



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES**

TEMA:

**Análisis de clusterización basado en el algoritmo K- Means Neighbourhood
para la clasificación de tipos de clientes web en una empresa de Retail E-
Commerce**

AUTORES:

Muñoz Franco Milena Gardenia;

Suárez Aragonés Genaro José

**Trabajo de integración curricular previo a la obtención del título de
LICENCIADO EN NEGOCIOS INTERNACIONALES**

TUTOR:

Ing. Carrera Buri, Félix Miguel Mgs.

Guayaquil, Ecuador

4 de septiembre del 2023



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES

CERTIFICACIÓN

Certificamos que el presente trabajo de integración curricular fue realizado en su totalidad por **Muñoz Franco Milena Gardenia** y **Suárez Aragonés Genaro José**, como requerimiento para la obtención del título de **Licenciado en Negocios Internacionales**.

TUTOR

f. _____
Ing. Carrera Buri, Félix Miguel Mgs.

DIRECTOR DE LA CARRERA

f. _____
Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.

Guayaquil, a los 4 días del mes de septiembre del año 2023



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES

DECLARACIÓN DE RESPONSABILIDAD

Nosotros, **Muñoz Franco Milena Gardenia;**
Suárez Aragonés Genaro José

DECLARAMOS QUE:

El Trabajo de Integración Curricular, **Análisis de clusterización basado en el algoritmo K-Means Neighbourhood para la clasificación de tipos de clientes web en una empresa de Retail E-Commerce**, previo a la obtención del título de **Licenciado en Negocios Internacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Integración Curricular referido.

Guayaquil, a los 4 días del mes de septiembre del año 2023

AUTORES

f.

Muñoz Franco Milena Gardenia

f.

Suárez Aragonés Genaro José



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA

CARRERA NEGOCIOS INTERNACIONALES

AUTORIZACIÓN

Nosotros, **Muñoz Franco Milena Gardenia;**
Suárez Aragonés Genaro José

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Integración Curricular, **Análisis de clusterización basado en el algoritmo K- Means Neighbourhood para la clasificación de tipos de clientes web en una empresa de Retail E-Commerce**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 4 días del mes de septiembre del año 2023

AUTORES

f.

Muñoz Franco Milena Gardenia

f.

Suárez Aragonés Genaro José



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA NEGOCIOS INTERNACIONALES

REPORTE URKUND

REPORTE URKUND 100%

URKUND

Documento	MUÑOZ FRANCO MILENA GARDENIA-SUAREZ ARAGONES GENARO JOSE.docx (D173178417)
Presentado	2023-09-28 01:25 (-05:00)
Presentado por	milenagmf@hotmail.com
Recibido	felix.carrera01.ucsg@análisis.urkund.com
Mensaje	TESIS MUÑOZ SUAREZ FINAL. Mostrar el mensaje completo

0% de estas 103 páginas, se componen de texto presente en 0 fuentes.

Ing. Félix Miguel Carrera Buri

AGRADECIMIENTO

En primer lugar, quiero agradecer a Dios por haberme dado la paciencia y sabiduría a lo largo de esta travesía académica, su dirección ha sido mi fortaleza para no desistir en momentos de adversidades.

Así mismo quiero extender mi gratitud a mi tutor Félix Carrera por su invaluable guía en todo este proceso. Su dedicación y el otorgarnos sus conocimientos fueron indispensables para poder llevar a cabo nuestro trabajo de investigación, superando cualquier dificultad que hayamos tenido en el camino.

Agradezco a mis padres, Gardenia y Edwin, por haber estado junto a mí en cada etapa que he tenido que superar, brindándome siempre sus mejores palabras y soporte cuando tengo momentos de vulnerabilidad. Este logro no habría sido posible sin su apoyo incondicional, orientación y aliento.

Quiero expresar un sincero agradecimiento a Sebastián, que, con su amor y apoyo constante, ha permitido que no me desenfoque de lo realmente importante y por lo que he trabajado todo este tiempo, gracias por tu comprensión en cada paso de esta tesis. Agradezco a Bruno, mi compañero de 4 patas, quien ha sido y será el ser más leal que tendré en todos los años que pueda tener la dicha de pasar con él.

Finalmente, más no menos importante, agradecer a todos mis compañeros quienes me han brindado su apoyo y compañía durante todos estos años, sin ustedes esta experiencia no hubiera sido tan gratificante como lo es ahora.

A cada uno de ustedes, les expreso mi más profunda gratitud. Esta tesis no solo refleja mi dedicación y esfuerzo, sino también el valioso aporte de cada uno hacia mi persona. Su apoyo ha sido esencial en este proceso personal y académico. Les doy las gracias por haber sido parte de este logro.

-Milena Gardenia Muñoz Franco

Este trabajo quiero agradecerles profunda y eternamente a mis padres, Genaro y Patricia piezas claves de mi formación tanto educativa como personal y que a pesar de que pasen los años cada día siguen haciéndolo, dotándome un poco más de su sabiduría y virtud, conozco de primera mano todo el incontable esfuerzo que han demandado para que yo pudiera estar donde estoy y nunca me bastaran las palabras para transmitir la gratitud que siento hacia ellos y lo feliz que me siento de que me hayan tocado como padres.

Quiero agradecer de igual manera a mis abuelos Luis y Ada que con su sabiduría me han guiado a lo largo de todo este arduo camino, cada uno de sus consejos y enseñanzas hoy en día doy gracias de haberlas recibido, a mis tíos Jaime y Silvia que han estado a mi lado prácticamente toda mi vida y de igual manera me han guiado a convertirme en la persona que soy hoy, a mi hermana Romina por brindarme su ayuda en todo este proceso, sacarme sonrisas y motivarme a crecer.

Agradezco a mis amigos por ayudarme a sobrellevar toda esta vida universitaria, la hicieron más amena, estoy feliz de haberlos conocidos a cada uno de ellos. Una mención especial a Belén Marmolejo quien me acompañó durante todo este trayecto y me ayudó a sobrellevar todas las adversidades que se me pusieron en el camino.

Agradezco a Markus Aurelio compañero fiel y leal que se mantuvo conmigo en numerosas noches y madrugadas de trabajo.

Agradezco a mis profesores, cada uno de ellos desde los de primer hasta los de octavo ciclo, todos ellos son los responsables del profesional que estoy por convertirme y acepto la carga de representar sus nombres y de la universidad donde quiera que vaya, una mención especial a mi tutor el Mgs. Félix Carrera que a más de ser un simple profesor ha sido un consejero de vida.

Y, por último, pero no menos importante agradezco a Dios por haberme bendecido con sabiduría y entendimiento para sobrellevar este trabajo, por bendecirme siempre y trazarme el camino que junto a su mano me lleve al éxito.

-Genaro José Suárez Aragonés

DEDICATORIA

A mis padres, quienes han sido mi soporte a lo largo de estos años, cada uno ha sabido formarme para ser la persona que soy ahora, todas las enseñanzas que dejaron en mí han contribuido en mi realización como ser humano, han permitido que pueda avanzar en cada etapa de mi vida con una sonrisa, independientemente de que tan duro haya sido el camino.

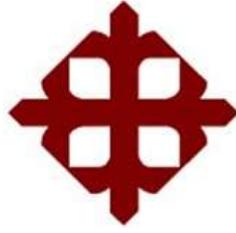
A mi abuela, Elsa, quien desde que era solo una bebé cuidó de mí, me ofreció todo lo que estaba en sus posibilidades e incluso a veces más de lo que podía, es el perfecto ejemplo de nobleza en una persona. Siempre será mi segunda mamá.

-Milena Gardenia Muñoz Franco

No llores porque terminé, sonrío porque sucedió, una frase dicha por uno de mis novelistas favoritos Gabriel García Márquez expresa con exactitud el sentimiento que me mana de mí en estos momentos, parece ayer el primer día de pre donde apoyado al brazo de mi madre buscaba con ansias el aula que iniciaría mi vida universitaria y eh aquí; 4 años después, a un paso de culminar este trayecto de mi vida y continuar hacia un nuevo capítulo, dar marcha atrás nunca fue una opción. Solo aceptar y seguir.

Luego de haber atravesado incontables problemas hasta una pandemia, este logro y todo lo que obtenga posteriormente en la vida se lo dedico principalmente a mi abuelita Ada, madre y mejor amiga que me acompañó a lo largo de mi infancia y adolescencia y que, aunque ahora este en la gracia divina de mi Dios, me sigue guiando y me cuida donde sea que vaya, a mis padres Genaro y Patricia que velan porque jamás me falte nada, y que cualquier cosa por más capricho que fuese lo tuviese, a mi hermana Romina mi motor para mejorar y alcanzar cualquier objetivo. A Belén Marmolejo pilar importante de mi vida que estuvo conmigo en todo momento compartiendo experiencias y formándonos juntos en este largo camino y que quede en constancia mediante este trabajo que con esfuerzo cualquier meta que se proponga se puede lograr, si cree que se puede, ya es el primer paso para lograrlo, les dedico igualmente este trabajo a mis tíos y abuelos parte importante de mí, que me sirven de motivación cuando las cosas se ponen difíciles y el simple hecho de que sé que puedo contar con ellos me apacigua y me hace querer seguir y mejorar.

-Genaro José Suárez Aragonés



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

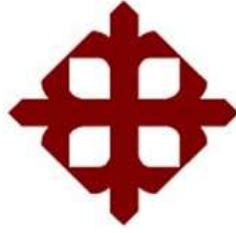
TRIBUNAL DE SUSTENTACIÓN

f.

Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.
DIRECTORA DE CARRERA

f. _____
COORDINADOR DEL ÁREA

f. _____
OPONENTE



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES**

CALIFICACIÓN

f.

Muñoz Franco Milena Gardenia

f.

Suárez Aragonés Genaro José

ÍNDICE GENERAL

Introducción.....	2
Problemática	5
Justificación	7
Alcance	9
Objetivos	10
Objetivo General	10
Objetivos Específicos	10
Capítulo I: Marco Teórico.....	11
1.1 Extract, Transform and Load.....	11
1.2 Inteligencia Artificial (IA).....	11
1.2.1 IA General	11
1.2.2 IA Aplicada.....	12
1.3 Tipos de Machine Learning	12
1.3.1 Aprendizaje Supervisado.	12
1.3.2 Aprendizaje No Supervisado.	13
1.3.3 Aprendizaje por Refuerzo.....	13
1.4 Agrupación en Clústeres Jerárquica.....	13
1.5 Agrupación en Clústeres K-Medias.....	14
1.6 Modelos de Mezcla Gaussiana	14
1.7 Relevancia del Machine Learning	14
1.8 Matemáticas Presentes en Machine Learning.....	14

1.9 Modelos de regresión	15
1.9.1 Calidad de los Datos.	16
1.10 Servidores de Base de Datos Relacionales	16
1.11 Kmeans Neighbor	16
1.12 Como Elegir el Valor de K	17
1.13 Clasificación de los Datos	18
1.14 Árbol de Decisiones Figura 2.	18
1.15 Bosques Aleatorios	19
1.15.1 Algoritmo de Bosque Aleatorio.....	20
1.15.2 Conocer el Funcionamiento del Algoritmo del Random Forest	20
1.16 Marco Conceptual.....	21
1.16.1 E-commerce	21
1.16.2 Clúster	21
1.16.3 Algoritmos	21
1.17 Marco Legal.....	22
Capítulo II: Metodología	23
2.1 Población y Muestra	23
2.2 Métodos y Materiales.....	23
2.3 K-Means.....	23
2.3.1. Set de Datos	23
2.3.2. Estandarización de los Datos	24
2.3.2. Estandarización Generalizada	24
2.4. Random Forest	26

2.4.1. Análisis del Proceso del Random Forest.....	26
2.4.2. Métodos de Ensemble	26
2.5 Construcción del Modelo (Data Train)	27
2.5.1. Out-of-Bag Error	28
2.5.2 Matriz de Confusión	28
2.5.3 Importancia de los Predictores.....	28
2.5.4 Evaluación del Modelo (Data Test)	29
Capítulo III: Análisis de Resultados	30
3.1 Modelo K Means Neighborhood	30
Conclusiones de Resultados.....	41
Conclusiones	43
Recomendaciones	44
Bibliografía	45
Anexos	50

Índice de Figuras

Figura 1. <i>Asignación de objetos a los centroides.</i>	35
Figura 2. <i>Árbol de decisión.</i>	36
Figura 3. <i>Captura de Rstudio vista de los datos.</i>	47
Figura 4. <i>Captura de R studio, variables cuantitativas.</i>	48
Figura 5. <i>Captura de R Studio visualización de graficas de clústeres óptimos.</i>	50
Figura 6. <i>Captura de R Studio Resultado arrojado para numero óptimo clústeres.</i>	51
Figura 7. <i>Cluster plot.</i>	51
Figura 8. <i>Cluster plot.</i>	52
Figura 9. <i>Cluster plot.</i>	53
Figura 10. <i>R Studio K-Means y K-Means clasificación.</i>	54
Figura 11. <i>Ventas totales.</i>	56
Figura 12. <i>Matriz de confusión.</i>	57
Figura 13. <i>Probabilidad por clasificación de cliente.</i>	58

Índice de Tablas

Tabla 1. <i>Problema de la calidad de datos.</i>	34
---	----

Índice de Anexos

Anexo A. <i>Script completo en R-Studio</i>	68
Anexo B. <i>Summary de los resultados.</i>	70
Anexo C. <i>Base de datos con aplicación de K-Means</i>	72
Anexo D. <i>Base de datos con aplicación de Random Forest.</i>	90

RESUMEN

En el mundo moderno y con la digitalización como pilar clave de nuestro día a día, saber e implementar las nuevas tecnologías que se desarrollan es de vital importancia pues son capaces de agilizar o mejorar los procesos que llevamos a cabo en nuestras actividades cotidianas. Sin lugar a duda el apogeo de la inteligencia artificial y sus derivados es la tendencia que hoy en día se desarrolla, y en el campo empresarial no es la excepción siendo infinita sus aplicaciones. Conocer las preferencias de tus clientes sin necesidad de tenerlos presentes representa una gran ventaja competitiva con respecto a la competencia, años atrás esta tarea se vería como algo imposible mas hoy en día con la implementación de nuevas tecnologías paso de ser un sueño a una realidad. En este trabajo se pretende demostrar como la implementación de un sistema de algoritmo de Machine Learning permitiría a las empresas Retail que están presentes en el E-commerce facilitar su clasificación de sus clientes basándose en características similares, esto mediante el modelo de K-Means Neighbour y para su validación mediante el modelo de árbol de decisión, dando consigo así agrupaciones de clústeres similares basados en variables como sería el tipo de pago, el tipo de envío, su uso o node cupones, el valor del envío y su número de transacciones y con ello poder posteriormente brindar descuentos, promociones o campañas dirigidas de mejor manera a los clientes que se les podría hacer más atractivo.

Palabras Claves: E-commerce, Retail, K-Means, Inteligencia de Negocios, Clusterización, Machine Learning.

ABSTRACT

In the modern world and with digitalization as a key pillar of our daily lives, knowing and implementing the new technologies that are being developed is of vital importance because they are able to streamline or improve the processes we carry out in our daily activities. Undoubtedly the heyday of artificial intelligence and its derivatives is the trend that is developing today, and in the business field is no exception being infinite its applications. Knowing the preferences of your customers without having to have them present represents a great competitive advantage over the competition, years ago this task would be seen as something impossible, but today with the implementation of new technologies it has gone from being a dream to a reality. In this work we intend to demonstrate how the implementation of a Machine Learning algorithm system would allow Retail companies that are present in the E-commerce to facilitate the classification of their customers based on similar characteristics, this through the K-means Neighbor model and for validation through the decision tree model, This will result in groupings of similar clusters based on variables such as the type of payment, the type of shipment, the use or not of coupons, the value of the shipment and the number of transactions and thus be able to subsequently provide discounts, promotions or campaigns targeted in a better way to customers that could be more attractive to them.

Keywords: E-commerce, Retail, K-means, Business Intelligence, Clustering, Machine Learning.

RÉSUMÉ

Dans le monde moderne et avec la numérisation comme pilier clé de nos vies quotidiennes, connaître et mettre en œuvre les nouvelles technologies qui sont développées est d'une importance vitale car elles sont capables de rationaliser ou d'améliorer les processus que nous menons dans nos activités quotidiennes. Sans aucun doute, l'apogée de l'intelligence artificielle et de ses dérivés est la tendance qui se développe aujourd'hui, et dans le domaine des affaires n'est pas une exception étant des applications infinies. Connaître les préférences de vos clients sans qu'ils soient présents représente un grand avantage compétitif par rapport à la concurrence. Il y a quelques années, cette tâche aurait été considérée comme quelque chose d'impossible, mais aujourd'hui, avec la mise en œuvre des nouvelles technologies, elle est passée du statut de rêve à celui de réalité. Cet article vise à démontrer comment la mise en œuvre d'un système d'algorithmes d'apprentissage automatique permettrait aux entreprises de vente au détail présentes dans le commerce électronique de faciliter la classification de leurs clients sur la base de caractéristiques similaires, en utilisant le modèle de voisinage K-Means et, pour validation, le modèle d'arbre de décision, Il en résultera des regroupements de groupes similaires basés sur des variables telles que le type de paiement, le type d'envoi, l'utilisation ou non de coupons, la valeur de l'envoi et le nombre de transactions, ce qui permettra par la suite de proposer des remises, des promotions ou des campagnes mieux ciblées sur les clients susceptibles de les rendre plus attrayants.

Mots-clés: E-commerce, commerce de détail, K-Means, Business Intelligence, Clustering, Machine Learning.

Introducción

Desde nuestra evolución nuestro rasgo más distintivo con el resto de las especies habitantes en la Tierra es sin lugar a duda nuestra inteligencia, nuestra capacidad de razonar, entender y tomar decisiones en base a nuestros pensamientos es lo que nos hace superior al resto y nos mantiene en la cúspide de la cadena evolutiva. Hoy en día con el nacimiento de la era digital y el mundo cada vez más conectado entre si se hace más presente que nunca la inteligencia artificial que busca mediante la tecnología dotarla a esta de nuestro conocimiento, inclusive haciéndola mejor que nosotros mismos, la inteligencia artificial es capaz de mejorar o automatizar los procesos que llevamos a cabo día a día ya sea en nuestra vida cotidiana como en el ámbito laboral.

Para Pedrero et al. (2021), el machine learning nace en la década de los cincuenta como medio para emular de manera computacional elementos del proceso cognitivo humano a través de reconocimiento de patrones y procesos de toma de decisiones más hoy en día es un amplio término que agrupa estrategias analíticas cuyo propósito es el desarrollo de algoritmos para extraer información de los datos para su posterior explicación, análisis o predicción. En el mundo empresarial el machine learning es cada vez más implementado y se vuelve cada vez más cotidiano su adopción y enseñanza, tal y como lo dice (Leyva, 2018). A diferencia de lo que pasaba hace 20 o 30 años, en la actualidad, las empresas son capaces de contar con datos más estructurados, ya que se apoyan en sistemas automatizados cada vez más accesibles lo que nos da a entender que manejar el machine learning en las empresas le da una importante ventaja competitiva con respecto a la competencia.

En nuestra era actual donde el comercio electrónico es cada vez más demandado, lograr llegar y abarcar la mayor cantidad de clientes es de suma prioridad para las empresas Retail, es así donde estas empresas se ven inmersas en una lucha constante donde comprender y segmentar a los clientes se vuelve vital importancia para su éxito en el mundo.

Por ejemplo, Thomas et al. (2019), en su artículo ‘Performance Analysis of Clustering Internal Validation Indexes with Asymmetric Clusters’ plantea a la clusterización de datos como una clasificación no supervisada de un conjunto de datos en grupos de manera aleatoria y así descubrir patrones y correlaciones de interés en la distribución de dichos datos, obteniendo así clústeres con datos similares y otros más con datos similares pero diferentes a los del otro clúster anterior. Partiendo de esto el algoritmo K-Means neighbor variante del algoritmo k-Means se presenta como el mejor candidato para ayudar a realizar una clusterización y clasificación de

los clientes para las empresas de Retail a través del comercio electrónico.

En el marco del Retail e-commerce, la clasificación de clientes desempeña un papel crucial en el desarrollo de estrategias de marketing y personalización de la experiencia del cliente. Al comprender las preferencias, comportamientos y necesidades de los clientes, las empresas de Retail e-commerce pueden ofrecer ofertas personalizadas, mejorar la retención de clientes y maximizar el valor de vida del cliente llegando a la fidelización de este. La clusterización basada en el algoritmo k-Means neighbor proporciona una forma eficiente de clasificar a los clientes en diferentes clústeres con características comunes, lo que facilita la toma de decisiones.

En García et al. (2022), se logra observar que, la aplicación de inteligencia artificial en el marketing se da en el desarrollo de experiencias digitales personalizadas y emocionalmente inteligentes basadas en su comportamiento, preferencias y emociones mismas de las personas, y bien ahora si juntamos la IA con el Machine learning este último nos puede brindar algunos beneficios en su uso como la ayuda en problemas de regresión y clasificación, visualización, agrupación, reducción de dimensionalidad, creación de reglas de asociación, agentes de aprendizaje reforzado entre muchas otras:

Como podemos ver el Machine learning está presente cada vez más en nuestras vidas cotidianas, hoy en día se le da numerosas aplicaciones a este, entre ellas podemos encontrar a los traductores inteligentes, muy aparte de los traductores convencionales que sin más traducen la palabra tal cual, hoy en día Google trabaja en aprendizaje automático para que los traductores cuando traduzcan varias palabras es decir un texto, esta traducción vengana orientada en el contexto del párrafo para una traducción más precisa, esto bajo el constante aprendizaje de patrones entre palabras similares, y no solo eso sino que el machine learning permite esta traducción tanto por medio de reconocimiento de texto como de voz, otro ejemplo claro que podemos constatar es el reconocimiento facial como medida de seguridad.

Al empezar los celulares con pantallas táctiles un avance que se le busco integrar era el desbloqueo con huella dactilar, tanto así que este llego a ser usado para desbloquear otras aplicaciones, de ahí partió el reconocimiento facial analizando la biométrica de cada usuario, llevando hasta tal punto de aprendizaje que hoy en día con el inicio de la pandemia del covid-19 podemos desbloquear nuestro dispositivos con la mascarilla puesta, el aprendizaje ha llegado

a tal punto que en la actualidad utilizamos este tipo de desbloqueo para poder iniciar en nuestras cuentas bancarias o inclusive para poder validar una compra que efectuemos.

