los distintos productos a ofrecer.

9.6.5. Fase V. Propuesta de la metodología de segmentación

Revisar los distintos modelos creados con la información recopilada y por medio de la matriz de confusión lograr determinar cuál modelo es el idóneo, la fase que no solo tenga mayor número de aciertos sino también menor número de aciertos en la aceptación del producto. Ya que la variable que se desea de aumentar es la de la aceptación.

9.7. Instrumentos

Los instrumentos utilizados en la investigación se describen a continuación.

9.7.1. Técnicas

La técnica a utilizar en esta investigación serán la comparación sistemática en sistema del cliente, en cada una de las tablas donde se tenga almacenada la información de afiliación o aceptación de los distintos productos.

9.7.2. Instrumentos

El instrumento que se utilizará será en tablas de información llamados *DATASETS* los cuales estarán almacenados en el servidor virtualizado que se contratará para este proyecto, esto debido a que cuenta con capacidad suficiente para poder guardar toda la información y también realizar la comparación del estado del cliente previo al lanzamiento de la comunicación y posterior al cliente.

9.7.3. Población y muestra

A continuación, se presentan los resultados obtenidos en la población entrevistada.

9.7.3.1. Población

La población objetivo de estudio comprende los clientes se tienen al corte del mes anterior al inicio de nuestra investigación de la institución bancaria.

9.8. Resultados esperados

El resultado esperado es un aumento en la efectividad de ventas producto de una mejor segmentación y ofrecimiento al cliente.

10. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN

En esta sección se especificará las técnicas de la estadística descriptiva que se utilizarán en cada una de las fases metodológicas de la investigación que permitirán tratar y analizar la información recolectada de la manera idónea para concluir los resultados que arroje la elaboración del estudio de campo.

El análisis de la información será utilizado durante el proceso para obtener el procedimiento de cambio eficiente. También se recopilará, organizará, presentará, analizará e interpretará datos. Para ello se utilizarán las siguientes herramientas de estadística descriptiva:

- Diagrama de Proceso
- Tablas
- Clasificación de actividades
- Lista de control de actividades a realizar

11. CRONOGRAMA

Figura 8. Cronograma de actividades



Fuente: elaboración propia, realizado con Project 2021.

12. RECURSOS NECESARIOS Y FACTIBILIDAD DEL ESTUDIO

El trabajo de estudio de investigación que se va a realizar es factible, ya que se posee con los recursos necesarios para poder llevar a cabo dicha investigación y se tiene la disponibilidad de los recursos tales como son intelectuales, humanos, financieros y físicos.

Para cumplir con los objetivos planteados se tiene la aprobación de la institución para proporcionarnos los recursos que a continuación describimos:

- Humanos

Se refiere a los colaboradores dentro del área de estudio que será investigado para la recolección de información complementaria como es los cuestionarios y entrevistas que se le impartirán y posteriormente reflejados en los resultados, también se involucra al asesor que es de apoyo para la resolución de conflictos y dudas que se pueden presentar en el camino de la investigación y como por último el investigador directo, ya que es quien se dedicará al 100 % en el estudio en análisis.

Se posee autorización de la subgerencia para el acceso de la información tales como descripción de las normativas, procesos, diagramas, historial estadístico en las múltiples operaciones de la unidad y registros de los colaboradores.

- **Financieros**

La investigación será financiada con el apoyo de la empresa beneficiada al 100 % de los gastos incurridos.

- **Materiales**

Son todos los materiales tangibles que apoyarán y dará validez a la investigación, como por ejemplo impresiones, maquinaria de cómputo, tinta, entre otros, que se utilizarán la cada una de las fases y cumplir con los objetivos.

Los recursos aproximados se dividen de la siguiente manera:

Tabla V. **Recursos**

Recurso	Descripción	Monto
Humano	Tiempo propio de inversión	Q 10,000.00
Humanos	Asistente apoyo base de datos	Q 5,000.00
Financieros	Investigador y asesor	Q 4,500.00
Físicos	Utilería	Q 935.00
Tecnológico	Servidor virtualizado	Q 15,000.00
Otros	Varios e imprevistos	Q1,000.00
TOTAL		Q 36,435.00

Fuente: elaboración propia.

La empresa proporcionará el 100 % de apoyo al investigador en los recursos físicos como son el uso de fotocopiadora, uso de equipo de cómputo y en la aprobación de uso de lapsos de tiempo para la realización del estudio.