Otro ejemplo claro presente en nuestra sociedad son las ordenes por comandos de voz, el GPS o Alexa aprende de lo que le decimos y lo interpreta para poder ayudarnos, como por ejemplo si le decimos al GPS que queremos ir a casa de Cristina, el GPS aunque no tenga una dirección en los mapas como tal asociara viajes pasados y trazara la ruta más cercana hasta la casa de Cristina sin necesidad de haber puesto la ruta o dirección exacta de lugar a donde queremos ir, los asistentes de sistemas operativos son otra novedad que integra machine learning, en estos podemos encontrar los asistentes de marcas de nombres como Siri en Apple, Bixby en Samsung y Cortana en Windows, su principal función es que se adapten a cada usuario brindándoles una experiencia única lo más parecido a una interacción con otro ser humano.

La inteligencia artificial al igual que en su tiempo fue el internet provee ser de una gran ayuda para el ser humano, tal y como lo dice Sosa (2023) según un informe de PwC la inteligencia artificial planea para el año 2030 contribuir al PIB mundial con \$15, billones, siendo el 45% de estas ganancias provenientes del incremento de la productividad y el 55% de los consumos de la urbe.

En el ámbito global Abbas y Farooque (2019) muestran como en Asia, China está apostando fuertemente por la inteligencia artificial, queriendo inclusive posicionarse como el mayor referente en este sector, con la continua implementación de esta se espera que su PIB se incremente de 0,8%-1,4% en los próximos años.

Acorde a lo anterior Lee (2018), plantea que China y Estados Unidos actualmente está incubando a los gigantes de la IA que serán los que en un futuro llevaran consigo el mercado mundial y se adueñaran de la riqueza de los consumidores de todo el mundo, a su vez también expone que la automatización de la inteligencia artificial prevé ser un problema para las economías en vías de desarrollo pues eliminaran la única ventaja que disponían, su mano de obra a bajo costo, lo que se espera que pase es que las empresas vuelvan a su país de origen y se instalen fabricas robotizadas eliminando así tiempos de entregas y gastos ya innecesarios de transporte logística y seguridad.

En Europa a pesar de estar por debajo en el desarrollo de la inteligencia artificial, si se ha enfocado mucho en dotarla a esta de un marco normativo y específico acorde a los estándares europeos estableciéndole límites o requisitos específicos (Casado, 2021)

Pasando a un marco local, Ecuador es conocido como un país biodiverso con una flora y fauna increíble más no un país que se destaque por su infraestructura tecnológica, pese a esto la inteligencia artificial se está adaptando al modelo de vida Ecuatoriano desempeñando un papel importante en temas de seguridad y salud esto principalmente dado por la pandemia del Covid-19 donde el país y el mundo entero se vio inmerso en digitalización masiva, donde pasamos de ir a una oficina para trabajar o ir a un centro comercial a realizar nuestras compras realizar estas actividades desde nuestra casas con la ayuda de una computadora o el celular, este cambio rotundo al principio no aceptado por algunos golpeo fuertemente pero a la larga trajo consigo numerosos beneficios pues procesos que se realizaban normalmente dentro de la empresa se halló que no era necesario que el trabajador estuviese ahí directamente para realizarlos permitiéndole ahorrar a la empresa tiempo y recursos.

En retrospectiva la inteligencia artificial junto al Machine Learning se posicionan como buenos candidatos para ayudar a clasificar los tipos de clientes en una empresa de Retail y si bien existen desafíos asociados a su implementación sus beneficios superan con creces a las dificultades, a lo largo de este trabajo se desarrollara más en profundo su aplicación y los resultados que se pueden obtener mediante el uso de esta en el contexto de las empresas de Retail, más específico en el comercio electrónico.

Problemática

Los E-commerce a lo largo de los años han ganado una fuerte popularidad en todo el mundo, vivimos en un mundo globalizado con constantes cambios y cada vez más empresas se suman a esta modalidad para poder seguir siendo competitivas en el mercado. Tal y como lo dice Orus (2023), el E-commerce en los últimos años y a raíz de la pandemia se ha convertido en una parte esencial de los comercios minoristas o Retail siendo así la única forma en que algunas empresas pudieron seguir generando ingresos cuando se activó la medida del confinamiento y así se evidencia un constante crecimiento el cual no ha frenado siendo aun mayor para los años 2021-2022.

Partiendo de esto, varias empresas se sumaron a esta modalidad y es aquí donde a pesar de contar con la tecnología y los datos o información necesaria que permitan realizar

una segmentación de mercado por medio de un modelo estadístico para poder atraer más clientes y generar mayor conversión en sus páginas, Ecuador y numerosos países del globo tienen el constante problema de que no contar con un personal debidamente capacitado que pueda hacer un correcto uso y análisis de la información.

Es evidente notar la necesidad de las empresas de contar con profesionales especializados que otorguen una información pertinente al área gerencial para que posterior a la revisión ellos puedan tomar las mejores decisiones en situaciones con mayor incertidumbre. Además, Villegas (2019) notó que, a pesar de tener una gran cantidad de datos, por más bueno que parezca, hay empresas que suelen caer en un laberinto sin salida al no saber cómo clasificar y analizar toda esta información que el sistema les brinda, esto hace que al final del camino los decisores no puedan llegar a un punto en específico en el que puedan concluir con una decisión para el problema que se les presenta.

La falta de conocimiento en el área estadística puede generar numerosas consecuencias negativas a la empresa; ya sea por no aprovechar de manera óptima los resultados o por un mal análisis de datos que conlleve a la toma de una mala decisión. En gran medida, es responsabilidad de cada organización contar con el personal adecuado para los análisis correspondientes, ya sea por medio de una buena selección de personal o por medio de capacitaciones al personal existente.

Al momento de clasificar los clientes para tener una mejor jerarquización de consumistas es importante disponer de personal apto para desarrollar estrategias y tácticas para captar más clientes y retener a los ya existentes, no obstante como se tomaran medidas sino hay reportes previos de los que guiarse, si bien los datos existen es necesario contar con candidatos cualificados para poder realizar un análisis completo de los resultados y es su deber transmitirlos hacia la empresa de una manera clara y concisa dado que desde su interpretación de los resultados se ejecutará una medida en la organización.

El Machine Learning rama de la inteligencia artificial se presenta como el mejor candidato para ayudar a las empresas en este reto, pero una vez la falta de conocimiento por parte del trabajador hacia estos temas es un inconveniente latente en empresas ecuatorianas y ya decisión de las compañías analizar si es rentable o no invertir recursos y tiempo en desarrollar estas habilidades.

Justificación

Las técnicas de minería de datos y aprendizaje automático han generado un interés notable y han ganado una creciente demanda en los últimos años, ya que entidades gubernamentales, empresas e investigadores buscan un aprovechamiento en su versatilidad y los resultados diversos que ofrecen estas herramientas ya que su uso no solo se limita a la resolución de problemas en el ámbito científico, sino que también proporciona una solución efectiva para abordar los desafíos específicos de empresas de diversos sectores, en particular las del sector servicios.

En general, la aplicación de minería de datos proporciona a toda estructura empresarial información y conocimientos en base a datos que respaldan la toma de decisiones empresariales para una mejor toma de decisiones en áreas diversas que abarcan tanto la planificación de productos y sus campañas de marketing, la expansión de la empresa, la adquisición de clientes y su fidelización en base a sus preferencias personales.

Mediante este trabajo, se busca lograr un procesamiento de datos más eficiente para una empresa de e-commerce, lo que ayuda a reducir el desorden general al permitir identificar y atender las necesidades de los diferentes grupos de clientes de manera más precisa. Así, se promueve una selección de avenidas y estrategias más adecuadas, basadas en prácticas o recorridos únicos, que sean más viables y beneficiosas para el negocio en su conjunto (Carchi, 2019).

En el contexto del comercio electrónico, comprender y clasificar a los clientes es fundamental para el éxito de una empresa, dado que, según Rodríguez (2022) esta segmentación de los clientes permite una mejor comprensión de sus características, preferencias y comportamientos de compra, a la vez que facilita la personalización de la experiencia del cliente, la oferta de productos relevantes y la implementación de estrategias de marketing más efectivas.

La clusterización basada en el algoritmo k-Means neighbor es de utilidad en estos procesos de clasificación al agrupar un conjunto de datos en clústeres, y en el contexto de una empresa, es crucial identificar y comprender cómo los productos o servicios llegarán a los diferentes tipos de clientes, lo que brinda una segmentación efectiva de los clientes en grupos homogéneos con características similares para facilitar la personalización de las ofertas, la implementación de estrategias de marketing dirigidas y la optimización de la

experiencia de compra en línea.

Como ventaja del empleo del algoritmo k-Means neighbor para la clusterización, Correa (2021) menciona que, este es una variante del algoritmo k-Means tradicional que mejora la precisión y robustez de la clusterización a través del concepto de vecinos más cercanos para asignar de manera más efectiva los puntos de datos a los centroides de los clústeres, lo que lo vuelve una opción adecuada para el análisis de datos para la clasificación de clientes.

Los beneficios que pueden lograrse a partir del empleo de esta técnica, además de lograr una identificación de segmentos de clientes con características similares. Según Puga (2020), Esto facilita la segmentación de la base de clientes en diferentes clústeres o segmentos, también se relacionan a la personalización de la experiencia del cliente al brindar una experiencia de compra más personalizada y también da oportunidad a adaptar las estrategias de marketing de manera más efectiva, a través de recomendaciones de productos relevantes, promociones específicas para cada segmento, mensajes personalizados y ofertas exclusivas, lo que aumenta la satisfacción del cliente y mejora la retención.

Una mejor comprensión de las preferencias de los clientes permite también la optimización de estrategias de marketing, ajustando los esfuerzos de publicidad y promoción para llegar a grupos específicos de clientes, así como el diseño de campañas de marketing más específicas para cada segmento y su respectiva asignación de recursos, tanto económicos como creativos mientras permite a los gerentes de la empresa una mejora de la gestión de inventario y la logística de la cadena de suministro, para poder predecir la demanda de productos, identificar patrones estacionales, optimizar el almacenamiento y la distribución de productos, reducir los costos de envío y mejorar la eficiencia en general.

Ya que las técnicas de análisis de datos son incipientes en el contexto ecuatoriano, una desventaja de aplicarlas puede consistir en la obtención de datos fiables, ya que en el contexto nacional recién empiezan a implementarse estas estructuras, ya que mientras más datos se procesen se garantizarán resultados precisos, también de importancia tal asegurarse de que los datos sean tratados de manera ética y legal, respetando la privacidad y los derechos de los clientes.

A pesar de ello, las ventajas de aplicar estas técnicas de procesamiento de información son más ventajosas ya que no solo ayuda al aspecto económico de la empresa, sino también, al considerar factores como la demanda, la competencia y las preferencias de los clientes, ayuda a identificar qué promociones y descuentos son más atractivos, maximizando el impacto de las ofertas promocionales, e incluso puede detectar comportamientos sospechosos que indiquen posibles fraudes o actividades fraudulentas en las plataformas de e-commerce, lo cual ayuda a proteger tanto a la empresa como a los clientes.

Mediante este trabajo, se busca contribuir también a la literatura existente, porque, aunque la clusterización y el algoritmo K-Means Neighbor han sido ampliamente estudiados, la aplicación específica de este algoritmo en el contexto de una empresa ecuatoriana de Retail e-commerce sentará precedentes para este tipo de actividades, brindando herramientas y respuestas a los planteamientos que surgen a partir de la aplicación de herramientas tecnológicas nuevas, proporcionando información valiosa para empresas del sector.

Alcance

La presente investigación está dirigida hacia las organizaciones locales e internacionales, con la finalidad de otorgar más información acerca del uso y los beneficios de algoritmos estadísticos para poder analizar los resultados del tratamiento de datos y variables de clientes en una empresa para poder llegar a una mejor toma de decisiones gerenciales.

Así mismo, para dar a conocer la importancia de entender el comportamiento, preferencias y las necesidades de los clientes a través de una óptima segmentación de estos, lo cual se puede desarrollar con la existencia de profesionales capacitados en el análisis de datos y en la implementación de soluciones con inteligencia de negocios.

Adicional, la investigación busca ser de interés para las personas que se interesen en el campo del análisis de datos y que sirva como apoyo y guía en el proceso de conocer las herramientas que le ofrece el business intelligence como instrumento para implementar estrategias de mejora en distintas áreas de una organización. Finalmente, también se dirige a estudiantes y profesores como un recurso académico que les permita reforzar el

conocimiento existente en campos como el data mining y el aprendizaje automático, así como también en futuras investigaciones académicas y profesionales.

Objetivos

Objetivo General

- Analizar e implementar una óptima clasificación de tipos de clientes web mediante la aplicación del concepto de Machine Learning a través del algoritmo K-Means Neighbor

Objetivos Específicos

- Analizar los conceptos bajo una amplia revisión literaria y su metodología para la clasificación mediante el aprendizaje supervisado y no supervisado
- Evaluar los modelos estadísticos y el rendimiento hacia la clasificación de tipos de clientes web.
- Optimizar la segmentación de los clientes web con el objetivo de tomar mejores estrategias de marketing para cada agrupación

Capítulo I: Marco Teórico

En la actual era digital, la gestión y análisis de datos son esenciales para decisiones efectivas en negocios. En el siguiente marco teórico, se exploran conceptos y metodologías que están directamente relacionados a la estadística como ciencia, analizando cómo convierte datos en conocimientos útiles y su impacto en la eficiencia operativa y ventaja competitiva empresarial.

1.1 Extract, Transform and Load

El proceso de transformación de datos a partir de un número ilimitado de fuentes, su organización y centralización en un repositorio único. Se trata de transformar la información para adaptarla a las necesidades de la empresa y luego cargarla en un sitio compartido para su consulta por las partes interesadas. Muchas empresas contienen información en plataformas como ERP, CRM u hojas de cálculo y el proceso ETL lo que logra es transformar esa información repartida en datos útiles para la limpieza, la transformación y la obtención de conocimiento que derive en acciones concretas. La premisa es canalizar los datos de distintos orígenes y transformarlos según las reglas de negocio cargándose en un almacén de datos (Prometeus Global Solutions, 2021)

El ETL es un proceso mediante el cual una herramienta de software permite capturar la información que genera las bases de datos propias de negocio, CRM, o de ERP de la compañía, y así luego transformarlas a nuestras necesidades y finalmente cargándola en la base analítica desde donde se explotará y se realizará su posterior análisis (Baena, 2016)

1.2 Inteligencia Artificial (IA)

Es la simulación de procesos de inteligencia humana por parte de máquinas, especialmente sistemas informáticos. Estos procesos incluyen el aprendizaje, razonamiento y la autocorrección (Rouse, 2021).

1.2.1 IA General

Realiza tareas como hablar, interpretar, reconocer palabras y objetos, y participar en negocios y asuntos comunitarios.

1.2.2 IA Aplicada

Se refiere a tecnologías sensoriales como los vehículos autónomos, que utilizan algoritmos para comprender patrones y diseños. Actualmente, las aplicaciones del algoritmo se han desarrollado tanto que podemos ejecutarlo desde teléfonos inteligentes y computadoras portátiles.

El Machine Learning tiene varias disciplinas arraigadas como lo es la minería de datos que se nutre de diversas áreas, como bases de datos, repositorios de datos, estadísticas, aprendizaje automático, visualización de datos, búsqueda y recuperación de información, y computación de alto rendimiento. Estos campos se combinan para desarrollar procesos esenciales que aplican métodos inteligentes con el fin de extraer y descubrir patrones en los datos. Se utilizan algoritmos como regresión lineal, regresión logística, asociación, lógica difusa, árboles de decisión y redes neuronales, entre otros. Estos métodos permiten obtener conocimientos tanto históricos como prospectivos, fundamentales para la toma de decisiones en un área de estudio o conocimiento determinado (Ferley & Gomez, 2014).

El aprendizaje automático incluye una rama de la informática, que está relacionada con el desarrollo de la inteligencia artificial y sirve para crear sistemas capaces de autoaprendizaje. Es un método de análisis de datos que automatiza la construcción de modelos analíticos mediante el uso de datos que abarcan diversas formas de información digital, incluidos números, palabras, clics e imágenes (Mendoza, 2020).

1.3 Tipos de Machine Learning

1.3.1 Aprendizaje Supervisado.

Los datos se etiquetan para indicar a la máquina qué patrones tiene que buscar. El sistema se entrena sobre un conjunto de datos etiquetados, con la información que se supone que tiene que determinar. Es posible que los datos ya estén clasificados de la forma en que se supone que debe hacerlo el sistema (Scientist, 2021).

De los modelos más conocidos de aprendizaje supervisado tenemos:

- Regresión lineal y logística.
- Máquina de vectores de soporte.
- Árboles de decisiones.
- Redes neuronales.

1.3.2 Aprendizaje No Supervisado.

En este tipo de aprendizaje no se utilizan valores ni etiquetas reales. Estos sistemas están destinados a esquemas de comprensión directa e información abstracta. Este es un modelo de problema llamado agrupamiento (clustering). Es un método de entrenamiento similar a cómo los humanos procesan la información (APD, 2019).

De los más conocidos no supervisados tenemos:

- K-Means
- Clusterización jerárquica

1.3.3 Aprendizaje por Refuerzo.

Su principal ventaja es que puede funcionar sin una gran cantidad de datos de entrenamiento. Solo necesitas una serie de indicaciones para aprender por ensayo y error. En contraste con el aprendizaje supervisado que se basa en un conjunto de datos que le dice a la máquina qué hacer, las recompensas se utilizan aquí para reforzar el comportamiento deseado (García, 2020).

A la hora de hablar del lenguaje no supervisado centrado más en nuestro desarrollo del tema podemos encontrar que este utiliza un sin número de algoritmos para poder ajustar los grupos en clústeres y asociaciones basándose en numerosos aspectos de similitud entre ellos.

1.4 Agrupación en Clústeres Jerárquica

Esto ocurre cuando la máquina agrupa las cosas que van juntas en un árbol de clústeres. Todos los datos son un grupo, luego se dividen en grupos cada vez más pequeños. Los datos pertenecerán a un conjunto en cascada de clústeres, desde los más genéricos

hasta los más específicos y estrechamente agrupados. Entonces, el resultado final será que usted verá cómo los diferentes subgrupos se relacionan entre sí o qué tan separados están.

1.5 Agrupación en Clústeres K-Medias

Este algoritmo separa los datos en distintos clústeres que no se etiquetaron en los datos. Según la precisión de la asociación los datos están más o menos cerca del centro. Los puntos de datos pueden pertenecer a un solo clúster. Una k más amplia significa que el grupo es más pequeño y con más granularidad. A cada clúster se le asigna una etiqueta de punto de datos.

1.6 Modelos de Mezcla Gaussiana

Sobre la base de una distribución de curva de campana normal, los clústeres de grupos se distribuyen a lo largo de las densidades normales previstas, mostrando subpoblaciones en los datos generales.

1.7 Relevancia del Machine Learning

El Machine Learning hoy en día está tomando suma relevancia en áreas empresariales de estructuración y automatización, esta rama de la inteligencia artificial se nutre de áreas de las matemáticas como pueden ser la estadística, las probabilidades, investigación de datos y recopilación de información para obtener patrones matemáticos, no obstante, también esta hace uso de áreas de ingeniería como es la física todo esto para poder desarrollar un buen modelo.

Tal y como lo dice Leyva (2018), entre los sectores que más se benefician se encuentran los sectores de la salud, público y educativo, es útil a su vez en el sector agrícola, automotriz, camaronero, etc. Indispensable en el mundo de las finanzas y de la contabilidad entre numerosas otras aplicaciones que se le puede dar.

1.8 Matemáticas Presentes en Machine Learning

Tanto supervisado como no supervisado, las matemáticas están presentes en cada uno de los modelos que engloba el Machine Learning, entre las áreas de matemáticas podemos encontrar al álgebra lineal usado en modelos de regresión lineal, regresión

logística, máquinas de soporte vectorial, árbol de decisión y Random Forest.

El cálculo es otro aspecto importante en el Machine Learning pues nos da acceso a temas relevantes como el descenso de gradiente que se usa para disminuir el error de los modelos no supervisados tal y como lo dice. El descenso de gradiente es un algoritmo que predice numéricamente donde una función genera sus valores mínimos.

No olvidar por otro lado las teorías de probabilidades para modelar la incertidumbre en los modelos de Machine Learning, agregado a esto la estadística que es la que se encarga de brindar los resultados y posteriormente su interpretación.

La estadística es la que se encarga de recolectar, ordenar, representar, analizar e interpretar los datos que se generen en una investigación de hechos, ya sea de un individuo o grupo y así lograr determinar una solución precisa o una estimación futura (Salazar & Del Castillo, 2018).

La estadística se desarrolla en dos áreas, la descriptiva y la inferencial. En la estadística descriptiva abarcan técnicas que permiten realizar análisis elementales de observaciones y un análisis de datos obtenidos y así poder dar una conclusión sobre el comportamiento de las variables estudiadas (Navarro, 2018). Cabe decir que al momento de hacer uso de estas técnicas se deben saber qué tipo de datos se cuentan ya sean cuantitativos como cualitativos, sin olvidar los gráficos que más se adecuen a su necesidad para su posterior interpretación.

Por otro lado, tenemos a la estadística inferencial que como su nombre mismo lo dice busca inferir mediante una muestra estadística el comportamiento de una población lo que podría traducirse que a las conclusiones obtenidas se le busca una nueva conclusión más amplia basada en la primera estas estimaciones se llevan a cabo mediante el uso de hipótesis las cuales posteriormente se validan o se refutan (López, 2018).

1.9 Modelos de regresión

La regresión es un modelo estadístico que permite cuantificar la relación que puede existir entre 2 o más variables pudiendo predecir valores de una variable dependiente, de una variable independiente. Existen 3 modelos principales los cuales son regresión simple lineal, regresión simple logística y regresión simple proporcional (Molina, et al., 2016)

1.9.1 Calidad de los Datos.

Antes de realizar cualquier modelo es necesario presentar unos buenos datos pues al final y cabo el resultado del modelo y su precisión se verá medido en base (Galway & Hanks,2011). Asimismo, como se citó en Muñoz, et al. (2018)) esta tabla ejemplifica el porqué del problema de la calidad de datos.

Tabla 1.

Problema de la calidad de datos.

Tipo Problema	Síntomas	Causas
Operacional	Existen, no son precisos o no son válidos	Problemas con la captura o transmisión de data
Conceptual	Los datos que están perdidos, que no son precisos, o no son válidos	Los datos no están bien definidos, o no son aptos para el uso previsto.
Organizacional	Hay problemas operacionales o conceptuales que son persistentes	Hay desconexiones entre la organización que recoge y que usa los datos.

Nota. Adaptado de (Muñoz et al, 2018).

1.10 Servidores de Base de Datos Relacionales

Es el primer nivel del cual se obtienen datos y se crea conocimiento. Este sistema se rige bajo el modelo cliente-servidor, que le brinda al usuario servicios de gestión, administración y protección de la información a través de conexiones de red, controladas por unos protocolos ya previamente definidos y a los que acceden los usuarios. Estos surgieron en la década de los 60 para poder ofrecer soluciones de manejo y administración de volúmenes de datos a grandes empresas (Prometeus Global Solutions, 2021).

1.11 Kmeans Neighbor

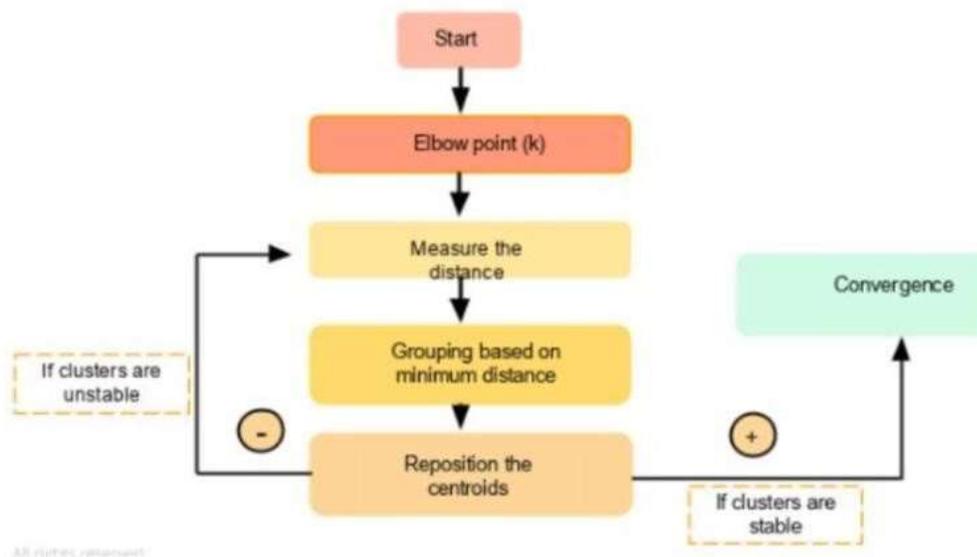
Kmeans Neighbor es un algoritmo que almacena los datos y posteriormente los clasifica en función de una medida de similitud, este agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o clúster, consta de 3 pasos una vez dispongamos de los datos.

Inicialización: una vez escogido el número de grupos, k , se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.

Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.

Figura 1.

Asignación de objetos a los centroides.



Actualización centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

1.12 Como Elegir el Valor de K

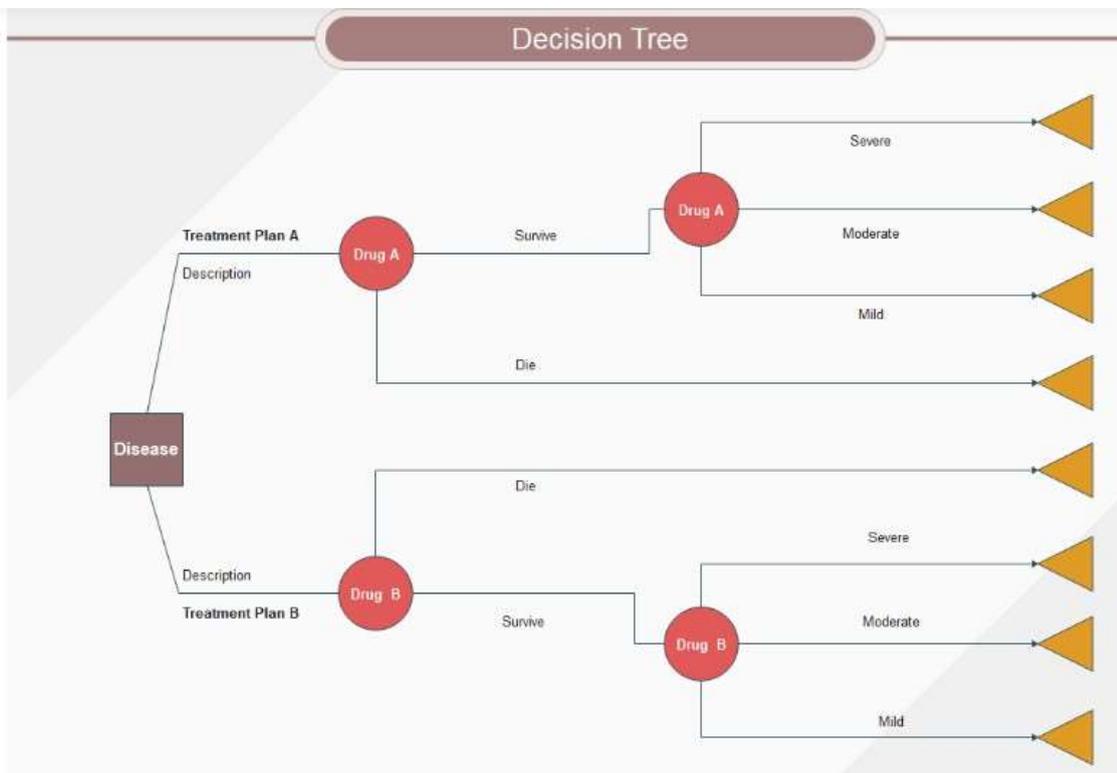
Para saber la cantidad de clúster en los datos se debe ejecutar el algoritmo por un valor de K , no existe un método definido, pero una de las métricas que se usa para comparar resultados es la distancia media entre los puntos de datos y su centroide, como el valor de la media disminuirá a medida que se aumente el valor de K se debe usar el momento donde la tasa de descenso de mueve del resto de una manera brusca y encontrar el punto "codo" de ahí el nombre de esta técnica (Navia, 2018).

1.13 Clasificación de los Datos

Cuando se llega a la etapa de clasificación es en donde se determina a que clúster ira cada punto. Para lograr todo esto se calcula la distancia que existe entre el punto en cuestión y cada uno de los centroides, la partencia de cada punto queda definido por la cercanía que existe entre este y el centroide formando así un clúster (Enriquez & Aracena, 2013).

1.14 Árbol de Decisiones Figura 2.

Árbol de decisión.



El árbol de decisiones es una técnica que emplea data tanto nominal, numérica y alfabética, generando una serie de resultados basados en una serie de problemas, es capaz de procesar data faltante o errónea, el árbol de decisión extrae información de grandes cantidades de data y la simplifica generando los posibles resultados los cuales luego se pueden usar para clasificar nuevamente dando paso a un Random Forest (Madan et al., 2018).

Como se puede observar en la ilustración 1 el árbol de decisiones genero las posibles soluciones a los tratamientos A y B, generando las probabilidades de sobrevivir o morir de cada uno y a su vez la probabilidad de si acaso sobrevives las consecuencias de este serán moderadas, severas o leves para cada uno de los tratamientos.

La construcción de los árboles de decisión se basa en la creación de escenarios donde los agentes interactúan entre sí de manera consecutiva lo que es distinto al análisis estático basado en donde los escenarios aquí se sucedían de manera simultánea. Por otro lado, los resultados que se arrojen se definen por sucesos que y decisiones presentes, dando consi go en el árbol las consecuencias de eventos probables que podrían pasar a futuro (Riascos & Molina, 2016)

Siguiendo el ejemplo de la gráfica de un árbol, el grupo de datos inicial o base es el nodo raíz, los conjuntos secundarios o subconjuntos particionados son nodos y los no particionados son ‘hojas’ estas partes del árbol son estructuradas a través de la secuencia de reglas de partición. Las ramas de los nodos son subgrupos creados a partir de los mismos. La finalidad de crear árboles de decisiones es organizar o clasificar un clúster de objetos con atributos o características diferentes o individuos en agrupaciones más homogéneas y reducidas (Rojas et al., 2014).

Estructura de un árbol de decisión

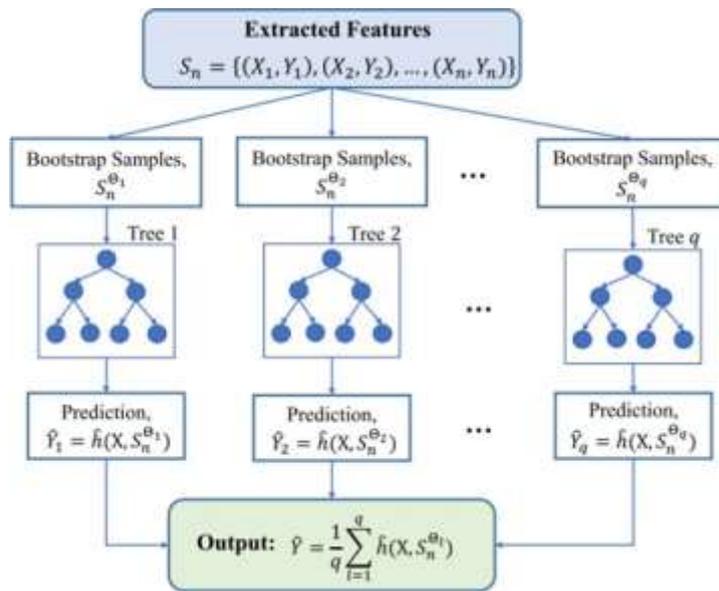
- Nodos: Identificadores de los atributos.
- Ramas: Posibles valores de los atributos.
- Hojas: Conjuntos ya clasificados y etiquetados con nombre de una clase.

1.15 Bosques Aleatorios

En términos simples podemos decir que el Random Forest es un conglomerado de árboles de decisiones.

Este es capaz de generar cientos o inclusive miles de árboles de decisiones que actúan como si fueran funciones de regresión y al momento de brindar la predicción final se le puede decir que es el promedio de las otras salidas de todos los árboles, El árbol crece mediante el proceso de aprendizaje de los datos introducidos, cada uno de esos árboles consta de nodos de decisión, Los nodos de decisión evalúan cada muestra ingresada por una función de prueba y error que luego pasa de rama en rama basándose en las características de cada una. X representa el vector de entrada conteniendo m características con $X = (X_1, X_2, \dots, X_m)$ Y , y el escalar de salida S_n el conjunto de entrenamiento que contiene n observaciones que se puede expresar como:

$$S_n = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}, X \in R^m, Y \in R$$



1.15.1 Algoritmo de Bosque Aleatorio

Este algoritmo se le puede decir que es una ampliación del método de agregación auto dominante, ya que hace uso de esta junto a la aleatoriedad de características para crear un bosque que no esté correlacionado de árboles de decisiones, esta aleatoriedad de características genera un subconjunto aleatorio de características lo que genera una baja correlación entre los árboles de decisión, esta es su principal diferencia pues mientras que los árboles de decisiones muestras todas las posibles divisiones de características a tomar, los bosques aleatorios solo toman un subconjunto de esas características (IBM, S.F.)

1.15.2 Conocer el Funcionamiento del Algoritmo del Random Forest

Para desarrollar un bosque aleatorio se logra siguiendo los siguientes pasos Cardellino (2021):

Paso 1: El algoritmo selecciona muestras en forma aleatoria de la base de datos proporcionada.

Paso 2: El algoritmo creará un árbol de decisión para cada muestra seleccionada. Luego obtendrá un resultado de predicción de cada árbol creado.

Paso 3: A continuación, se realizará la votación para cada resultado previsto. Para un problema de clasificación, usará la moda, y para un problema de regresión, usará la media.

Paso 4: Y finalmente, el algoritmo seleccionará el resultado de predicción más votado como predicción final.

1.16 Marco Conceptual

1.16.1 E-commerce

El e-commerce o comercio electrónico es definido como la compra y venta de información, productos y/o servicios a través de medios electrónicos. Esto incluye el intercambio de datos y fondos. Se define también como la aplicación de las TIC's en la cadena de valor de un negocio desde su origen hasta su final y que facilita la conexión de modelos Business to Business, Business to Consumers y Consumers to Business, con el objetivo de reducir costos que giran en torno al producto, la logística y la atención al cliente. El e-commerce es un campo multidisciplinario que incluye un conjunto de operaciones técnicas, de comunicación, seguridad informática, cobros, pagos, propiedad intelectual, compras, logística y otros (Gómez & Ramos, 2011).

1.16.2 Clúster

Según Andoin (2020), un clúster se define como una agrupación de datos que comparten relación entre ellos. Los clústeres suelen ser conocidos por su forma, dependiendo el objetivo de la búsqueda, como en clústeres elipsoidales, que su centro de masas está en su interior.

1.16.3 Algoritmos

Un algoritmo es una secuencia de instrucciones codificadas en un software que, al ser ejecutadas, requieren que el usuario tome las acciones necesarias para solucionar un problema específico (Choque & Encizo, 2020). Los algoritmos informáticos operan mediante la entrada y salida de datos. Al recibir los datos de entrada, el sistema analiza la información proporcionada y ejecuta las instrucciones adecuadas para obtener el resultado deseado (Revilla, 2021).

Secuencia Lineal. El algoritmo progresa mediante la ejecución secuencial de tareas o declaraciones.

Condiciona. El algoritmo realiza una elección entre dos opciones, basándose en condiciones predefinidas, es decir, si X es igual a 5, se ejecuta la acción Y (Bernal, 2020).

Bucle. El algoritmo consta de una serie de instrucciones que se repiten múltiples veces en secuencia (Manrique, 2021).

1.17 Marco Legal

En el ámbito legal se han desarrollado varios planes de acuerdo con la legislación, entre ellos: Plan Nacional de Telecomunicaciones y Tecnologías de la Información del Ecuador 2016-2021. Acuerdos marco de cooperación interinstitucional celebrados con varias unidades estructurales, p. Ministerio de Telecomunicaciones y Sociedad de la Información, Educación Superior, Ciencia, Tecnología e Innovación. Plan Nacional de Desarrollo (2013-2017).

Ley de Telecomunicaciones. Código de Organización Socioeconómica del Conocimiento, la Creatividad y la Innovación. Así como artículos sobre el desarrollo de la investigación científica y la innovación tecnológica, las TIC y diversos campos del conocimiento en la Constitución de la República del Ecuador, donde se puede destacar (Art. 281, Art. 385, Art. 387, Art. 423, Art. 388, Art. 313) que permiten alinearse y ser partícipes de la adopción de estas tecnologías emergentes para el cambio y adopción de una cultura de transformación digital para un mejor desarrollo a nivel empresarial en el ámbito local con proyección internacional, lo que se convierte en un impacto positivo

Capítulo II: Metodología

2.1 Población y Muestra

En el presente trabajo de investigación, al usar aprendizaje automatizado, no existen conceptos de población y muestra.

2.2 Métodos y Materiales

Se usará como métodos los algoritmos de Machine Learning de K-Means Neighbors y Random Forest. Así mismo, se mencionarán conceptos de Regresión para el pronóstico de la Clasificación.

2.3 K-Means

2.3.1. *Set de Datos*

Para la ejecución del modelo a emplear se va a utilizar una base de datos sin clasificación. Una base de datos no relacionadas es la óptima para el modelo K-Means ya que tiene una mejor escalabilidad y está pensada para grandes cantidades de datos lo que lo hace al K-Means el candidato perfecto ya que este está diseñado para encontrar patrones y hacer predicciones de grandes cantidades de datos.

Entre las ventajas que se le pueden atribuir a este tipo de bases podemos encontrar:

Que son más flexibles a la hora de realizar esquemas y bosquejos de información, pero recordando que no siguen al lenguaje de consulta estructurado (SQL) el cual se puede utilizar como método execute para crear y manipular directamente las bases de datos jet y crear consultas SQL de paso a través, para así poder manipular bases remotas de clientes – servidor (Poirier & Gabillaud, 2019)

Este tipo de base se usa para almacenar datos no estructurados como semiestructurados. A diferencia de una base de datos con clasificación esta no garantiza las cualidades ACID, es decir la atomicidad, consistencia, integridad y durabilidad

2.3.2. *Estandarización de los Datos*

Se usará la estandarización de los datos para poder ajustar los valores medidos en diferentes escalas respecto a una escala común.

Se realizará la estandarización de datos debido a que K-Means está pensado para datos continuos, no para datos discretos, si no se la realizara probablemente los centros de los clústeres serían anómalos.

Min-Max. En este escalamiento se busca mover los valores hacia la media de la columna, en esta el resultado se da por la resta del valor menos el valor mínimo entre la diferencia entre el valor máximo y mínimo de la columna según esta formula

Z Score. Este es el resultado de restar cada valor entre la media poblacional y dividirlo entre la desviación estándar. Cabe destacar que cuando se hace una estandarización por variable es mejor usar Min-Max ya que esta intenta acercar los valores a la media, pero cuando existan outliers o datos atípicos que son relevantes es mejor usar Z score ya que no se pierde mucho su impacto.

2.3.2. *Estandarización Generalizada*

En esta estandarización los valores no se transforman individualmente, sino que lo hacen en relación con otros valores ya presentes en la tabla de datos. Es importante estandarizar los datos ya que en los algoritmos de clustering como es el K-Means se usa la distancia euclidiana para medir la similitud, esto debido a que minimiza la distancia media entre los puntos y los centroides, con la estandarización de datos los criterios de distancia se puedan aplicar de mejor manera cuando se comparan características y así la influencia de uno no se vea exagerada con relación al otro debido a las diferencias de sus escalas.

Matemáticamente podemos decir que la distancia euclidiana es la longitud de un segmento de línea entre los dos puntos

En una sola dimensión podemos decir que la distancia entre dos puntos en la línea real es la diferencia numérica de sus coordenadas, es decir:

La misma fórmula, pero elevada al cuadrado y luego dividida para la raíz cuadrada da el mismo resultado, pero quita cualquier valor negativo previo y pone su absoluto

Cuando se habla de dos dimensiones o más (p&q) pasan a tener coordenadas siendo $(p_1, p_2 - q_1, q_2)$ En n dimensional Espacio Euclidiano, ahí la distancia es de:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Posterior a la estandarización se procede a realizar una matriz de distancias para poder ver que variable tiene correlación con otra, se entiende como matriz de distancia a una matriz cuadrada en la cual los elementos representan la distancia entre puntos.

Esta matriz debe ser estrictamente simétrica de tamaño NxN, esto dado en conjunto de N puntos en el espacio euclidiano, un espacio geométrico que responde a las 3 dimensiones

Se realiza la matriz de distancia en cada prueba y muestra del entrenamiento del modelo debido a que el resultado predicho depende de esta, una vez hecha esta matriz el algoritmo selecciona las k de muestras de entrenamiento que son las más cercanas a la muestra de prueba y así predecir en función del valor mayoritario si fuese para clasificación o promedio si fuese para regresión.

Donde k es el número de centroides seleccionados, n es el tamaño del conjunto de entrenamiento y es el número de dimensiones.

Luego de esto se procede a realizar la selección optima de k números de clusters, esto se lo realiza por medio el método del codo (elbow method) en ese método estamos variando el número de k desde 1 a N clusters y así poder ver la mejor distribución de clusters.

En cada valor de k se le está calculando la suma de cuadrados de la distancia (WCSS), Esta se entiende que es la suma de la distancia al cuadrado entre cada punto y el centroide en un grupo, el valor de la suma de cuadrados de la distancia es mayor cuando $k = 1$, cuando ya es hora de analizar el grafico que el grafico cambiara drásticamente en un punto creando el asemejo al dobles de un codo y es ahí en ese corte el números óptimo de clústeres a elegir para ese data set y a partir desde ese punto el grafico se moverá casi paralelo al eje x.

2.4. Random Forest

El algoritmo de Random Forest es conocido como una técnica de aprendizaje supervisado que produce diversos árboles de decisión a partir de un conjunto de datos de entrenamiento. Los resultados obtenidos se fusionan con la finalidad de obtener un modelo más robusto con relación a las salidas de cada árbol por separado (Jesús, 2020). Cada árbol se obtiene mediante dos etapas:

1. Se produce un número significativo de árboles de elección utilizando el grupo de datos. Cada árbol incluye un conjunto aleatorio de variables m (indicadores) de manera que m sea menor que M (donde M es el número total de predictores)
2. Cada árbol se desarrolla hasta alcanzar su expansión máxima.

Cada árbol que surge del algoritmo de Bosque Aleatorio incorpora un conjunto de observaciones seleccionadas al azar (a través de la técnica estadística Bootstrap, que implica obtener muestras de una población en las que una misma observación puede estar presente en más de una muestra).

Las observaciones que no son consideradas en la construcción de los árboles, también conocidas como ‘out of the bag’ se emplean para validar el modelo. Las salidas de todos los árboles se agrupan en una salida final, ‘Y’ conocida como ensamblado, que se logra mediante algún criterio.

2.4.1. Análisis del Proceso del Random Forest

El modelo del algoritmo Random Forest está estructurado por un *ensemble* o conjunto de árboles de decisión individuales, cada uno capacitado utilizando una muestra al azar obtenida por medio de bootstrapping aplicado a los datos originales de entrenamiento. Esto deja implícito que cada árbol se entrena con unos datos levemente variados. En cada árbol específico, las observaciones se distribuyen a lo largo de las ramificaciones (nodos) para crear la estructura del árbol hasta llegar a un nodo final. La predicción de una nueva observación se logra al combinar las predicciones de todos los árboles únicos que componen el modelo.

2.4.2. Métodos de Ensemble

Todos los modelos de aprendizaje estadístico y Machine Learning enfrentan la problemática de equilibrio entre varianza y bias.

El término bias o sesgo, se refiere a la medida en la que las predicciones de un modelo

se distancian en promedio de los valores reales Indica la habilidad del modelo para comprender la auténtica relación entre los predictores y la variable de respuesta. El concepto de varianza alude a la cantidad en que el modelo se altera en función de los datos empleados en su entrenamiento. En un escenario ideal, el modelo no debería sufrir cambios significativos debido a pequeñas fluctuaciones en los datos de entrenamiento; si esto sucede, implica que el modelo está memorizando los datos en lugar de captar la auténtica relación entre los predictores y la variable de respuesta.

A modo de ejemplo, un modelo de árbol con una gran cantidad de nodos tiende a modificar su configuración incluso ante mínimos cambios en los datos de entrenamiento, lo que resulta en una alta varianza. Conforme el modelo se vuelve más complejo, aumenta su capacidad para adaptarse a las observaciones, lo que resulta en una disminución del sesgo y una mejora en la capacidad predictiva. Sin embargo, al alcanzar cierto nivel de complejidad, surge el problema de sobreajuste (overfitting), en el cual el modelo se ajusta en exceso a los datos de entrenamiento, lo que limita su habilidad para realizar predicciones precisas en nuevas observaciones. El modelo óptimo es aquel que logra un equilibrio ideal entre el sesgo y la varianza.