13. REFERENCIAS

1. Andrieu, C., de Freitas, N., Doucet, A., y Jordan, M. (s.f.). *An Introduction to MCMC for Machine Learning*. *Machine Learning*, 50(1/2), 5–43. Recuperado de <http://www.cs.ubc.ca/~nando/papers/mlintro.pdf>.
2. Asensi, E., Boronat, L., Tomás, D. y Vicedo, J. (2006). *Desarrollo de un corpus de entrenamiento para sistemas de búsqueda de respuestas basados en aprendizaje automático*. *Procesamiento Del Lenguaje Natural*, 37(37), 59-66. Recuperado de <http://sepln.org/revistasepln/revista/37/08.pdf>.
3. Bernal, D. (2017). *Ajuste de particiones planas mediante diagramas de Voronoi discretos*. [Mensaje de un blog]. Recuperado de <http://oa.upm.es/47124>.
4. Bringas, P., Lopez, I. y Grueiro, I. (2014). *Visión artificial basada en aprendizaje automático para la categorización de defectos superficiales en fundición*. *Dyna*, 89(3), 325-332. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=4677519>.
5. Díaz, C., Morales, E. y Pérez, C. (2009). *Support vector machine model for regression applied to the estimation of the creep rupture stress in ferritic steels*. *Revista Facultad De Ingenieria-universidad De Antioquia*(47), 53-58. Recuperado de http://scielo.org.co/scielo.php?script=sci_arttext&pid=s0120-62302009000100005.

6. Elvira, J., Cívico, F., Cabrera, R., Osuna, M., Cabrera, J. y Olivares, R. (2017). *Actividades extraescolares y rendimiento académico en alumnos de Educación Secundaria. Electronic Journal of Research in Educational Psychology*, 4(8), 35-46. Recuperado de http://investigacion-psicopedagogica.org/revista/articulos/8/espanol/art_8_82.pdf.
7. Figuera, D. (2004). *Métodos cuantitativos para la toma de decisiones*. [Mensaje de un blog]. Recuperado de http://exa.unne.edu.ar/informatica/evalua_ant/metodos_cuantitativo.pdf.
8. Gonzalez, D. (2011). *Algoritmos de clustering paralelos en sistemas de recuperación de información distribuidos*. [Mensaje de un blog]. Recuperado de <https://riunet.upv.es/handle/10251/11234>.
9. Hernández, A. (2019). *Online Clustering con STREAMING K-MEANS usando SPARK Streaming*. [Mensaje de un blog]. Recuperado de <http://oa.upm.es/56397>.
10. Illescas, P., Rizo, D., Iñesta, J. y Ramirez, R. (2011). *Learning melodic analysis rules*. [Mensaje de un blog]. Recuperado de <http://grfia.dlsi.ua.es/repositori/grfia/pubs/276/mml2011-melan-final.pdf>.
11. Ley de Bancos y Grupos Financieros. (2002). *Ley Orgánica del Banco de Guatemala*. Guatemala: Ministerio de Fianzas Públicas.
12. Mecánico, I. y Diaz, A. (2006). *Selección de un servomotor y transmisión por el método de las potencias transitorias*. [Mensaje de un blog].

Recuperado de <https://dialnet.unirioja.es/descarga/articulo/4832259.pdf>.

13. Moreno, F. (2004). *Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano*. (Tesis de doctorado). Universidad de Alicante, España. Recuperado de <https://rua.ua.es/dspace/bitstream/10045/11790/1/Moreno-Seco-Francisco.pdf>
14. Núñez, E., Steyerberg, E. y Núñez, J. (2011). Estrategias para la elaboración de modelos estadísticos de regresión. *Revista Espanola De Cardiologia*, 64(6), 501-507. Recuperado de <https://revespcardiol.org/es-estrategias-elaboracion-modelos-estadisticos-regresion-articulo-s0300893211003502>.
15. Ortega, M. (2008). *Kernelized graph matching and clustering*. [Mensaje de un blog]. Recuperado de <https://dialnet.unirioja.es/servlet/tesis?codigo=21378>.
16. Quevedo, F. (2011). *Medidas de tendencia central y dispersión*. *Medwave*, 11(03). Recuperado de <https://medwave.cl/link.cgi/medwave/series/mbe04/4934>.
17. Roca, J., Bueno, A. y Sancho, J. (2013). *Exploiting Diversity of Neural Network Ensembles based on Extreme Learning Machine*. *Neural Network World*, 23(5), 395-409. Recuperado de <http://nnw.cz/obsahy13.html>.

18. Rubio, A., Fernández, V., Unanue, R., y Herranz, S. (2005). *Evaluación del clustering de páginas web mediante funciones de peso y combinación heurística de criterios*. *Procesamiento Del Lenguaje Natural*, 35(35), 417-424. Recuperado de <http://sepln.org/revistasepln/revista/35/51.pdf>.
19. Sánchez, A., Fernández, F., Valero, C., Muñoz, M., Rodríguez, A. López, y Espejo, I. (2009). *Estadística Descriptiva y Probabilidad: (Teoría y problemas)*. [Mensaje de un blog]. Recuperado de <https://libros.metabiblioteca.org/handle/001/140>.
20. Steve Fortune (junio de 1986). *Numerical stability of algorithms for line arrangements*. *Actas del séptimo simposio anual sobre geometría computacional*, 1(1), 334-341. Recuperado de <https://dl.acm.org/doi/10.1145/109648.109685>
21. Tapis, X. (2001). *El marketing aplicado a las ONGD: coherencias e incoherencias en relación con la educación para el desarrollo*. *Comunicar*, 8(16), 103-114. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/185291.pdf>.