Las técnicas de *ensemble* unen varios modelos para formar uno nuevo con el propósito de alcanzar un equilibrio entre el sesgo y la varianza, obteniendo predicciones superiores a las de cada uno de los modelos originales. Dos de los enfoques de ensemble más empleados son:

Bagging. Se configuran diversos modelos, cada uno entrenado con un conjunto único de datos de entrenamiento. Al realizar predicciones, todos los modelos en el conjunto contribuyen con sus pronósticos. La predicción final se obtiene al calcular el promedio de todas las predicciones (en el caso de variables continuas) o seleccionando la clase más común (en el caso de variables categóricas). Los modelos de Bosques Aleatorios permanecen en esta categoría.

Boosting. Se configuran en serie varios modelos simples conocidos como "aprendices débiles" (weak learners), de manera que cada modelo aprende de los errores del previo. En última instancia, al igual que en el enfoque de Bagging, se calcula el promedio de las predicciones de todos los modelos o se selecciona la clase más común.

2.5 Construcción del Modelo (Data Train)

Se proporciona un resumen del modelo que ha sido entrenado utilizando el 70 % o 75% de los datos, y luego se utiliza este resumen para entrenar un modelo de Bosque Aleatorio con un

número determinado de árboles. En cada uno de estos árboles se requiere seleccionar una cantidad específica de variables para cada división.

La estimación de la tasa de error OOB se calcula desde las observaciones fuera de la bolsa.

2.5.1. *Out-of-Bag Error*

Debido a la característica del proceso de bagging, es viable calcular el error de prueba sin requerir la implementación de técnicas de validación cruzada (cross-validation). La particularidad de que los árboles se ajusten usando conjuntos de datos generados mediante bootstrapping implica que, en promedio, cada ajuste se realiza utilizando aproximadamente dos tercios de las observaciones de origen. El tercio restante es denominado como out-of-bag (OOB).

Si se realiza un seguimiento de las observaciones utilizadas en la construcción de cada árbol durante el proceso de bagging, es factible anticipar la respuesta de la observación i al emplear aquellos árboles en los cuales dicha observación ha quedado excluida, promediándolos. Mediante esta metodología, se pueden obtener predicciones para las n observaciones y, en base a estas, calcular el error cuadrático medio fuera de la bolsa (OOB-mean square error, para regresión) o el error de clasificación fuera de la bolsa (OOB-classification error, para árboles de clasificación).

2.5.2 *Matriz de Confusión*

Después de calcular la estimación del error fuera de la bolsa (OOB) para los datos en cuestión, es esencial realizar un análisis de la matriz de confusión. En esta matriz se anotan los valores predichos del Random Forest en contraste con los valores observados.

2.5.3 *Importancia de los Predictores*

Si bien es cierto que el método de *bagging* logra mejorar la capacidad predictiva en comparación con los modelos basados en un solo árbol, esto conlleva un compromiso: la capacidad de interpretar el modelo se ve reducida. Debido a su naturaleza como una combinación de múltiples árboles, no es factible obtener una representación gráfica simple del

modelo y no resulta inmediato identificar visualmente cuáles predictores son más relevantes. No obstante, se han ideado nuevas estrategias para medir la importancia de los predictores, lo que convierte a los modelos de bagging (Random Forest) en una herramienta extremadamente poderosa, no solo para la predicción, sino también para la exploración analítica.

2.5.4 Evaluación del Modelo (Data Test)

Para llevar a cabo la validación del modelo, se procede a crear la matriz de confusión utilizando los datos de prueba o los datos de validación.

Capítulo III: Análisis de Resultados

Pasos para la creación de modelos y ejecución.

Para el desarrollo de los modelos se hará uso del software estadístico “R studio” y la información a usarse es extraída de una empresa retail.

3.1 Modelo K Means Neighborhood

En primera instancia se realizará el modelo de machine learning K-MeansNeighborhood para poder evaluar a los clientes y poder categorizarlos en base a sus similitudes. Para poder realizar lo anterior primero se procede a llamar a la base y a nombrarla con un nombre para poder trabajar a partir de ello tal y como se muestra a continuación:

```
Clientes <- read.csv2("../Data/BASE TESIS - PRUEBA 299OBS.csv")
```

Esta función mediante el comando read.csv2 carga la base con la cual se va a trabajar y Tía es el nombre que se le acredita para poder trabajar la base como un objeto.

Figura 3.

Captura de Rstudio vista de los datos.

X	Client.Document	Nombre	GEN	EDAD	Payment.System.Name	Address.Type	Coupon	Shipping.Value	X.Transacciones	Ventas.Totales
1	1207101229	ABEL BARRAGAN DIAZ	MASCULINO	43		4	1	0	1	169.00
2	950129433	Abraham Jurado	MASCULINO	29		4	1	0	2	1418.00
3	929221976	Abraham Muñoz	MASCULINO	23		4	0	0	1	399.00
4	1212082334	adolfo ibarra	MASCULINO	37		1	1	0	1	108.00
5	940821569	Adrian fabricio Suarez del valle	MASCULINO	29		1	1	0	1	134.00
6	942709775	Adrian Pm	MASCULINO	25		4	1	0	2	1520.00
7	929883881	ADRIANA RODRIGUEZ PILAY	FEMENINO	36		4	1	0	1	194.00
8	1208350889	ALEIDA MARIA SOLOZANO ARTEAGA	FEMENINO	27		3	1	0	2	567.00
9	104671964	Alex Amador Sanchez Morales	MASCULINO	25		1	1	0	2	2889.00
10	911306260	Alex Suarez Carr	MASCULINO	47		3	1	0	1	148.00
11	1054895929	Alex Espinosa	MASCULINO	50		1	0	0	1	157.00
12	903660133	alex raque	MASCULINO	36		4	0	0	1	84.00
13	954062188	ALEXANDER JOSUE VERA VALEZ	MASCULINO	50		1	0	0	1	43.00
14	1206542529	Alexandra del valle Ruiz chica	FEMENINO	35		4	1	0	2	3014.00
15	1759523244	Alexandro Padra	FEMENINO	46		4	1	0	1	13.00
16	605901099	Alejo Joel Wilza Lopez	MASCULINO	39		4	0	0	1	84.00
17	704573583	alison castro	FEMENINO	21		4	1	0	2	7864.00
18	1206165464	Alison Italy Pizarri Jimenez	FEMENINO	36		4	1	0	1	80.00
19	917419442	Anabel Jimenez	FEMENINO	57		4	0	0	1	79.00
20	1719613902	Jordenson Herrera	MASCULINO	49		4	1	0	1	39.00
21	1315278513	Andrés Anzures	MASCULINO	46		1	1	0	1	40.00
22	2000021562	Josber Guerrero	MASCULINO	34		4	0	0	1	30.00
23	1719182188	ANDRES VILLARREAL	MASCULINO	42		4	1	0	1	229.00
24	309475135	andy andrade	MASCULINO	48		3	1	0	1	149.00
25	93248919	Anelise Viquez Tamayo	FEMENINO	40		4	0	0	1	97.00
26	929690428	ANGELCHOIZ	MASCULINO	39		3	1	0	1	299.00

Esta es una vista no completa de todos los datos, pero permite mostrar cómo se ve la base en r studio.

Al momento de trabajar con el algoritmo K-Means es importante recordar que esta solo se puede trabajar con variables cuantitativas, es por ello que se cómo siguiente paso se procederá a eliminar las variables cualitativas creándose así un nuevo dataset.

Cientes1 <- Clientes[,c(-1,-2,-3,-4)]

Como se ve en el código se eliminaron las variables 1,2,3 y 4, las cuales eran X, client.document, Nombre y GEN dándonos consigo ahora una base de 7 variables solamente con respecto a las 11 que eran originalmente.

Figura 4.

Captura de R studio, variables cuantitativas.



Luego de realizar los pasos anteriores es momento de instalar las librerías que se van a utilizar en la implementación del algoritmo, esta es un conglomerado de funciones y códigos que facilitan la programación del código para agilizar la ejecución del modelo

Aquí en K-Means se harán uso de:

Tidyverse: Esta librería es una colección de paquetes que está diseñada para la Datascience

Cluster: Esta librería permite ejecutar métodos para el análisis de conglomerados

Factoextra: Mediante el uso de este paquete se facilita la extracción y visualización de los resultados dados por los análisis exploratorios multivariados.

Nbclust: Este paquete es el que nos va a permitir determinar cuál es el numero óptimo de clústeres en el dataset

Tidyr: Este paquete diseñado por r mismo permite ordenar los datos “sucios” del dataset y convertirlos en objetos de datos en R mismo.

Si es la primera vez que se hace uso de estas librerías se debe instalarlas mediante el comando “**install.package()**”, una vez ya habiéndolas descargadas para poder llamarlas y estar lista para su uso se debe hacer mediante el comando “**Library()**”.library(tidyverse) library(cluster) library(factoextra)library(NbClust) library(tidyr)

Una vez hecho estos se precederá a realizar una escalación de datos con el fin de

normalizar los datos, Esta normalización se realiza con el fin de que todas las variables a trabajarse se les de la misma relevancia y prevenir que se produzca algún error a la hora de correr el programa, esta escalación se la realizará mediante el uso de la función “Scale”

```
Cientes1Esc <- scale(Cientes1)
```

Una vez normalizado los datos se procederá a conocer el numero óptimo de clústeres en los que se dividirá el k means

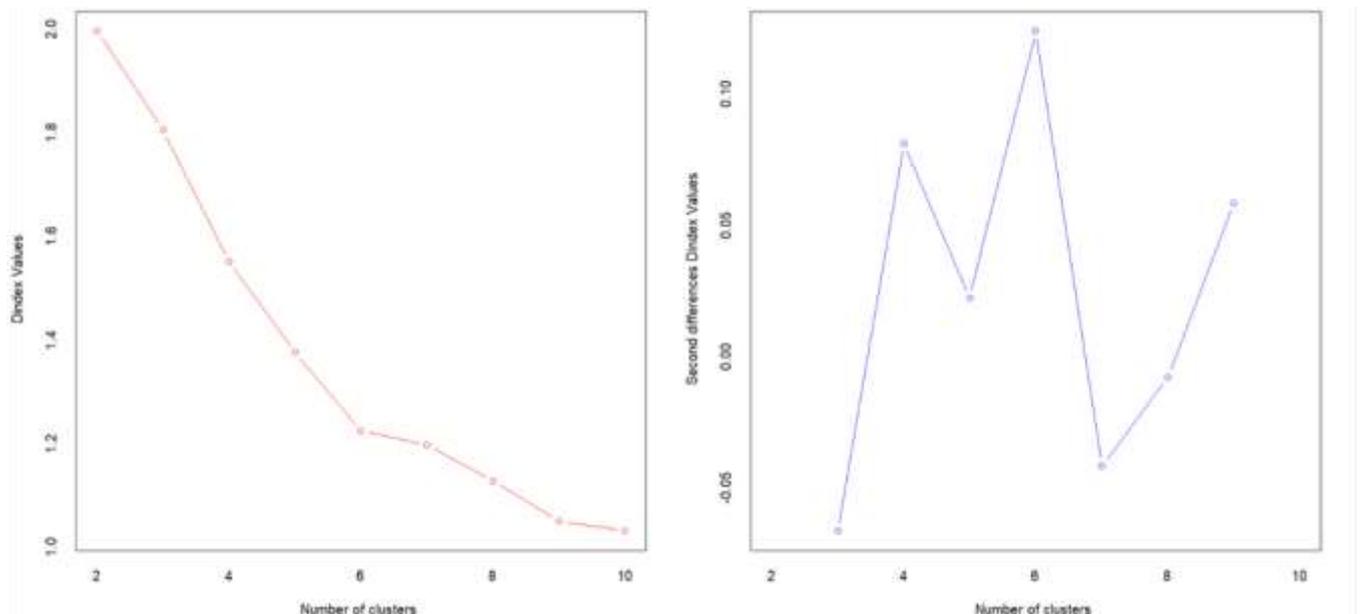
```
resnumclust<-NbClust(Cientes1Esc, distance = "euclidean", min.nc=2, max.nc=10,  
method = "kmeans", index = "alllong")
```

Aquí se hace uso de la función NbClust para poder hallar el numero óptimo de clusters, se hace uso de la distancia euclidiana con el fin de minimizar la distancia media entre los puntos y los centroides, el valor mínimo de clústeres posibles es 2 y el valor máximo de clústeres a por evaluar será 10, el método que se usará será K means y el índice de medición será todo el tiempo.

Una vez ejecutado el programa se cargarán 2 datos en r studio:

Figura 5.

Captura de R Studio visualización de graficas de clústeres óptimos.



En el primer grafico correspondiente a la gráfica del método de Hubert se puede ver que el corte conocido comúnmente como el “Codo” de la función el punto donde la medición

deja de decrecer y los valores tienden a mantenerse casi iguales, en esta se observa que el numero óptimo de clústeres es 6, en el segundo el grafico del índice D se busca un pico significativo en el aumento de la medición, en este caso el valor óptimo de clústeres es de 3, ahora entraríamos en duda de cual valor seria el óptimo a escoger para realizar el k means sin embargo R studio evaluando todos los índices indica que la mejor opción a escogerse será 3 clústeres

Ahora bien, ya conociendo el número de clústeres óptimos se procederá a calcular el K-Means:

Figura 6.

Captura de R Studio Resultado arrojado para el numero óptimo de clústeres.

```
Among all indices:
5 proposed 2 as the best number of clusters
7 proposed 3 as the best number of clusters
1 proposed 4 as the best number of clusters
7 proposed 6 as the best number of clusters
6 proposed 9 as the best number of clusters
1 proposed 10 as the best number of clusters

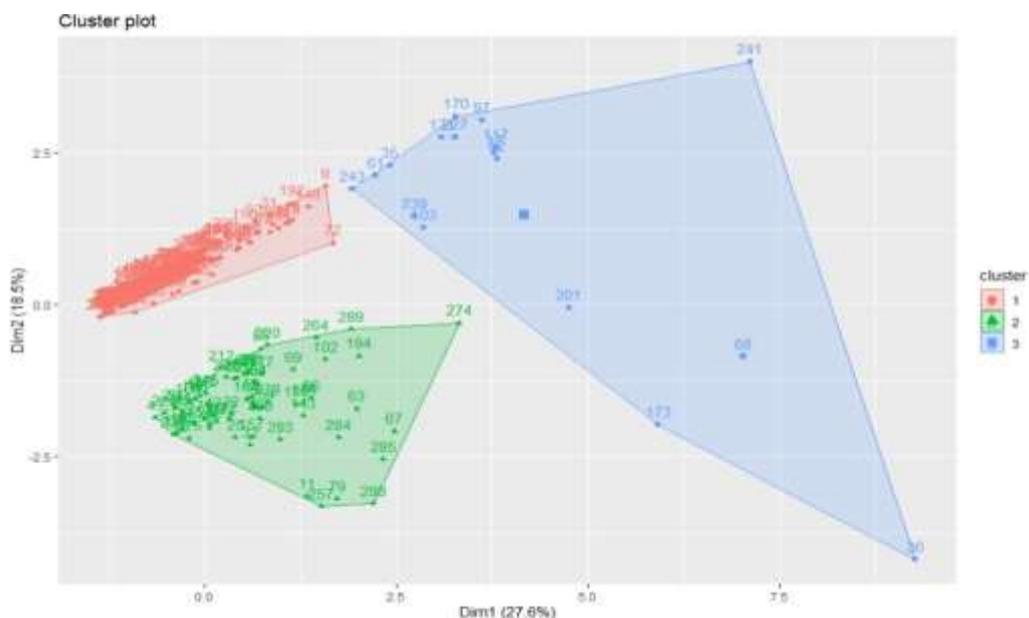
***** Conclusion *****

According to the majority rule, the best number of clusters is 3
```

Con 3 centroides, para ello se ejecutará el siguiente código

```
k3 <- kmeans(Clientes1Esc, centers = 3, nstart = 25)
```

Una vez ya teniendo los resultados se procederá a graficarlo para poder visualizarlo de una mejor manera, se hará por medio de la función “Fviz_cluster”



fviz_cluster(k3, data = Clientes1Esc)Figura 7.

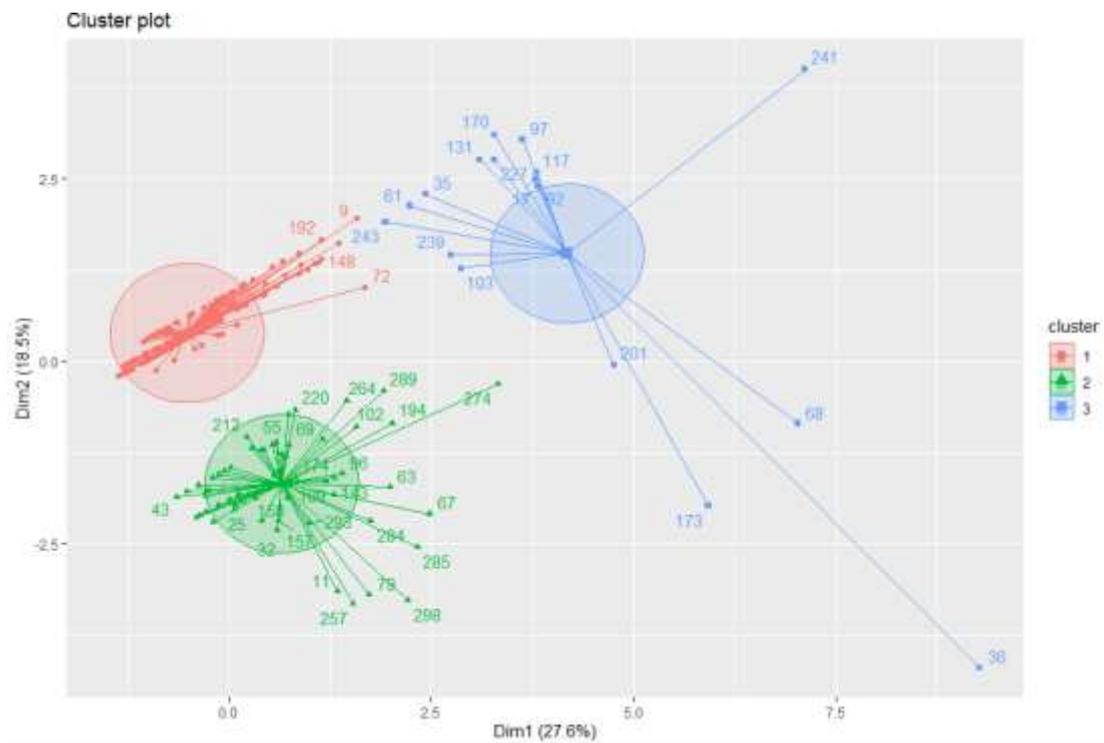
Cluster plot.

En este primer gráfico generado por el algoritmo K-Means, con el código de `fviz_clúster`, se lo conoce como gráfico original, el cual nos indica la diferenciación de a que clúster pertenece cada uno de los datos, los cuales están clasificados por semejanzas en características y necesidades.

fviz_cluster(k3, data = Clientes1Esc, ellipse.type = "euclid",repel = TRUE,star.plot = TRUE) #ellipse.type= "t", "norm", "euclid"

Figura 8.

Cluster plot.



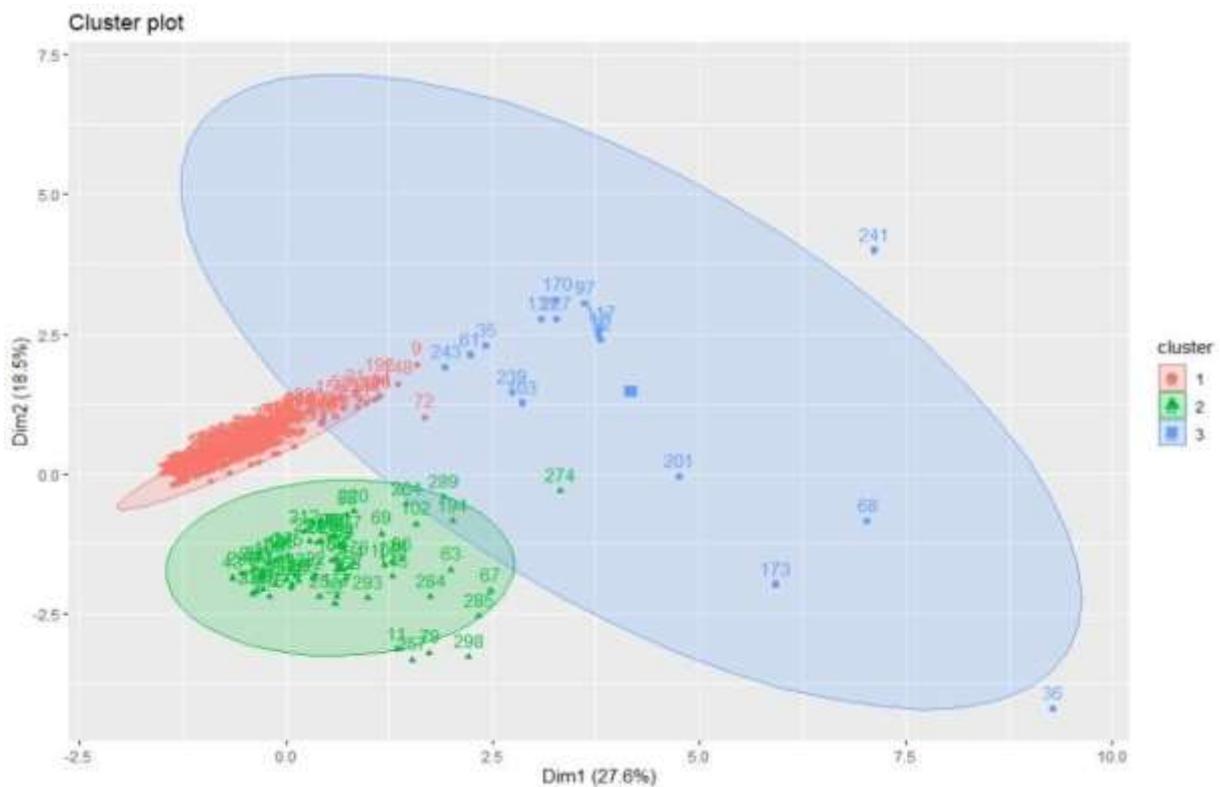
En esta parte del código podemos observar el método a usar que en este caso es la distancia euclidiana, la cual se usa como una medida de dispersión de un conjunto para la agrupación de los clústeres. Así mismo, se agregará un código para generar el gráfico de forma de estrella en la que se pueda apreciar la tendencia de los datos en cada clúster de acuerdo con sus centroides.

Por ejemplo, podemos ver que por un lado hay datos del clúster 1 que tienen tendencia

a migrar al clúster número 3, sin embargo, por otro lado, se puede apreciar que los clústeres 2 y 3 tienen datos que no migran a un clúster ya existente, sino que tienen tendencia a migrar a un posible nuevo subconjunto.

fviz_cluster(k3, data = Clientes1Esc, ellipse.type = "norm") Figura 9.

Cluster plot.



En este tercer gráfico, se aprecia como los clústeres se transforman a en agrupaciones más robustas donde se puede visualizar un pronóstico de a donde pueden migrar los clústeres en un futuro.

Como se puede observar, con respecto a esta lógica, se prevé que un 5% o 10% de los datos de los clústeres 1 y 2 tienden a migrar al clúster 3. Así mismo podemos hacer mención a la gráfica anterior con respecto al clúster 3, que contiene ciertos datos dispersos, que probablemente puedan ser parte de un nuevo subgrupo.

Como siguiente paso se procederá a trasponer los resultados obtenidos en el K-Means a la base original, esto con el fin de poder validarla mediante el algoritmo de randomforest.

Clientes\$Kmeans <- k3\$cluster

Una vez teniendo los valores del k means como nueva variable de la base original se procederá a cambiar la variable categórica ahora formada y cambiarla para la clasificación por medio de la función “factor”, esto con el fin de que cuando se ejecute la clasificación su visualización sea más fácil

```
Clientes$Kmeans <- factor(Clientes$Kmeans,levels = c(1,2,3),  
labels = c("Bajo","Medio","Alto"))
```

Dicha conversión se la daría por medio de la función “levels” y “labels” en la cual se cambiaria el valor numérico 1 por bajo, 2 por medio y 3 por alto tal y como se muestra a continuación.

Figura 10.

R Studio K-Means y K-Means clasificación.



Al igual que el k means se eliminar las variables de que no representan relevancia en el modelo.

```
Clientes2 <- Clientes[,c(-1,-2,-3)]
```

Ahora se procederá a realizar el árbol de decisiones para la clasificación, para ello se necesitarán de nuevas librerías

Caret: El paquete caret (Classification and regression training) contiene un conjunto de funciones que intentan facilitar el proceso de métodos para la regresión y clasificación. Este paquete incluye herramientas como:

- Train and Test
- Cross Validation
- Ajuste por remuestreo (modelos)
- Selección de variables, entre otras

Este paquete se utilizará para llevar a cabo la clasificación del modelo, es el encargo de la división de datos de prueba y entrenamiento.

Rpart: El paquete rpart (Recursive Partitioning and Regression Trees) permite hacer arboles de decisión a partir de tablas, es decir, empezando de un conjunto de datos, y una fórmula de predicción, puede crear un árbol que pueda pronosticar la clasificación. Mediante esta paquetería se dispondrá de la función que permita crear el modelo del árbol de decisión.

Rpart.plot: Este paquete sirve como complemento de Rpart, el cual nos permitirá realizar gráficos o representaciones de los resultados obtenidos, como es el árbol de clasificación.

RandomForest: Este paquete será el encargo de crear el modelo de pronóstico.

library(caret) library(rpart) library(rpart.plot)

library(randomForest)

Una vez instalado las paqueterías requeridas se procederá al cálculo del árbol de clasificación por ello se procederá a entrenar los datos de la nueva base. Hay que mencionar que como el objetivo del modelo es predecir con que probabilidad se asignará cada clúster se debe disponer de data de entrenamiento que serviría como data de prueba y así esta podrá ser usada para facultar la eficacia del randomforest al momento de predecir. Este entrenamiento se logrará mediante el uso de la función “createDataPartition” incluida en la paquetería “caret”, la partición seleccionada para el modelo es de 80%/20% debido a la cantidad de datos que se poseen.

entrenamiento <- createDataPartition(Clientes2\$Kmeans, p= 0.8, list = F)

Una vez entrenados los datos se procede a crear el modelo mediante la función rpart

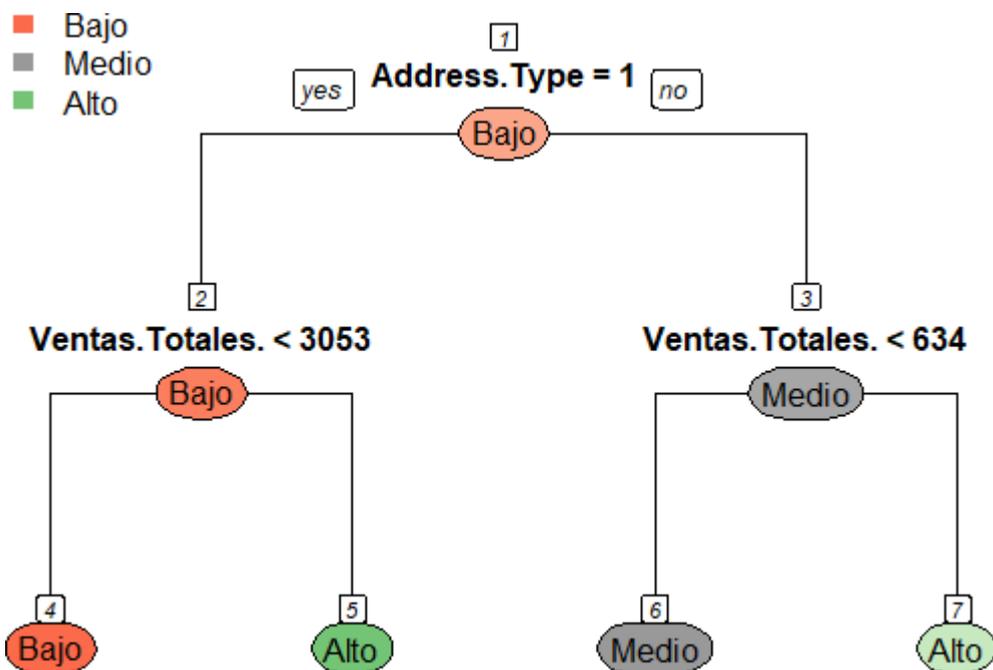
```
Cientes2Ar <- rpart(Kmeans~., data = Clientes2)
```

Como siguiente paso luego de haber concluido con el modelo se debe graficar el árbol para luego poder empezar con randomforest

```
rpart.plot(Cientes2Ar, type=1, digits = -1, extra= 0, cex = 1, nn=TRUE, fallen.leaves = TRUE)
```

Figura 11.

Ventas totales.



Luego de haber ejecutado el código este es el árbol de decisión que arrojo el modelo, como resultado podemos encontrar que las variables principales en las cuales se basó para el pronóstico fueron Address Type y las ventas totales, el árbol asumió a yes or no como el 1 y 0 de las variables y siendo estas 1 Pickup (Retiro en tienda) y 0 residencial (Domicilio), cada uno partía de un rango bajo tal y como se muestra, una vez continuando con las ramas ahí es donde empezaba la diferenciación de clientes donde si se escogiera la rama del yes -1- (pickup) y las ventas aquí fuesen mayor a \$3053 este cliente pasaría a ser considerado a rango alto pero si sus compras eran inferiores a \$3053 este seguiría siendo considerado de rango bajo, si por el otro lado se escogiese la rama del no el -0- (residencial) y las ventas aquí fueran mayores a \$634 el cliente pasaría a ser considerado de rango alto, pero si a su vez las ventas de este cliente no superarían los \$634 pasaría a ser considerado a rango medio.

Luego de tener listo el árbol de decisión es momento de comenzar a crear el modelo de pronóstico, para ellos se usará la función randomforest

```
Clientesa2Rf <- randomForest(x= Clientes2[entrenamiento, 2:8],
y
=Clientes2[entrenamiento, 9], xtest = Clientes2[-entrenamiento, 2:8], ytest = Clientes2[-
entrenamiento,9], ntree = 20000, importance=T,keep.forest=T)
```

Figura 12.

Matriz de confusión.

```
OOB estimate of error rate: 0.42%
Confusion matrix:
      Bajo Medio Alto class.error
Bajo  174     0   0  0.00000000
Medio   0    52   0  0.00000000
Alto    0     1  13  0.07142857
Test set error rate: 1.69%
Confusion matrix:
      Bajo Medio Alto class.error
Bajo   43     0   0  0.00000000
Medio  0    13   0  0.00000000
Alto   1     0   2  0.33333333
```

Antes de pronosticar la clasificación, el código ejecutado de random forest genera la matriz de confusión, la cual detalla los valores predichos del Random Forest con los actuales, así mismo nos da el valor estimado del error fuera de la bolsa o out-of-bag error que es de un 0.42%, quiere decir que es un buen modelo para pronosticar.

Al hacer el entrenamiento, visualizamos que se usa un 80% en entrenamiento y un 20% en prueba, debido a la cantidad de datos con los que se está trabajando, esto es con el fin de que el entrenamiento se lleve la mayor cantidad de datos y tienda a ser más robusto el modelo.

Como se puede observar se hace uso de la función randomForest para poder predecir con que probabilidad se va a clasificar el K- means previamente hecho, se hará un entrenamiento de la data para ser evaluada con las pruebas a un tamaño de 2000 árboles, la función T,keep,forest la validamos para que sea True debido a que la data de salida la queremos mantener para luego con ayuda de la base Clientes 2 evaluar su probabilidad con la que se ejecutará.

```
Clientes2$Probabilidad <- predict(Clientes2Rf,Clientes2,type = "prob")
```

13.

Probabilidad por clasificación de cliente.

Kmeans	Probabilidad[,"Bajo"]	Probabilidad[,"Medio"]	Probabilidad[,"Alto"]
Bajo	0.99920	0.00080	0.00000
Bajo	0.99220	0.00060	0.00720
Medio	0.08610	0.91150	0.00240
Bajo	0.99870	0.00130	0.00000
Bajo	0.99890	0.00110	0.00000
Bajo	0.97895	0.00140	0.01965
Bajo	0.99935	0.00060	0.00005
Bajo	0.99275	0.00100	0.00625
Bajo	0.94230	0.00095	0.05675
Bajo	0.99930	0.00070	0.00000
Medio	0.01455	0.98495	0.00050
Medio	0.00890	0.99090	0.00020
Medio	0.02380	0.97620	0.00000
Bajo	0.87315	0.00015	0.12670
Bajo	0.99515	0.00485	0.00000
Medio	0.00725	0.99275	0.00000
Alto	0.02055	0.01680	0.96265

Para concluir con el modelo se agregará los resultados obtenidos en el random forest por medio de variables a la base de datos empleada (Clientes2) esto con el fin de poder observar de manera tangible los resultados con los que se validará de manera porcentual la probabilidad de clasificación ya sea esta baja, media o alta.

Conclusiones de Resultados

Dentro de esta investigación se estableció como objetivo principal el análisis e implementación de modelos o algoritmos como el Kmeans y Random Forest para una óptima segmentación de tipos de clientes web en una empresa e-commerce. Para poder llegar a un resultado sólido con respecto a la viabilidad de esta metodología, se siguió un proceso que inicia con la recolección de datos de clientes, siguiendo con la aplicación de los modelos en mención, y finalizando con un análisis de los hallazgos. Esta base de datos cuenta con las variables más relevantes en relación con sus intereses y su frecuencia de compra en la plataforma, esto con el fin de poder obtener mayor asertividad al momento de elegir a qué segmento de clientes se va a dirigir una campaña publicitaria y que de esta manera, se pueda contar con clientes que se estén estrechamente fidelizados a la página web.

El algoritmo de k means neighbor fue implementado a partir de la base de datos proporcionada por la empresa, se realizó la estimación del número correcto de clústeres para distribuir estos datos, basándose en características similares de las variables evaluadas, las variables que se usaron para poder desarrollar este modelo fueron el género del cliente, el método de pago que usaba, el tipo de entrega de mercadería, su uso o no de cupones, el costo de envío por courier o delivery, el número de transacciones realizadas en el periodo y las ventas totales que realiza el cliente.

La ejecución del Kmeans Neighbor mostró la partición de los clústeres, se puede observar que el clúster en donde residen la mayor cantidad de los datos (Clientes) es el clúster 1 es decir el de los clientes de rango bajo, siguiendo en la lista en segundo lugar de mayor cantidad de clientes incluidos es el clúster 2 es el de los clientes de rango medio y finalmente el clúster con menos cantidad de clientes es el clúster 3 el de los clientes de rango alta, se puede notar que la mejor relación de datos la podemos encontrar en el clúster 1, la forma en que logramos corroborar esto es que los centroides son calculados por la media del total de datos para esa clasificación, y como se puede ver más exacto en el gráfico 8 del cluster plot de k means las líneas de tendencias de los datos aunque sí se dispersan, sus líneas se mantienen relativamente cercanas a su centroide, esto no se refleja en el resto de los clústeres evaluados, es más podemos notar como en el clúster 3 hay datos que se alejan excesivamente de su centroide siendo valores muy alejados a la media. También de acuerdo con la dispersión que presenta el cluster 2 y 3 podemos concluir que es muy probable que en un futuro junto a nuevos datos se llegue a formar un nuevo clúster para un nuevo tipo de cliente “super alto” podría ser.

Posterior a esto, ya contando con una clasificación de los datos generada por la aplicación del Kmeans Neighbor, se procedió a comprobar que esta segmentación esté correctamente dividida, esto se logró en primer lugar generando un árbol de decisión que sea el más adecuado a usar para llevar a cabo el pronóstico de la clasificación, pues de no hacerla inteligencia artificial iba a escoger cualquier árbol de decisión independientemente que sea o no el mejor para el modelo, este árbol fue el encargado, mediante el entrenamiento y prueba de datos, de arrojar qué variables eran las más relevantes en la clasificación de los clústeres y hacia qué cluster tornaba más relevancia cada una de ellas. La selección del árbol se llevó a cabo sobre un total de 299 árboles resaltando como más importantes las variables tipo de entrega de mercadería y ventas totales de cliente.

Mediante el uso de otra técnica de Machine Learning, que fue el Random Forest, se pudo pronosticar con qué probabilidad se le iba a asignar cada cluster a los clientes de la base de datos. Como primer paso, se generó una matriz de confusión la cual arrojó una estimación del error menor a 1%, que indica que tan bueno es el modelo por pronosticar, que, en este caso, sí lo es. Luego de esto, se hizo el entrenamiento y prueba de los datos, que finalmente dieron como resultado la predicción de la clasificación por clústeres, es decir, tenemos el porcentaje de acierto de cada partición, que como pudimos observar en los análisis de resultados, la mayor parte de los clientes tienen un 98% de probabilidad de estar en el clúster asignado por el algoritmo de Kmeans.

Es importante considerar que dicho modelo se puede enfrentar a problemas a la hora de clasificar correctamente los datos si es que no se llegase a contar con una buena base de datos y nos encontrásemos en realidad con una base con datos faltantes o muchos valores alejados entre sí. De ser este el caso realizar técnicas artificiales como datacleaning o sustituir los datos faltantes por la media de la variable podrían ser métodos que se podrían implementar con el fin de obtener una clasificación más precisa.

Conclusiones

En conclusión, podemos decir que este trabajo de investigación ha abordado de manera profunda los objetivos específicos planteados, los cuales fueron estructurados con relación a la necesidad de la empresa, que, en este caso, era el poder conocer a sus clientes para así poder clasificarlos en sus bases de datos, de acuerdo con sus características e intereses.

A través de una revisión literaria exhaustiva, se pudo dar paso a un análisis sólido de conceptos básicos como la estadística hasta conceptos como Machine Learning, que no son tan conocidos ni explorados por el talento humano hoy en día. Esta investigación no solo nos sirvió como base de conocimiento, sino también como una herramienta para poder reconocer las metodologías idóneas que nos ayude a obtener un mejor resultado en la segmentación de los clientes.

Además, se pudo evaluar la aplicabilidad y rendimiento de los modelos a implementar, como el K-Means Neighbors y Random Forest, los cuales dieron como resultado una óptima clasificación de tipo de clientes y un pronóstico de segmentación de estos. Se pudo observar que, además de su funcionalidad, fueron buenos modelos a implementar en este trabajo de investigación, ya que se pudo obtener un porcentaje de error menor al 1%, lo que significa robustez en los prototipos.

El logro clave de este estudio fue haber alcanzado una optimización exitosa en el proceso de clasificación de los clientes de la empresa, dando como resultado reducción de tiempo, eficiencia en la implementación de recursos y un mejor plan de inversión monetaria en campañas, que conllevan a un aumento del tráfico en la página web y por ende una mayor probabilidad de conversión de compra dependiendo de la finalidad de cada campaña elaborada por el departamento de marketing.

En resumen, podemos decir que esta investigación ha logrado alcanzar el cumplimiento de sus objetivos de manera global. Toda esta información además de dar soporte a las áreas que analizan a profundidad estas metodologías, puede dar pautas a distintas empresas que busquen optimizar sus procesos internos como la segmentación de diferentes bases de datos, con el fin de obtener una evaluación crítica a partir de la aplicación de distintos modelos estadísticos orientados al análisis de resultados.

Recomendaciones

Para un correcto desenvolvimiento a la hora de realizar futuras investigaciones se recomienda tener en claro los conceptos que engloba el machine learning, Como es de conocimiento el machine learning dispone de numerosos algoritmos algunos similares entre sí, es importante repasar los fundamentos de cada uno y analizar cuál sería el más óptimo para el estudio a realizar, existen un sin números de lenguajes y formas de programar, es importante que se escoja el que más cómodo se le haga al usuario, en este caso se hizo uso del lenguaje R pues es el que se usa para programar y ejecutar los modelos en la app R-studio.

Con el propósito de mejorar los modelos realizados y hacerlos más preciso el modelo mitigando el error se recomienda trabajar con una base de datos apropiada con la menor cantidad de outliers (valor atípico) o datos faltantes para que de esta manera no se tenga que recurrir a técnicas digitales como el “data cleaning” o trasponer los datos faltantes con la media de los valores de la variable, contar con la cantidad de datos necesaria es otro punto en el que se podría profundizar para un mejor resultado del modelo, a los entendidos de la materia sabrán que entre más data se disponga el resultado será más beneficioso.

Gracias a esta investigación se pudo demostrar la exactitud con la que los modelos clasificaron a los clientes y con que probabilidad se los valido a cada uno, se corroboró que la inteligencia artificial rara vez falla siendo esta no la excepción, la implementación de este análisis en empresas E-commerce sería fructífero debido a que permitiría ver hacia dónde va la tendencia de los clientes y anticiparse con descuentos por compras online o físicas, enfocarse en productos hacia cierto rango de edad, ver qué forma de pago deberían priorizar más, y un sin números de aspectos más que se pudieran implementar todo dependiendo de hacia dónde lo planeo orientar pero siguiendo el mismo principio.

Bibliografía

- Abbas, H., & Farooque, J. (27 de November de 2019). Reverse logistics practices in Indian pharmaceutical supply chains: a study of manufacturers. *Logistics Systems and Management*, 35(1), 72-89. Obtenido de <https://www.inderscienceonline.com/doi/pdf/10.1504/IJLSM.2020.103863>
- Anderson, M., Carvalho, P., & Fernández, L. (Marzo de 2020). Improving solar forecasting using Deep Learning and Portfolio Theory integration. *Energy*, 195. doi:10.1016/j.energy.2020.117016
- Assefa, S., Ejigu, A., & Namera, G. (Diciembre de 2021). A structural equation analysis of export marketing adaptation strategies on export performance: evidence from textile and garment exporting enterprises in Ethiopia. *Innovation*(67). Obtenido de https://www.researchgate.net/profile/Abebe-Alemu-2/publication/357810435_A_structural_equation_analysis_of_export_marketing_adaptation_strategies_on_export_performance_evidence_from_textile_and_garment_exporting_enterprises_in_Ethiopia/links/61e066a78d338
- Barnett, M., Irene, H., & Husted, B. (27 de January de 2020). Beyond Good Intentions: Designing CSR Initiatives for Greater Social Impact. *Journal of Management*, 46(6), 937-964. doi:10.1177/0149206319900539
- Benites, C., & Arteaga, E. (2022). *Aplicación de modelos de machine learning para clasificación de clientes en una empresa*. Tesis Doctoral, Universidad Central del Ecuador, Quito. Obtenido de <http://www.dspace.uce.edu.ec/bitstream/25000/27960/1/UCE-FCE-CPO-CONSUELO%20BENITES-EDGAR%20ARTEAGA.pdf>
- Bernal, C. (2020). *Metodología de la Investigación: administración, economía, humanidades y ciencias sociales* (Vol. III). Bogotá D.C.: Pearson Educación. Obtenido de <https://abacoenred.com/wp-content/uploads/2019/02/El-proyecto-de-investigaci%C3%B3n-F.G.-Arias-2012-pdf.pdf>
- Bo, X., Ayokunle, O., Qing, C., Linlin, X., & Yong, L. (10 de September de 2018). Conceptualising the state of the art of corporate social responsibility (CSR) in the construction industry and its nexus to sustainable development. *Journal of Clear*

Production, 195, 340-353. doi:10.1016/j.jclepro.2018.05.157

Caballero, R., Kebaier, A., Scavino, M., & Tempone, R. (29 de Junio de 2020). A Derivative Tracking Model for Wind Power Forecast Error. *King Abdullah Univerity Journal*. Obtenido de <https://repository.kaust.edu.sa/handle/10754/664091>

Carchi, I. (2019). Modelo de Negocios E-Commerce para Potenciar la Comercialización de los Spas en el Cantón Cuenca, Provincia del Azuay. (*Tesis de maestría*). Universidad Tecnológica Empresarial de Guayaquil, Guayaquil. Obtenido de <http://biblioteca.uteg.edu.ec:8080/bitstream/handle/123456789/999/Modelo%20de%20negocios%20E-Commerce%20para%20potenciar%20la%20comercialización%20de%20los%20Spas%20en%20el%20Cantón%20Cuenca%20C%20Provincia%20del%20Azuay.pdf?sequence=3&isAllowed=y>

Dogu, E., & Albayrak, E. (3 de Mayo de 2018). Criteria evaluation for pricing decisions in strategic marketing management using an intuitionistic cognitive map approach. *Soft Computing*, 22, 4989–5005. Obtenido de <https://link.springer.com/article/10.1007/s00500-018-3219-5>

Elbayoumi, I., Elkhwesky, Z., & Ramkissoon, H. (2021). A content analysis for government's and hotels' response to COVID-19 pandemic in Egypt. *Tourism and Hospitality Research*, 22(1), 42-59. doi:10.1177/14673584211002614

Estrella, J. (2023). *Análisis Situacional y Propuesta de Plan Estratégico de Marketing para el Área Comercial de la Empresa Alimentos Don Diego de la ciudad de Quito para el período 2014-2018*. Tesis, Universidad Técnica Particular de Loja, Área Administrativa, Loja. Obtenido de <https://dspace.utpl.edu.ec/bitstream/20.500.11962/21220/1/Estrella%20Caizagano%20Juan%20Carlos.pdf>

Faria, T., Neves, D., & Viera, M. (Mayo de 2018). Assessing the Agricultural Commodities Transportation in the State of Mato Grosso: Scenarios Destined to Exportation. *Revista Produção E Desenvolvimento*, IV(2), 31-41. doi:10.32358/rpd.2018.v4.282

Franco, S., Giuliano, M., Cappa, F., & Del Chiappa, G. (July de 2020). Are you good enough? CSR, quality management and corporate financial performance in the hospitality

- industry. *International Journal of Hospitality Management*, 88, 1-12.
doi:10.1016/j.ijhm.2019.102395
- García, A. (Diciembre de 2018). Algunas Aplicaciones del Diagnóstico de Rorschach en el
Ámbito Forense. *Asociación Argentina de Psicodiagnóstico de Rorschach*, XXXVIII(2),
17-23. Obtenido de <http://www.asoc-arg-rorschach.com.ar/revistas/Revista-Rorschach-Diciembre-2018.pdf#page=15>
- Gómez, M., & Ramos, E. (2011). *Las nuevas tecnologías de la información y el Ecommerce, algunas implicaciones regulatorias y éticas*. Obtenido de Universidad Juárez Autónoma de Tabasco:
https://upcommons.upc.edu/bitstream/handle/2099/13620/G%C3%B3mez_Lemus_Marbella_Araceli.pdf
- Gyenge, B., Mate, Z., & Vida, I. (25 de Marzo de 2023). A New Strategic Marketing Management Model for the Specificities of E-Commerce in the Supply Chain. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(4).
doi:10.3390/jtaer16040064
- Henriquez Rojas, J., & Aracena Pizarro, D. (2013). Segmentation of license plates through clustering techniques in external environments. Obtenido de https://www.ingeniare.cl/index.php?option=com_ingeniare&view=va&aid=124&vid=64&lang=en
- Hernández, C. (2018). *Técnicas de capacitación y su influencia en la productividad laboral en el sector de las pequeñas y medianas empresas*. Obtenido de Universidad Privada del Norte de Perú:
<https://repositorio.upn.edu.pe/bitstream/handle/11537/22136/Hernandez%20Bruno%20c%20Carmen%20Amelia%20.pdf?sequence=2&isAllowed=y>
- Jastrzebski, S., Szymczak, M., & Fort, S. (21 de Febrero de 2020). The Break-Even Point on Optimization Trajectories of Deep Neural Networks. *Machine Learning*.
doi:10.48550/arXiv.2002.09572
- Joaquín Abellán García, F. A. (2022). Modelo predictivo de redes neuronales para estimar la resistencia a compresión de hormigones con materiales cementantes suplementarios y agregados reciclados. *Scielo*.