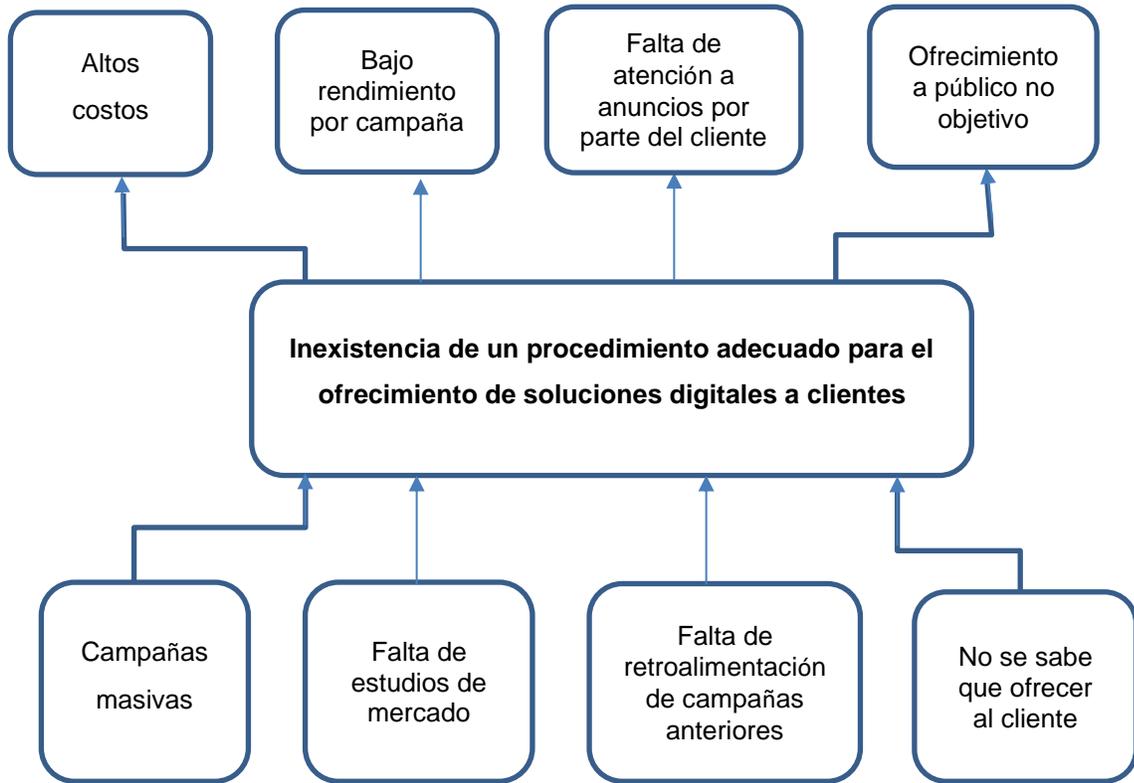
14. APÉNDICES

Apéndice 1. Matriz de coherencia

Formulación del problema	Objetivos	Hipotesis	Variables	Indicadores	Técnicas e instrumentos	Metodología
La inexistencia de un procedimiento adecuado para el ofrecimiento de soluciones digitales a clientes acrecentando los costos de obtención de clientes	<p>General: Proponer un proceso de segmentación para la venta de productos digitales a clientes existentes en una institución bancaria</p> <p>Identificar la forma en que la empresa hace la segmentación de clientes existentes</p> <p><</p> <p>Determinar las ventajas que tiene la utilización de soluciones digitales en la segmentación de ventas en una empresa bancaria para los clientes existentes</p>	La propuesta de un nuevo proceso de segmentación para la venta de productos digitales a clientes existentes aumentara la mejora de los resultados y la rentabilidad del área	<p>Cuales son las mejoras características a tomar en cuenta para evaluar una correcta segmentación</p> <p>Tipos de productos a tomar en cuenta en la evaluación para ofrecimiento</p>	<p>Cantidad de productos aceptados versus cantidad de productos ofrecidos</p> <p>Cantidad de productos vendidos por tipo vs total de productos vendidos.</p>	<p>Los polígonos de Thiessen</p> <p>Diagramas de Voronoi</p>	<p>Metodo de investigación</p> <p>El metodo de investigación es descriptivo</p> <p>Diseño de investigación</p> <p>El diseño correlacional</p>

Fuente: elaboración propia, empleando Excel 2021.

Apéndice 2. **Árbol de problemas**



Fuente: elaboración propia, empleando Word 2021.

Plantilla Hoja de verificación sistemas a evaluar		Logo de la empresa				
Proceso	Verificación variables a utilizar					
Nombre de observador						
Localización	Laboratorio de cómputo					
Fecha						
Sistema / Categoría	Demográficos	Geográficas	Psicográficas	Socioeconómicas	Conductuales	Otra
AS400A						
AS400 B						
Gestor de colas						
CRM						
Gestor de ventas						
Otro						
Total						

Fuente: elaboración propia, empleando Excel 2021.