- Khuntia, S., & Pattanayak. (13 de March de 2018). Adaptive market hypothesis and evolving predictability of bitcoin. *Economics Letters*, 167, 26-28.
doi:10.1016/j.econlet.2018.03.005
- Kirchebner, J., Sonnweber, M., Nater, U., & Günther, M. L. (January de 2022). Stress, Schizophrenia, and Violence: A Machine Learning Approach. *Journal of Interpersonal Violence*, XXXVII(2), 602-622. doi:10.1177/088626052091364
- Kumar, D. (2018). Application of Statistics in HR Research. *Statistical Tools and Analysis in Human Resources Management*, 1-34. doi:10.4018/978-1-5225-4947-5.ch005
- Kumar, R., Raiswa, S., Susoban, G., & Sekar, R. (25 de October de 2022). Consumer's response to CSR activities: Mediating role of brand image and brand attitude. *Corporate Social Responsibility and Environmental Management*, 26(2), 377-387. doi:10.1002/csr.1689
- Marlowe, Ho, Carey, & Chadick. (2020). Employing standardized risk assessment in pretrial release decisions: Association with criminal justice outcomes and racial equity. *Law and Human Behavior*, 44(5), 361–376. doi:10.1037/lhb0000413
- Mendieta, M., Erazo, J., & Narváez, C. (2020). Gestión por competencias: herramienta clave para el rendimiento laboral del talento humano del sector hospitalario. *Revista Arbitrada Interdisciplinaria Koinonía*, 5(10), 287-304. doi:10.35381/r.k.v5i10.696
- Molina, M., Pascual, P. M., Martínez, A., Fernández, M., Pretus, S., Armelles, E., González, C. (25 de May de 2016). *Una relación simple. Modelos de regresión simple*. Recuperado el 10 de November de 2022, de AnestesiaR:<https://anestesiario.org/2016/una-relacion-simple-modelos-regresion-simple/>
- Orellana, I., Pinos, L., Tonon, L., Reyes, M., & Cevallos, E. (Junio de 2022). Analysis of business closure in the manufacturing sector of Ecuador, period 1901 2021. *Ecos de Economía*, 24(50), 45-79. doi:10.17230/ecos.2020.50.3
- Orus, A. (17 de Abril de 2023). *Statista*. Obtenido de <https://es.statista.com/temas/9072/comercio-electronico-en-el-mundo/#editorsPicks>
- Pedrero, V., Reynaldos-Grandón, K., Ureta-Achurra, J., & Cortez-Pinto, E. (2021). Overview of machine learning and its application in the management of emergency services. *Revista medica de chile*.

Poirier, J., & Gabillaud, J. (2019). SQL Server 2019. ENI.

Regiani, E., Tavana, M., Schwengber, C., Bouzon, M., & Carísio, S. (18 de November de 2020).

A grey-DEMATEL approach for analyzing factors critical to the implementation of reverse logistics in the pharmaceutical care process. *Environmental Science and Pollution Research*, 28, 14156–14176. Obtenido de <https://link.springer.com/article/10.1007/s11356-020-11138-8>

Rodríguez, G. (2022). Método de algoritmo de clúster para el análisis del perfil. *Método de algoritmo de clúster para el análisis del perfil*, 12(2), 160-184. doi:<https://doi.org/10.15517/eci.v12i2.50456>

Rojas, C., Medina Quispe, & Meneses Villegas. (2014). Modelo aumentado de árbol de decisión utilizando mapas autoorganizados. *Scielo*.

Sánchez, S., & Villamarín, E. (2021). *Aplicación web basado en machine learning para el fortalecimiento de la gestión y logística de servicio a domicilio en la empresa MobilFoods de la provincia de Santo Domingo de los Tsáchilas*. Tesis de Postgrado, Pontificia Universidad Católica del Ecuador Sede Santo Domingo. Obtenido de https://issuu.com/pucesd/docs/4434-_4446sandra_patricia_s_nchez_ram_rez

Thomas, J., Santos, M., Mora, M., & Duro, M. (2019). Performance Analysis of Clustering Internal. *IEEE LATIN AMERICA TRANSACTIONS*.

Yi, L., Changfu, Z., Maitane, B., Nanini-Maury, E., C.-W. Chan, J., Van den Bossche, P. Noshin, O. (2018). Random forest regression for online capacity estimation of lithium-ion. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0306261918315010>

Anexos

Anexo A. Script completo en R-Studio.

```
Clientes <- read.csv2("../DATA/BASE TESIS - PRUEBA 299OBS.csv")

#Se crea un nuevo dataset para el kmeans
Clientes1 <- Clientes[,c(-1,-2,-3,-4)]

library(tidyverse)
library(cluster)
library(factoextra)
library(NbClust)
library(tidyr)

#Se escala los datos
Clientes1Esc <- scale(Clientes1)

#Se calcula el número óptimo de clústeres
resnumclust<-NbClust(Clientes1Esc, distance = "euclidean", min.nc=2, max.nc=10, method =
"kmeans", index = "alllong")
NCluster=NbClust(Clientes1Esc,distance="euclidean",min.nc=2,max.nc=15,method="kmeans")
#Calculamos el K-means con un centro de 3
k3 <- kmeans(Clientes1Esc, centers = 3, nstart = 25)
k3

head(k3)
#Representamos gráficamente el K-means
fviz_cluster(k3, data = Clientes1Esc)
fviz_cluster(k3, data = Clientes1Esc, ellipse.type = "euclid",repel = TRUE,star.plot = TRUE)
#ellipse.type= "t", "norm", "euclid"
fviz_cluster(k3, data = Clientes1Esc, ellipse.type = "norm")
```

```
#Pasamos los resultados a la base original
```

```
Clientes$Kmeans <- k3$cluster
```

```
#Etiquetamos los clústeres por clasificación
```

```
Clientes$Kmeans <- factor(Clientes$Kmeans,
```

```
  levels = c(1,2,3),
```

```
  labels = c("Bajo","Medio",
```

```
    "Alto"))
```

```
Clientes2 <- Clientes[,c(-1,-2,-3)]
```

```
#Calculamos un árbol de clasificación
```

```
library(caret)
```

```
library(rpart)
```

```
library(rpart.plot)
```

```
library(randomForest)
```

```
entrenamiento <- createDataPartition(Clientes2$Kmeans, p= 0.8, list = F)
```

```
#Paso 3. Creamos el modelo
```

```
Clientes2Ar <- rpart(Kmeans~., data = Clientes2)
```

```
Clientes2Ar
```

```
#Paso 4. Graficamos el árbol
```

```
rpart.plot(Clientes2Ar, type=1, digits = -1, extra= 0, cex = 1, nn=TRUE, fallen.leaves = TRUE)
```

```
#Paso 5. Creamos el modelo de pronóstico
```

```
Clientes2Rf <- randomForest(x= Clientes2[entrenamiento, 2:8], y =Clientes2[entrenamiento,  
9],
```

```
  xtest = Clientes2[-entrenamiento, 2:8], ytest = Clientes2[-entrenamiento,9],
```

```
  ntree = 20000, importance=T,keep.forest=T)
```

```
Clientes2Rf
```

#Paso 5. Predecimos con qué probabilidad se asigna cada clúster

```
Cientes2$Probabilidad <- predict(Cientes2Rf,Cientes2,type = "prob")
```

Anexo B. Summary de los resultados.

```
*** : The Hubert index is a graphical method of determining the number of clusters.  
      In the plot of Hubert index, we seek a significant knee that corresponds to a  
      significant increase of the value of the measure i.e the significant peak in Hubert  
      index second differences plot.
```

```
*** : The D index is a graphical method of determining the number of clusters.  
      In the plot of D index, we seek a significant knee (the significant peak in Dindex  
      second differences plot) that corresponds to a significant increase of the value of  
      the measure.
```

```
*****
```

```
* Among all indices:  
* 5 proposed 2 as the best number of clusters  
* 7 proposed 3 as the best number of clusters  
* 1 proposed 4 as the best number of clusters  
* 7 proposed 6 as the best number of clusters  
* 6 proposed 9 as the best number of clusters  
* 1 proposed 10 as the best number of clusters
```

```
***** Conclusion *****
```

```
* According to the majority rule, the best number of clusters is 3
```

```
K-means clustering with 3 clusters of sizes 217, 65, 17
```

```
Cluster means:
```

	EDAD	Payment.System.Name	Address.Type	Coupon	Shipping.Value	X.Transacciones
1	0.10622370	-0.02544839	0.54680587	0.006820253	-0.2977205	-0.1318960
2	-0.08586957	0.01432132	-1.82268622	-0.047632020	0.5941858	-0.1763081
3	-1.02758945	0.27008322	-0.01072168	0.095063905	1.5284283	2.3577333

Ventas.Totales.

1	-0.2156583
2	-0.2189954
3	3.5901500

```
Clustering vector:
```

```
[1] 1 1 2 1 1 1 1 1 1 2 2 2 1 1 2 3 1 2 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 3 3 1 1 1 1 1 2 2 1 2 2 1 1 1  
[50] 1 2 1 1 2 2 1 1 1 2 1 3 2 2 1 1 1 2 3 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 2 1 1 2 3 1  
[99] 1 1 1 2 3 1 1 2 1 1 2 1 1 1 1 1 1 2 1 3 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1  
[148] 1 1 1 2 1 2 1 1 1 2 2 1 1 1 1 1 2 1 1 1 1 3 1 2 3 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1  
[197] 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 3 1 2  
[246] 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1 2 2 1 1 1 2 1 2 1 1 1 1 2 2 1 1 1 2 1 2 1 2 2  
[295] 1 1 1 2 1
```

```
within cluster sum of squares by cluster:
```

```
[1] 743.7188 294.6480 329.7038  
(between_SS / total_SS = 34.4 %)
```

```
Available components:
```

```
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"  
[7] "size"         "iter"         "ifault"
```



```

> head(Clientes2)
  GEN EDAD Payment.System.Name Address.Type Coupon Shipping.Value X.Transacciones Ventas.Totales.
1 MASCULINO 42 4 1 0 0 0 1 169
2 MASCULINO 29 4 1 0 0 0 2 1418
3 MASCULINO 23 4 0 0 0 0 1 399
4 MASCULINO 37 1 1 0 0 0 1 188
5 MASCULINO 29 1 1 0 0 0 1 134
6 MASCULINO 25 4 1 0 0 0 2 1520
Kmeans Probabilidad.Bajo Probabilidad.Medio Probabilidad.Alto
1 Bajo 0.99970 0.00030 0.00000
2 Bajo 0.98070 0.00355 0.01575
3 Medio 0.06490 0.93425 0.00085
4 Bajo 0.99655 0.00345 0.00000
5 Bajo 0.99930 0.00070 0.00000
6 Bajo 0.93690 0.01315 0.04995
>

```

Anexo C. Base de datos con aplicación de K-Means.

	X	Client. Document	Nombre	GEN	EDA D	Payment .System. Name	Addr ess.T ype	Cou pon	Shipp ing.V alue	X.Tra nsacci ones	Vent as.To tales.	Kme ans
1	1	12071 01229	ABEL BARRAGAN DIAZ	MAS CULI NO	42	4	1	0	0	1	169	Bajo
2	2	95012 6433	Abraham Jurado	MAS CULI NO	29	4	1	0	0	2	1418	Bajo
3	3	92522 1376	Abraham Muñoz	MAS CULI NO	23	4	0	0	0	1	399	Me dio
4	4	13120 82314	adolfo donoso	MAS CULI NO	37	1	1	0	0	1	188	Bajo
5	5	94083 3569	Adrian fabricio Ibarra del valle	MAS CULI NO	29	1	1	0	0	1	134	Bajo
6	6	94270 9775	Adrian Pin	MAS CULI NO	25	4	1	0	0	2	1520	Bajo
7	7	92866 3681	ADRIANA RODRIGUEZ PILAY	FEM ENI NO	36	4	1	0	0	1	184	Bajo
8	8	13083 50899	ALEIDA MARIA SOLORZANO ARTEAGA	FEM ENI NO	27	3	1	0	0	2	567	Bajo
9	9	10467 7604	Alex Amador Sanchez Morales	MAS CULI NO	25	1	1	0	0	2	2699	Bajo
10	10	93130 6260	Alex Buenas Cuvi	MAS CULI NO	47	3	1	0	0	1	149	Bajo
11	11	10504 80589	Alex Espinosa	MAS CULI NO	50	1	0	0	23	1	157	Me dio
12	12	80366 0133	alex vaque	MAS CULI NO	36	4	0	0	5	1	84	Me dio

13	13	95406 2188	ALEXANDER JOSUE VERA VÉLEZ	MAS CULI NO	50	1	0	0	3.5	1	43	Me dio
14	14	12065 42589	Alexandra del cisne Ruiz chica	FEM ENI NO	35	4	1	0	0	2	3014	Bajo
15	15	17599 23244	Alexandra Piedra	FEM ENI NO	46	4	1	0	0	1	13	Bajo
16	16	60590 1099	Alexis Joel Velata López	MAS CULI NO	39	4	0	0	4.5	1	84	Me dio
17	17	70457 3583	alisson castro	FEM ENI NO	21	4	1	0	5	2	7864	Alto
18	18	12061 65464	Allison Haily Putan Jimenez	FEM ENI NO	36	4	1	0	0	1	80	Bajo
19	19	91741 9442	Anabel Jimenez	FEM ENI NO	37	4	0	0	0	1	79	Me dio
20	20	17196 13992	Anderson Herrera	MAS CULI NO	49	4	1	0	0	1	39	Bajo
21	21	13152 78513	Andrés Anzules	MAS CULI NO	46	1	1	0	0	1	40	Bajo
22	22	20000 31563	Andres Guerrero	MAS CULI NO	34	4	0	0	4.5	1	30	Me dio
23	23	17191 82188	ANDRES VILLARREAL	MAS CULI NO	42	4	1	0	0	1	229	Bajo
24	24	92947 5135	andy andrade	MAS CULI NO	48	3	1	0	0	1	149	Bajo
25	25	93248 0619	Anette Vasquez Tamayo	FEM ENI NO	40	4	0	0	8	1	97	Me dio
26	26	92993 0428	ANGEL CHOEZ	MAS CULI NO	39	3	1	0	0	1	259	Bajo
27	27	18014 86851	angel zamora	MAS CULI NO	46	3	1	0	0	1	59	Bajo
28	28	13116 92485	Angela Cusme Zambrano	FEM ENI NO	33	1	1	0	0	1	121	Bajo
29	29	17224 16748	Angela Naula	FEM ENI NO	49	4	1	0	0	1	39	Bajo

30	30	12044 75378	Angela Sanchez	FEM ENI NO	23	4	1	1	0	1	302	Bajo
31	31	91799 2950	ANGELICA MORAN	FEM ENI NO	24	1	1	0	0	2	416	Bajo
32	32	70399 6603	ANGELITA MAGALI CAMINOS DAVILA	FEM ENI NO	40	4	0	0	10	1	180	Me dio
33	33	92490 6175	Angie Guerrero	FEM ENI NO	31	4	1	1	0	1	5	Bajo
34	34	92967 4091	Angioly Michel Vera Magallanes	MAS CULI NO	21	4	1	0	0	1	112	Bajo
35	35	95430 1156	ANTHONY BAQUE PEREZ	MAS CULI NO	25	3	1	0	0	4	338	Alto
36	36	92393 9185	Anthony del	MAS CULI NO	23	4	0	0	70	4	1690	Alto
37	37	17169 78844	Antonio Gruezo	MAS CULI NO	25	4	1	0	0	1	15	Bajo
38	38	11037 83542	Argenys Valdiviezo	MAS CULI NO	34	1	1	0	0	1	309	Bajo
39	39	10084 6237	Ariosto trelles cabrera	MAS CULI NO	23	1	1	0	0	1	99	Bajo
40	40	95165 0472	arlet moran	FEM ENI NO	34	4	1	0	0	1	99	Bajo
41	41	94292 7005	Axel Orrala Ramirez	MAS CULI NO	21	1	1	0	0	1	107	Bajo
42	42	90603 4129	Beatriz Arellano Castillo	FEM ENI NO	38	4	0	0	3.5	1	66	Me dio
43	43	13127 71387	BELEN TORO	FEM ENI NO	50	4	0	0	0	1	159	Me dio
44	44	92663 6481	BERNARDO JOSUE ENCALADA ESCOBAR	MAS CULI NO	37	4	1	0	0	1	79	Bajo
45	45	93096 4721	Betsy Salazar	FEM ENI NO	44	4	0	0	3.5	1	93	Me dio

46	46	70638 9640	BRYAN GAMBOA	MAS CULI NO	42	4	0	0	5	1	65	Me dio
47	47	21004 10972	Bryan Jaramillo Calderon	MAS CULI NO	21	3	1	0	0	1	499	Bajo
48	48	92698 9427	Byron David Coloma Fuentes	MAS CULI NO	37	4	1	0	0	1	269	Bajo
49	49	92971 9417	Byron ismael Hernández gomez	MAS CULI NO	45	1	1	0	0	1	228	Bajo
50	50	92969 9288	CAMACHO ANDRADE LUIS ENRRIQUE	MAS CULI NO	30	3	1	0	0	1	299	Bajo
51	51	91880 1119	Carla Lorena Paladines Martinez	FEM ENI NO	43	4	0	0	0	1	69	Me dio
52	52	93069 8501	Carla Mina	FEM ENI NO	25	1	1	0	0	1	134	Bajo
53	53	12075 53387	Carlos Caregua	MAS CULI NO	31	1	1	0	0	2	724	Bajo
54	54	92680 4675	Carlos Alvia	MAS CULI NO	32	1	0	0	4.5	1	422	Me dio
55	55	92508 1952	CARLOS ANTONIO JORDAN ORRALA	MAS CULI NO	23	1	0	0	0	1	497	Me dio
56	56	90793 2875	CARLOS ARTURO PATIÑO FLORES	MAS CULI NO	21	4	1	0	0	1	149	Bajo
57	57	11049 94346	carlos Jaramillo	MAS CULI NO	40	4	1	0	0	1	20	Bajo
58	58	80386 2382	Carlos Jerez	MAS CULI NO	50	3	1	0	0	1	149	Bajo
59	59	17133 81794	Carlos Lenin Castillo Salgado	MAS CULI NO	29	4	0	0	3.5	1	403	Me dio
60	60	95784 8401	carlos macias	MAS CULI NO	49	3	1	0	0	1	159	Bajo
61	61	92671 0120	Carlos Vera	MAS CULI NO	35	4	1	0	0	2	6234	Alto

62	62	10021 58507	Cecilia Duarte	FEM ENI NO	27	4	0	0	0	1	331	Me dio
63	63	92109 9214	Celeni Serrano Toala	FEM ENI NO	21	1	0	0	14.5	1	483	Me dio
64	64	91877 1080	César Augusto Cedeño galarza	MAS CULI NO	41	4	1	0	0	1	69	Bajo
65	65	17240 84320	CHIPANTASIG Edwin	MAS CULI NO	32	3	1	0	0	1	319	Bajo
66	66	92707 5903	Christian Olaya soriano	MAS CULI NO	50	3	1	0	0	1	45	Bajo
67	67	70508 0521	Cindy Elizabeth Andrade Santana	FEM ENI NO	22	1	0	0	20	1	718	Me dio
68	68	91848 5756	Claudia Illescas Laines	FEM ENI NO	20	4	0	1	30	2	8676	Alto
69	69	11025 32833	Clemencia Guarnizo	FEM ENI NO	23	1	0	0	5	1	544	Me dio
70	70	94034 5176	Cristhan Chipre	MAS CULI NO	36	2	1	1	0	2	77	Bajo
71	71	92941 4225	Cristhian Alava	MAS CULI NO	42	4	0	0	3.5	1	243	Me dio
72	72	94420 1011	CRISTHOFER ALFONSO VEGA	MAS CULI NO	27	1	1	0	8	2	1292	Bajo
73	73	23500 26700	Cristian Abad	MAS CULI NO	22	1	1	0	0	1	323	Bajo
74	74	91980 9400	Cristian garcia	MAS CULI NO	28	4	1	0	0	1	318	Bajo
75	75	90583 5864	Cristina Franco	FEM ENI NO	45	4	1	0	0	1	189	Bajo
76	76	17157 32150	Cristina Rhon Martin	FEM ENI NO	23	4	1	0	0	1	99	Bajo
77	77	16004 15937	CRISTINA ZAMBRANO	FEM ENI NO	45	3	1	0	0	1	179	Bajo

78	78	92456 4479	CRISTOPHER PAREDES	MAS CULI NO	35	1	1	0	0	1	350	Bajo
79	79	11051 15529	Damaris Garcia Moreno	FEM ENI NO	38	4	0	0	23	1	172	Me dio
80	80	90658 0931	Daniel Urgelles Falconi	MAS CULI NO	38	4	1	0	0	1	229	Bajo
81	81	10034 16409	Daniela Caluguillin	FEM ENI NO	49	3	1	0	0	1	149	Bajo
82	82	17293 56756	Daniela Pinto Rodriguez	FEM ENI NO	30	4	1	0	0	1	119	Bajo
83	83	95349 6437	Danny Loor	MAS CULI NO	50	3	1	0	0	1	49	Bajo
84	84	75028 7476	DARIO ANDRES MEDINA RODRIGUEZ	MAS CULI NO	50	3	1	0	0	1	149	Bajo
85	85	93112 7211	Dario Chang	MAS CULI NO	27	4	1	0	0	2	398	Bajo
86	86	17274 43267	Dario Espinosa	MAS CULI NO	30	4	1	0	0	1	379	Bajo
87	87	21005 55966	Dario garcia	MAS CULI NO	22	4	1	0	0	1	129	Bajo
88	88	13130 13920	DARIO VITE	MAS CULI NO	38	4	1	0	0	1	279	Bajo
89	89	17249 69819	Darwin Zambrano	MAS CULI NO	27	3	1	0	0	1	140	Bajo
90	90	13148 53530	david alcivar	MAS CULI NO	40	1	1	0	0	1	94	Bajo
91	91	10446 6909	David mosquera	MAS CULI NO	40	4	1	0	0	1	229	Bajo
92	92	17218 88236	David Rosero	MAS CULI NO	23	4	1	0	6	2	7980	Alto
93	93	15005 75731	denice ledesma	FEM ENI NO	22	4	0	0	0	1	298	Me dio

94	94	17279 23607	Denis Tituaña	MAS CULI NO	35	3	1	0	0	1	300	Bajo
95	95	17164 88026	DEXY DIAZ	FEM ENI NO	48	3	1	0	0	1	149	Bajo
96	96	95270 3114	Diana Ortiz	FEM ENI NO	27	1	0	0	10	1	549	Me dio
97	97	92008 9539	diana rodriguez	FEM ENI NO	32	4	1	0	0	2	9969	Alto
98	98	95278 3389	Diana Suárez	FEM ENI NO	32	3	1	0	0	1	499	Bajo
99	99	92671 6077	Diego Perez sarango	MAS CULI NO	44	4	1	0	0	1	199	Bajo
100	100	95158 4978	Diego Aguirre	MAS CULI NO	43	1	1	0	0	1	188	Bajo
101	101	17222 97593	DIEGO ORTIZ LOAQUIZA	MAS CULI NO	38	1	1	0	0	1	239	Bajo
102	102	92582 7768	Dieguito15 Acosta Zamora	MAS CULI NO	27	4	0	0	3.5	2	462	Me dio
103	103	92438 4332	digna trinidad ayala abad	FEM ENI NO	26	4	1	0	10	2	4574	Alto
104	104	70242 4391	DINO YANEZ MOROCHO	MAS CULI NO	45	4	1	1	0	1	66	Bajo
105	105	12038 08819	Doriam lascano vera	MAS CULI NO	21	4	1	0	0	1	20	Bajo
106	106	12029 24732	Doris Campi Fabre	FEM ENI NO	39	4	0	0	0	1	89	Me dio
107	107	13132 12662	DORIS ELIZABETH SOLORZANO MOREIRA	FEM ENI NO	43	4	1	0	0	1	174	Bajo
108	108	13505 90244	DUVANY DELGADO	MAS CULI NO	48	3	1	0	0	1	149	Bajo
109	109	17149 90015	Edgar Andres Vinueza	MAS CULI NO	26	4	0	0	7.5	1	807	Me dio

110	110	17172 08860	Edgar Giovanny Armijos Bravo	MAS CULI NO	44	4	1	0	0	1	89	Bajo
111	111	94268 0356	Edhison Yagual Vega	MAS CULI NO	35	4	1	0	0	1	19	Bajo
112	112	24500 37052	EDISON GUALE	MAS CULI NO	44	3	1	0	0	1	169	Bajo
113	113	13145 25229	Edison Rodríguez Baque	MAS CULI NO	41	4	1	0	0	1	269	Bajo
114	114	60467 9696	Edison Arias	MAS CULI NO	34	4	1	0	0	1	149	Bajo
115	115	17245 15919	Edison Mauricio Tumbaco Toapanta	MAS CULI NO	37	4	0	0	0	1	279	Me dio
116	116	94039 8480	Edison Morales	MAS CULI NO	33	4	1	0	0	1	348	Bajo
117	117	92905 3262	Edison Roberto Vargas vera	MAS CULI NO	28	4	1	0	4.5	2	8914	Alto
118	118	91641 7678	Edith Acuña	FEM ENI NO	27	4	1	0	0	2	379	Bajo
119	119	94206 5004	Edson Pazmiño Mutre	MAS CULI NO	33	4	1	0	0	1	299	Bajo
120	120	50321 3613	Eduardo Andrango	MAS CULI NO	27	4	1	0	0	1	142	Bajo
121	121	19005 61620	Eduardo Aníbal Alarcón Marzilla	MAS CULI NO	26	4	1	0	0	1	149	Bajo
122	122	17253 56610	Edwin Gavilánes Damian	MAS CULI NO	33	1	1	0	0	1	134	Bajo
123	123	50287 0371	Edwin Chuquitarco Delgado	MAS CULI NO	26	4	1	0	0	1	489	Bajo
124	124	13052 89553	Edwin Ponce Minaya	MAS CULI NO	28	4	1	0	0	1	99	Bajo
125	125	91105 3981	efren roberto preciado cheme	MAS CULI NO	49	4	0	0	3.5	1	41	Me dio

126	126	18051 11539	Elian paul Castro brito	MAS CULI NO	36	1	1	0	0	1	67	Bajo
127	127	91989 9641	Eliana Yomayra Calderón Orozco	FEM ENI NO	25	4	1	0	0	1	137	Bajo
128	128	92445 5645	Elizabeth Pinto	FEM ENI NO	46	4	1	0	0	1	42	Bajo
129	129	90351 4347	ELVIA HERRERA	FEM ENI NO	49	4	1	0	0	1	50	Bajo
130	130	70629 2323	ELVIS JESUS DIAZ CARDENAS	MAS CULI NO	44	4	1	0	0	1	269	Bajo
131	131	91937 9776	Elvis Lindao Muñoz	MAS CULI NO	21	3	1	0	0	2	6916	Alto
132	132	60340 2660	Emma Inés Granizo Espinoza	FEM ENI NO	40	4	1	0	0	1	12	Bajo
133	133	92024 5776	Emmanuel Moran	MAS CULI NO	28	3	1	0	0	1	499	Bajo
134	134	95779 7590	Erick Anchundia	MAS CULI NO	25	1	1	0	0	1	364	Bajo
135	135	92677 6428	Erick Suarez	MAS CULI NO	49	3	1	0	0	1	149	Bajo
136	136	17256 75266	Erika Panizo	FEM ENI NO	38	1	1	0	0	1	242	Bajo
137	137	10044 81576	Erika Castro	FEM ENI NO	48	3	1	0	0	1	149	Bajo
138	138	17260 49578	Esteban Proaño Mendieta	MAS CULI NO	45	1	1	0	0	1	66	Bajo
139	139	10399 0909	Eugenio Cabrera	MAS CULI NO	34	4	1	0	0	2	114	Bajo
140	140	13112 75828	Evelyn Gissela Aspiazu Macias	FEM ENI NO	38	4	1	0	0	1	179	Bajo
141	141	70522 0705	Fabian Gallardo Loaiza	MAS CULI NO	32	4	1	0	0	1	17	Bajo

142	142	95084 9125	Fabricio Ponce	MAS CULI NO	36	4	1	0	0	1	69	Bajo
143	143	91439 8193	Fanny Solis Carbo	FEM ENI NO	22	4	0	0	9.5	1	103	Me dio
144	144	13987 6833	Felipe Antonio Avila tuarez	MAS CULI NO	50	4	1	0	0	1	47	Bajo
145	145	94216 9020	felix quinto	MAS CULI NO	28	4	1	0	0	1	299	Bajo
146	146	80258 3617	Fernanda Chica Intriago	FEM ENI NO	37	4	0	0	0	1	79	Me dio
147	147	91921 7398	Fernando Antonio AGUAS ASINC	MAS CULI NO	47	4	1	1	0	2	36	Bajo
148	148	95381 5271	Fernando Minaya	MAS CULI NO	33	3	1	0	0	3	717	Bajo
149	149	70394 4033	Fernando neptali González macaa	MAS CULI NO	39	4	1	0	0	1	188	Bajo
150	150	17165 65559	Fernando Trujillo Carrillo	MAS CULI NO	32	1	1	0	0	2	484	Bajo
151	151	17115 10766	Fernando Villegas Barros	MAS CULI NO	22	4	0	0	0	1	149	Me dio
152	152	93043 6647	Fernando Yperty	MAS CULI NO	44	1	1	0	0	1	255	Bajo
153	153	95144 3357	Fiorella Torres Jivaja	FEM ENI NO	47	4	0	1	0	1	151	Me dio
154	154	92009 6815	Florentino Alvarez	MAS CULI NO	48	4	1	0	0	1	32	Bajo
155	155	70479 2944	Francisco CASTANEDA	MAS CULI NO	42	4	1	0	0	1	177	Bajo
156	156	17198 21520	Francisco Ocampo	MAS CULI NO	32	4	1	0	0	1	379	Bajo
157	157	92388 5487	Francisco Salvatierra Vargas	MAS CULI NO	38	4	0	0	9	1	219	Me dio

158	158	13126 62271	Francisco Vera Coox	MAS CULI NO	33	4	0	0	7	1	396	Me dio
159	159	13150 57685	Franklin Arturo Ortiz Mejia	MAS CULI NO	36	3	1	0	0	1	189	Bajo
160	160	13002 18243	FRESSIA POGGI	FEM ENI NO	40	4	1	0	0	1	179	Bajo
161	161	17205 14981	Frixion Ricardo Vera Zambrano	MAS CULI NO	40	4	1	0	0	1	61	Bajo
162	162	92892 4877	GABRIEL TUMBACO	MAS CULI NO	49	3	1	0	0	1	152	Bajo
163	163	17142 55377	Gabriela Díaz	FEM ENI NO	37	4	1	0	0	2	174	Bajo
164	164	91556 5519	Galo Crespo Martínez	MAS CULI NO	28	4	0	0	3.5	1	103	Me dio
165	165	22000 78067	Gary Alfredo Galeas Castillo	MAS CULI NO	30	4	1	0	0	1	129	Bajo
166	166	13099 10774	Gema Alava Aray	FEM ENI NO	39	4	1	0	0	1	78	Bajo
167	167	40187 0902	Génesis Cangas	FEM ENI NO	23	1	1	0	0	1	379	Bajo
168	168	92175 9130	Geomayra Morales Navas	FEM ENI NO	50	4	1	0	0	1	50	Bajo
169	169	12030 98254	GEORGINA ZAMORA VARAS	FEM ENI NO	30	4	1	0	0	2	338	Bajo
170	170	91707 9741	Geovanna Bolla Acosta	FEM ENI NO	33	1	1	0	0	2	8711. 9	Alto
171	171	17585 38662	Gerardo Cabrita	MAS CULI NO	47	4	1	0	0	1	11	Bajo
172	172	17080 74131	Gerardo campaña	MAS CULI NO	39	4	0	0	4.5	1	84	Me dio
173	173	12048 64472	Glenda Liliana Medina Ricacho	FEM ENI NO	28	4	0	0	35	2	5882	Alto
174	174	91332 3770	GLENDA MARIBEL	FEM ENI NO	28	1	0	0	10	1	387	Me dio

191	191	80314 2678	ISAAC BRIONES	MAS CULI NO	33	4	1	0	0	1	344	Bajo
192	192	17597 90494	Isaac Ramirez	MAS CULI NO	25	1	1	0	0	2	1391. 13	Bajo
193	193	13056 08919	Isabel de los Angeles Suárez Cobeña	FEM ENI NO	20	1	1	0	0	1	462	Bajo
194	194	91312 2248	Ismael Figueroa Parra	MAS CULI NO	20	4	0	0	5.25	2	462	Me dio
195	195	92077 3058	Ivette Miranda Salazar	FEM ENI NO	47	4	0	0	3.5	1	33	Me dio
196	196	92997 9904	Jacinto Alexander Cedeño López	MAS CULI NO	35	4	1	0	0	1	6	Bajo
197	197	95215 5133	Jair Alexander Arias Loor	MAS CULI NO	24	4	1	0	0	1	31	Bajo
198	198	13132 04545	Jair Sanchez Bravo	MAS CULI NO	43	4	1	0	0	1	259	Bajo
199	199	13511 78015	jandry cornejo	MAS CULI NO	48	3	1	0	0	1	149	Bajo
200	200	92043 1046	Jasmin Ortiz Mera	FEM ENI NO	23	4	1	0	0	1	120	Bajo
201	201	91504 4028	Javier Ruben León Naranjo	MAS CULI NO	24	4	0	0	13.5	2	7093	Alto
202	202	95317 2921	JAYLENE PEÑAFIEL	FEM ENI NO	29	3	1	0	0	1	398	Bajo
203	203	94102 9332	Jean Carlos Alvarado	MAS CULI NO	27	3	1	0	0	1	129	Bajo
204	204	12071 68053	Jefferson Parraga	MAS CULI NO	45	3	1	0	0	1	269	Bajo
205	205	13132 60240	jefferson Cabeza	MAS CULI NO	31	3	1	0	0	1	369	Bajo
206	206	12502 69972	jeffre valentin ruiz urrutia	MAS CULI NO	30	4	1	0	0	1	288	Bajo
207	207	17233 27092	Jennifer Jama	FEM ENI NO	42	3	1	0	0	1	179	Bajo

208	208	12059 23376	JENNIFER León	FEM ENI NO	46	4	1	0	0	1	45	Bajo
209	209	92294 8880	JENNIFER LILIAN CHIPRE VELASCO	FEM ENI NO	40	3	1	0	0	3	226	Bajo
210	210	17231 51138	Jennifer Osorio Moreira	FEM ENI NO	35	4	1	0	0	1	299	Bajo
211	211	96177 1177	jennifer ponce	FEM ENI NO	21	4	1	0	0	1	99	Bajo
212	212	92584 3492	jennifer velasco	FEM ENI NO	34	1	0	0	0	1	350	Me dio
213	213	91282 5841	JENNY ALVARADO	FEM ENI NO	49	3	1	0	0	1	149	Bajo
214	214	91657 4098	jenny zuñiga	FEM ENI NO	33	1	1	1	0	1	319	Bajo
215	215	95545 7726	Jericko Villacis Bazurto	MAS CULI NO	36	3	1	0	0	1	259	Bajo
216	216	23001 92305	Jessenia Pereira	FEM ENI NO	27	3	1	0	0	1	149	Bajo
217	217	13087 96885	Jessica Cornejo	FEM ENI NO	24	4	1	0	0	1	349	Bajo
218	218	17197 08487	Jessica Guaranda	FEM ENI NO	25	4	1	0	0	1	148	Bajo
219	219	92077 6945	Jesus Camba	MAS CULI NO	35	1	1	0	0	1	485	Bajo
220	220	96516 9162	Jesús mauricio Cañas blandon	MAS CULI NO	21	1	0	0	0	1	498	Me dio
221	221	17596 24974	Jheniffer Lara	FEM ENI NO	35	4	1	0	0	1	928	Bajo
222	222	92426 1910	jhon jacome capa	MAS CULI NO	27	1	1	0	0	1	134	Bajo
223	223	14008 59185	Jhonatan fernando Salinas rodriguez	MAS CULI NO	43	4	1	0	0	1	229	Bajo

224	224	17250 98121	Jhonatan Tisalema	MAS CULI NO	24	3	1	0	0	1	99	Bajo
225	225	21007 80200	Jhonny Moreira	MAS CULI NO	36	4	1	0	0	1	69	Bajo
226	226	17240 77902	Jimmy Javier Logaña Cadena	MAS CULI NO	43	4	1	0	0	1	193	Bajo
227	227	91832 5838	Jimmy Morales	MAS CULI NO	20	4	1	0	0	3	4882. 5	Alto
228	228	21005 71435	Joffre Damian Zambrano Camacho	MAS CULI NO	36	1	1	0	0	2	244	Bajo
229	229	92709 4987	joffre Escalante	MAS CULI NO	49	3	1	0	0	3	50	Bajo
230	230	70523 1355	johana chuquimarca	FEM ENI NO	39	4	1	1	0	1	6	Bajo
231	231	95535 7496	JOHN CHICHA DE RIVERA	MAS CULI NO	25	3	1	0	0	1	149	Bajo
232	232	92185 8460	JOHNNY DECIMAVILLA PLUAS	MAS CULI NO	35	3	1	0	0	1	139	Bajo
233	233	13129 84162	Jonathan Zambrano	MAS CULI NO	38	4	1	0	0	1	79	Bajo
234	234	95096 1391	jordy pivaque villamar	MAS CULI NO	22	1	1	0	0	1	619	Bajo
235	235	95070 6143	Jordy Winsthon Pinargote Rivera	MAS CULI NO	29	4	1	1	0	1	474	Bajo
236	236	91920 4347	Jorge Coello Vargas	MAS CULI NO	50	4	0	0	3.5	1	38	Me dio
237	237	92117 0742	Jorge Lumisaca Cunduri	MAS CULI NO	46	4	1	0	0	1	36	Bajo
238	238	92017 0024	Jorge Perdomo Cevallos	MAS CULI NO	30	3	1	0	0	1	338	Bajo
239	239	17148 41689	Jorge Albuja	MAS CULI NO	20	1	1	0	10	2	3091. 6	Alto

240	240	93149 3001	JORGE JOAO AVILA MIELES	MAS CULI NO	41	4	1	0	0	1	269	Bajo
241	241	12055 56507	Jorge Leon	MAS CULI NO	29	4	1	0	10	7	4793	Alto
242	242	12056 99513	Jorge Leon	MAS CULI NO	24	4	1	0	0	1	99	Bajo
243	243	13098 63627	JORGE LUIS GARCIA PONCE	MAS CULI NO	26	4	1	0	0	2	4215	Alto
244	244	14004 88415	Jorge Macao	MAS CULI NO	22	4	1	0	0	2	867	Bajo
245	245	92931 0498	Jorge Miguel Guerra Merino	MAS CULI NO	34	4	0	0	0	1	129	Me dio
246	246	17049 51456	Jorge Salazar	MAS CULI NO	49	4	1	0	0	1	40	Bajo
247	247	10708 3446	jorge saquipay	MAS CULI NO	42	4	1	0	0	1	259	Bajo
248	248	92195 9961	Jose Cervantes	MAS CULI NO	47	3	1	0	0	1	42	Bajo
249	249	13014 97770	Jose Garcia	MAS CULI NO	45	1	1	0	0	1	228	Bajo
250	250	60493 5056	José Merino	MAS CULI NO	43	4	0	0	0	1	79	Me dio
251	251	10015 88845	José Moreno	MAS CULI NO	47	4	1	0	0	1	25	Bajo
252	252	13125 74468	José Parrales	MAS CULI NO	46	4	1	0	0	1	45	Bajo
253	253	13012 99382	José Alfredo Arguello Valdiviezo	MAS CULI NO	42	4	1	0	0	1	199	Bajo
254	254	91964 3882	jose delgado torres	MAS CULI NO	36	1	0	0	8	1	277	Me dio
255	255	16003 70850	Jose Flor	MAS CULI NO	42	4	1	0	0	1	90	Bajo
256	256	96347 7062	Jose Guillen	MAS CULI NO	42	4	1	0	0	2	1791. 29	Bajo

257	257	18035 29229	Jose Hidalgo Tirado	MAS CULI NO	43	4	0	0	23	1	202	Me dio
258	258	92309 4486	Jose Jacinto Martillo Holguin	MAS CULI NO	30	4	1	0	0	1	124	Bajo
259	259	60333 2024	Jose Larrea Davalos	MAS CULI NO	24	4	1	0	0	1	339	Bajo
260	260	17180 06602	José Luis Collaguazo	MAS CULI NO	33	4	1	1	0	1	10	Bajo
261	261	13095 20615	JOSE LUIS GARCIA PULIDO	MAS CULI NO	26	4	0	0	0	1	329	Me dio
262	262	17535 47254	José Miguel Mora Lumbi	MAS CULI NO	25	4	1	0	0	1	458	Bajo
263	263	96261 5902	JOSE SANCHEZ SANCHEZ	MAS CULI NO	23	4	0	0	0	1	407	Me dio
264	264	95160 9445	Joselyne Gabriela Robles Zavala	FEM ENI NO	24	4	0	0	0	2	574	Me dio
265	265	95835 4888	JOSELY RONQUILLO SALGUERO	FEM ENI NO	26	3	1	0	0	1	499	Bajo
266	266	92910 3570	Josue Jaramillo	MAS CULI NO	39	1	1	0	0	1	93	Bajo
267	267	23509 87786	Josue Steven Gamboa Ponce	MAS CULI NO	39	1	1	0	0	1	67	Bajo
268	268	92344 6892	Juan carlos Cabrera	MAS CULI NO	28	3	1	0	0	1	518	Bajo
269	269	93174 8529	Juan Carlos Chancay tomala	MAS CULI NO	49	3	1	0	0	1	20	Bajo
270	270	70325 5257	Juan javier Cajamarca nieto	MAS CULI NO	50	4	1	0	0	1	22	Bajo
271	271	70345 7598	juan pablo morales macas	MAS CULI NO	46	4	1	0	0	1	59	Bajo
272	272	13110 37061	Juan Vera	MAS CULI NO	41	4	0	0	4.5	1	74	Me dio
273	273	17118 07550	Juliana Ferreira	FEM ENI NO	50	4	1	0	0	1	32	Bajo

274	274	13114 34813	Julio Cesar Rodriguez	MAS CULI NO	23	1	0	0	10	3	653	Me dio
275	275	17133 64956	Julio Coronel Coronel	MAS CULI NO	48	4	0	0	5	1	39	Me dio
276	276	92702 9900	Julio Hidalgo Villegas	MAS CULI NO	38	1	1	0	0	2	164	Bajo
277	277	12070 72834	Julio Morales Urquiza	MAS CULI NO	38	4	1	0	0	1	69	Bajo
278	278	80051 1164	JULIO RAMON ANDRADE MOSQUERA	MAS CULI NO	29	4	0	0	5.26	1	615	Me dio
279	279	24000 00986	Julio Ricardo Suarez Gonzalez	MAS CULI NO	38	3	1	0	0	1	162	Bajo
280	280	12068 27196	Justin Cevallos	MAS CULI NO	46	4	0	0	0	1	26	Me dio
281	281	92529 9810	Karem Gaínza Cornejo	FEM ENI NO	39	4	1	0	0	1	189	Bajo
282	282	13136 33008	Karen Del Valle Solórzano	FEM ENI NO	24	1	1	0	0	1	944	Bajo
283	283	95031 5119	Karen Pincay	FEM ENI NO	26	4	1	0	0	1	99	Bajo
284	284	13063 06588	Karina Bravo	FEM ENI NO	43	4	0	0	15	2	224	Me dio
285	285	93148 2269	Karina Fanny NIETO	FEM ENI NO	29	1	0	0	23	1	442	Me dio
286	286	80266 8970	Karla Mercado	FEM ENI NO	49	4	1	0	0	1	45	Bajo
287	287	91739 8372	Katherine Velasco	FEM ENI NO	27	1	1	0	0	1	363	Bajo
288	288	50172 0668	KATTIA INÉS Navas Padilla	FEM ENI NO	29	4	1	0	0	1	30	Bajo
289	289	91627 8740	Katty Eloísa Villao Romero	FEM ENI NO	21	1	0	0	3.5	2	339	Me dio
290	290	13072 15283	KATTY MARGOT	FEM ENI NO	35	4	1	0	0	2	102	Bajo

5	MAS CULI NO	29	1	1	0	0	1	134	Baj o	0.99890	0.00110	0.00000
6	MAS CULI NO	25	4	1	0	0	2	1520	Baj o	0.97895	0.00140	0.01965
7	FEM ENIN O	36	4	1	0	0	1	184	Baj o	0.99935	0.00060	0.00005
8	FEM ENIN O	27	3	1	0	0	2	567	Baj o	0.99275	0.00100	0.00625
9	MAS CULI NO	25	1	1	0	0	2	2699	Baj o	0.94230	0.00095	0.05675
10	MAS CULI NO	47	3	1	0	0	1	149	Baj o	0.99930	0.00070	0.00000
11	MAS CULI NO	50	1	0	0	23	1	157	Me dio	0.01455	0.98495	0.00050
12	MAS CULI NO	36	4	0	0	5	1	84	Me dio	0.00890	0.99090	0.00020
13	MAS CULI NO	50	1	0	0	3,5	1	43	Me dio	0.02380	0.97620	0.00000
14	FEM ENIN O	35	4	1	0	0	2	3014	Baj o	0.87315	0.00015	0.12670
15	FEM ENIN O	46	4	1	0	0	1	13	Baj o	0.99515	0.00485	0.00000
16	MAS CULI NO	39	4	0	0	4,5	1	84	Me dio	0.00725	0.99275	0.00000
17	FEM ENIN O	21	4	1	0	5	2	7864	Alt o	0.02055	0.01680	0.96265
18	FEM ENIN O	36	4	1	0	0	1	80	Baj o	0.99435	0.00560	0.00005
19	FEM ENIN O	37	4	0	0	0	1	79	Me dio	0.08735	0.91265	0.00000
20	MAS CULI NO	49	4	1	0	0	1	39	Baj o	0.99890	0.00110	0.00000
21	MAS CULI NO	46	1	1	0	0	1	40	Baj o	0.99715	0.00285	0.00000

22	MAS CULI NO	34	4	0	0	4,5	1	30	Medio	0.02130	0.97850	0.00020
23	MAS CULI NO	42	4	1	0	0	1	229	Bajo	0.99960	0.00040	0.00000
24	MAS CULI NO	48	3	1	0	0	1	149	Bajo	100.000	0.00000	0.00000
25	FEM ENINO	40	4	0	0	8	1	97	Medio	0.00750	0.99250	0.00000
26	MAS CULI NO	39	3	1	0	0	1	259	Bajo	0.99975	0.00025	0.00000
27	MAS CULI NO	46	3	1	0	0	1	59	Bajo	0.99990	0.00010	0.00000
28	FEM ENINO	33	1	1	0	0	1	121	Bajo	0.99890	0.00040	0.00070
29	FEM ENINO	49	4	1	0	0	1	39	Bajo	0.99890	0.00110	0.00000
30	FEM ENINO	23	4	1	1	0	1	302	Bajo	0.96975	0.02985	0.00040
31	FEM ENINO	24	1	1	0	0	2	416	Bajo	0.98225	0.00680	0.01095
32	FEM ENINO	40	4	0	0	10	1	180	Medio	0.00800	0.99155	0.00045
33	FEM ENINO	31	4	1	1	0	1	5	Bajo	0.99690	0.00310	0.00000
34	MAS CULI NO	21	4	1	0	0	1	112	Bajo	0.99180	0.00595	0.00225
35	MAS CULI NO	25	3	1	0	0	4	338	Alto	0.95130	0.00595	0.04275
36	MAS CULI NO	23	4	0	0	70	4	1690	Alto	0.01560	0.29885	0.68555
37	MAS CULI NO	25	4	1	0	0	1	15	Bajo	0.99920	0.00075	0.00005
38	MAS CULI NO	34	1	1	0	0	1	309	Bajo	0.99920	0.00060	0.00020

39	MAS CULINO	23	1	1	0	0	1	99	Bajo	0.98860	0.01130	0.00010
40	FEMENINO	34	4	1	0	0	1	99	Bajo	0.99715	0.00285	0.00000
41	MAS CULINO	21	1	1	0	0	1	107	Bajo	0.98730	0.01045	0.00225
42	FEMENINO	38	4	0	0	3,5	1	66	Medio	0.00875	0.99125	0.00000
43	FEMENINO	50	4	0	0	0	1	159	Medio	0.10710	0.89290	0.00000
44	MAS CULINO	37	4	1	0	0	1	79	Bajo	0.98330	0.01670	0.00000
45	FEMENINO	44	4	0	0	3,5	1	93	Medio	0.01005	0.98995	0.00000
46	MAS CULINO	42	4	0	0	5	1	65	Medio	0.00855	0.99145	0.00000
47	MAS CULINO	21	3	1	0	0	1	499	Bajo	0.97915	0.01465	0.00620
48	MAS CULINO	37	4	1	0	0	1	269	Bajo	0.99435	0.00565	0.00000
49	MAS CULINO	45	1	1	0	0	1	228	Bajo	0.99980	0.00020	0.00000
50	MAS CULINO	30	3	1	0	0	1	299	Bajo	0.99360	0.00640	0.00000
51	FEMENINO	43	4	0	0	0	1	69	Medio	0.10305	0.89695	0.00000
52	FEMENINO	25	1	1	0	0	1	134	Bajo	0.99995	0.00005	0.00000
53	MAS CULINO	31	1	1	0	0	2	724	Bajo	0.99205	0.00160	0.00635
54	MAS CULINO	32	1	0	0	4,5	1	422	Medio	0.01045	0.98955	0.00000
55	MAS CULINO	23	1	0	0	0	1	497	Medio	0.09225	0.90725	0.00050

56	MASCULINO	21	4	1	0	0	1	149	Bajo	0.98435	0.01340	0.00225
57	MASCULINO	40	4	1	0	0	1	20	Bajo	0.99775	0.00225	0.00000
58	MASCULINO	50	3	1	0	0	1	149	Bajo	0.99245	0.00755	0.00000
59	MASCULINO	29	4	0	0	3,5	1	403	Medio	0.00825	0.99115	0.00060
60	MASCULINO	49	3	1	0	0	1	159	Bajo	0.99925	0.00075	0.00000
61	MASCULINO	35	4	1	0	0	2	6234	Alto	0.19960	0.00000	0.80040
62	FEMENINO	27	4	0	0	0	1	331	Medio	0.08615	0.91365	0.00020
63	FEMENINO	21	1	0	0	14,5	1	483	Medio	0.01300	0.96080	0.02620
64	MASCULINO	41	4	1	0	0	1	69	Bajo	0.99775	0.00225	0.00000
65	MASCULINO	32	3	1	0	0	1	319	Bajo	0.99900	0.00085	0.00015
66	MASCULINO	50	3	1	0	0	1	45	Bajo	0.99325	0.00675	0.00000
67	FEMENINO	22	1	0	0	20	1	718	Medio	0.00780	0.97880	0.01340
68	FEMENINO	20	4	0	1	30	2	8676	Alto	0.00635	0.19475	0.79890
69	FEMENINO	23	1	0	0	5	1	544	Medio	0.01355	0.97865	0.00780
70	MASCULINO	36	2	1	1	0	2	77	Bajo	0.98570	0.00250	0.01180
71	MASCULINO	42	4	0	0	3,5	1	243	Medio	0.00970	0.99030	0.00000
72	MASCULINO	27	1	1	0	8	2	1292	Bajo	0.42230	0.23615	0.34155

73	MAS CULI NO	22	1	1	0	0	1	323	Baj o	0.98460	0.01530	0.00010
74	MAS CULI NO	28	4	1	0	0	1	318	Baj o	0.99210	0.00785	0.00005
75	FEM ENIN O	45	4	1	0	0	1	189	Baj o	0.99940	0.00060	0.00000
76	FEM ENIN O	23	4	1	0	0	1	99	Baj o	0.99275	0.00675	0.00050
77	FEM ENIN O	45	3	1	0	0	1	179	Baj o	0.99995	0.00005	0.00000
78	MAS CULI NO	35	1	1	0	0	1	350	Baj o	0.99955	0.00030	0.00015
79	FEM ENIN O	38	4	0	0	23	1	172	Me dio	0.00780	0.99035	0.00185
80	MAS CULI NO	38	4	1	0	0	1	229	Baj o	0.99935	0.00065	0.00000
81	FEM ENIN O	49	3	1	0	0	1	149	Baj o	100.000	0.00000	0.00000
82	FEM ENIN O	30	4	1	0	0	1	119	Baj o	0.99965	0.00035	0.00000
83	MAS CULI NO	50	3	1	0	0	1	49	Baj o	0.99465	0.00535	0.00000
84	MAS CULI NO	50	3	1	0	0	1	149	Baj o	0.99245	0.00755	0.00000
85	MAS CULI NO	27	4	1	0	0	2	398	Baj o	0.98905	0.00395	0.00700
86	MAS CULI NO	30	4	1	0	0	1	379	Baj o	0.99780	0.00220	0.00000
87	MAS CULI NO	22	4	1	0	0	1	129	Baj o	0.98305	0.01670	0.00025
88	MAS CULI NO	38	4	1	0	0	1	279	Baj o	0.99475	0.00525	0.00000
89	MAS CULI NO	27	3	1	0	0	1	140	Baj o	0.99970	0.00030	0.00000

90	MASCULINO	40	1	1	0	0	1	94	Bajo	0.99930	0.00070	0.00000
91	MASCULINO	40	4	1	0	0	1	229	Bajo	0.99930	0.00070	0.00000
92	MASCULINO	23	4	1	0	6	2	7980	Alto	0.02360	0.00960	0.96680
93	FEMENINO	22	4	0	0	0	1	298	Medio	0.08210	0.91715	0.00075
94	MASCULINO	35	3	1	0	0	1	300	Bajo	0.99870	0.00100	0.00030
95	FEMENINO	48	3	1	0	0	1	149	Bajo	100.000	0.00000	0.00000
96	FEMENINO	27	1	0	0	10	1	549	Medio	0.01100	0.98595	0.00305
97	FEMENINO	32	4	1	0	0	2	9969	Alto	0.17615	0.00000	0.82385
98	FEMENINO	32	3	1	0	0	1	499	Bajo	0.99865	0.00120	0.00015
99	MASCULINO	44	4	1	0	0	1	199	Bajo	0.99940	0.00060	0.00000
100	MASCULINO	43	1	1	0	0	1	188	Bajo	0.99945	0.00055	0.00000
101	MASCULINO	38	1	1	0	0	1	239	Bajo	0.99990	0.00010	0.00000
102	MASCULINO	27	4	0	0	3,5	2	462	Medio	0.02360	0.91785	0.05855
103	FEMENINO	26	4	1	0	10	2	4574	Alto	0.03545	0.01540	0.94915
104	MASCULINO	45	4	1	1	0	1	66	Bajo	0.98620	0.01380	0.00000
105	MASCULINO	21	4	1	0	0	1	20	Bajo	0.99150	0.00625	0.00225
106	FEMENINO	39	4	0	0	0	1	89	Medio	0.09315	0.90685	0.00000

1 0 7	FEM ENIN O	43	4	1	0	0	1	174	Baj o	0.99850	0.00150	0.00000
1 8	MAS CULI NO	48	3	1	0	0	1	149	Baj o	100.000	0.00000	0.00000
1 9	MAS CULI NO	26	4	0	0	7,5	1	807	Me dio	0.01105	0.98255	0.00640
1 0	MAS CULI NO	44	4	1	0	0	1	89	Baj o	0.99565	0.00435	0.00000
1 1	MAS CULI NO	35	4	1	0	0	1	19	Baj o	0.99745	0.00170	0.00085
1 2	MAS CULI NO	44	3	1	0	0	1	169	Baj o	0.99990	0.00010	0.00000
1 3	MAS CULI NO	41	4	1	0	0	1	269	Baj o	0.99980	0.00020	0.00000
1 4	MAS CULI NO	34	4	1	0	0	1	149	Baj o	0.99735	0.00265	0.00000
1 5	MAS CULI NO	37	4	0	0	0	1	279	Me dio	0.09495	0.90505	0.00000
1 6	MAS CULI NO	33	4	1	0	0	1	348	Baj o	0.99950	0.00050	0.00000
1 7	MAS CULI NO	28	4	1	0	4,5	2	8914	Alt o	0.03525	0.02665	0.93810
1 8	FEM ENIN O	27	4	1	0	0	2	379	Baj o	0.99010	0.00290	0.00700
1 9	MAS CULI NO	33	4	1	0	0	1	299	Baj o	0.99940	0.00060	0.00000
1 0	MAS CULI NO	27	4	1	0	0	1	142	Baj o	0.99875	0.00125	0.00000
1 1	MAS CULI NO	26	4	1	0	0	1	149	Baj o	0.99660	0.00235	0.00105
1 2	MAS CULI NO	33	1	1	0	0	1	134	Baj o	0.99905	0.00025	0.00070
1 3	MAS CULI NO	26	4	1	0	0	1	489	Baj o	0.99105	0.00790	0.00105

1 2 4	MAS CULI NO	28	4	1	0	0	1	99	Baj o	0.99875	0.00120	0.00005
1 5	MAS CULI NO	49	4	0	0	3,5	1	41	Me dio	0.01215	0.98785	0.00000
1 6	MAS CULI NO	36	1	1	0	0	1	67	Baj o	0.99885	0.00115	0.00000
1 7	FEM ENIN O	25	4	1	0	0	1	137	Baj o	0.99965	0.00030	0.00005
1 8	FEM ENIN O	46	4	1	0	0	1	42	Baj o	0.99945	0.00055	0.00000
1 9	FEM ENIN O	49	4	1	0	0	1	50	Baj o	0.99970	0.00030	0.00000
1 0	MAS CULI NO	44	4	1	0	0	1	269	Baj o	0.99975	0.00025	0.00000
1 1	MAS CULI NO	21	3	1	0	0	2	6916	Alt o	0.16615	0.00285	0.83100
1 2	FEM ENIN O	40	4	1	0	0	1	12	Baj o	0.99820	0.00180	0.00000
1 3	MAS CULI NO	28	3	1	0	0	1	499	Baj o	0.99905	0.00095	0.00000
1 4	MAS CULI NO	25	1	1	0	0	1	364	Baj o	0.99970	0.00030	0.00000
1 5	MAS CULI NO	49	3	1	0	0	1	149	Baj o	100.000	0.00000	0.00000
1 6	FEM ENIN O	38	1	1	0	0	1	242	Baj o	0.99990	0.00010	0.00000
1 7	FEM ENIN O	48	3	1	0	0	1	149	Baj o	100.000	0.00000	0.00000
1 8	MAS CULI NO	45	1	1	0	0	1	66	Baj o	0.99920	0.00080	0.00000
1 9	MAS CULI NO	34	4	1	0	0	2	114	Baj o	0.98615	0.00125	0.01260
1 0	FEM ENIN O	38	4	1	0	0	1	179	Baj o	0.99920	0.00080	0.00000

1 4 1	MAS CULI NO	32	4	1	0	0	1	17	Baj o	0.99885	0.00065	0.00050
1 4 2	MAS CULI NO	36	4	1	0	0	1	69	Baj o	0.99740	0.00255	0.00005
1 4 3	FEM ENIN O	22	4	0	0	9,5	1	103	Me dio	0.01395	0.97595	0.01010
1 4 4	MAS CULI NO	50	4	1	0	0	1	47	Baj o	0.98360	0.01640	0.00000
1 4 5	MAS CULI NO	28	4	1	0	0	1	299	Baj o	0.99730	0.00265	0.00005
1 4 6	FEM ENIN O	37	4	0	0	0	1	79	Me dio	0.08735	0.91265	0.00000
1 4 7	MAS CULI NO	47	4	1	1	0	2	36	Baj o	0.98355	0.01190	0.00455
1 4 8	MAS CULI NO	33	3	1	0	0	3	717	Baj o	0.98250	0.00400	0.01350
1 4 9	MAS CULI NO	39	4	1	0	0	1	188	Baj o	0.99820	0.00180	0.00000
1 5 0	MAS CULI NO	32	1	1	0	0	2	484	Baj o	0.98715	0.00225	0.01060
1 5 1	MAS CULI NO	22	4	0	0	0	1	149	Me dio	0.09840	0.90085	0.00075
1 5 2	MAS CULI NO	44	1	1	0	0	1	255	Baj o	0.99990	0.00010	0.00000
1 5 3	FEM ENIN O	47	4	0	1	0	1	151	Me dio	0.10440	0.89560	0.00000
1 5 4	MAS CULI NO	48	4	1	0	0	1	32	Baj o	0.99875	0.00125	0.00000
1 5 5	MAS CULI NO	42	4	1	0	0	1	177	Baj o	0.99950	0.00050	0.00000
1 5 6	MAS CULI NO	32	4	1	0	0	1	379	Baj o	0.99790	0.00160	0.00050
1 5 7	MAS CULI NO	38	4	0	0	9	1	219	Me dio	0.00760	0.99240	0.00000

158	MASCULINO	33	4	0	0	7	1	396	Medio	0.01195	0.98765	0.00040
159	MASCULINO	36	3	1	0	0	1	189	Bajo	0.99990	0.00010	0.00000
160	FEMENINO	40	4	1	0	0	1	179	Bajo	0.99900	0.00100	0.00000
161	MASCULINO	40	4	1	0	0	1	61	Bajo	0.99890	0.00110	0.00000
162	MASCULINO	49	3	1	0	0	1	152	Bajo	0.99945	0.00055	0.00000
163	FEMENINO	37	4	1	0	0	2	174	Bajo	0.98765	0.00325	0.00910
164	MASCULINO	28	4	0	0	3,5	1	103	Medio	0.01025	0.98585	0.00390
165	MASCULINO	30	4	1	0	0	1	129	Bajo	0.99900	0.00100	0.00000
166	FEMENINO	39	4	1	0	0	1	78	Bajo	0.99420	0.00580	0.00000
167	FEMENINO	23	1	1	0	0	1	379	Bajo	0.98355	0.01635	0.00010
168	FEMENINO	50	4	1	0	0	1	50	Bajo	0.98360	0.01640	0.00000
169	FEMENINO	30	4	1	0	0	2	338	Bajo	0.99325	0.00095	0.00580
170	FEMENINO	33	1	1	0	0	2	8711,9	Alto	0.22165	0.00020	0.77815
171	MASCULINO	47	4	1	0	0	1	11	Bajo	0.99790	0.00210	0.00000
172	MASCULINO	39	4	0	0	4,5	1	84	Medio	0.00725	0.99275	0.00000
173	FEMENINO	28	4	0	0	35	2	5882	Alto	0.00260	0.14060	0.85680
174	FEMENINO	28	1	0	0	10	1	387	Medio	0.00820	0.98845	0.00335

175	MAS CULI NO	34	4	0	0	0	1	128	Medio	0.10120	0.89880	0.00000
176	MAS CULI NO	36	4	1	0	0	1	199	Bajo	0.99945	0.00050	0.00005
177	FEMENINO	28	1	0	0	3,5	1	299	Medio	0.02280	0.97535	0.00185
178	MAS CULI NO	28	4	1	0	0	2	1992,7	Bajo	0.97260	0.00055	0.02685
179	MAS CULI NO	32	4	1	0	0	1	449	Bajo	0.99655	0.00290	0.00055
180	MAS CULI NO	44	4	1	0	0	1	239	Bajo	0.99965	0.00035	0.00000
181	MAS CULI NO	21	3	1	0	0	1	359	Bajo	0.98840	0.00665	0.00495
182	MAS CULI NO	47	4	1	0	0	1	35	Bajo	0.99860	0.00140	0.00000
183	MAS CULI NO	33	3	1	0	0	1	298	Bajo	0.99810	0.00190	0.00000
184	MAS CULI NO	26	4	1	0	0	1	289	Bajo	0.98710	0.01185	0.00105
185	MAS CULI NO	42	1	1	0	0	1	188	Bajo	0.99975	0.00025	0.00000
186	MAS CULI NO	49	4	1	0	0	1	159	Bajo	0.99295	0.00705	0.00000
187	MAS CULI NO	38	4	1	0	0	1	279	Bajo	0.99475	0.00525	0.00000
188	MAS CULI NO	43	4	1	0	0	1	69	Bajo	0.99275	0.00725	0.00000
189	MAS CULI NO	43	4	1	0	0	1	89	Bajo	0.99310	0.00690	0.00000
190	MAS CULI NO	27	4	1	0	0	1	139	Bajo	0.99875	0.00125	0.00000
191	MAS CULI NO	33	4	1	0	0	1	344	Bajo	0.99950	0.00050	0.00000

192	MASCULINO	25	1	1	0	0	2	1391,13	Bajo	0.99210	0.00135	0.00655
193	FEMENINO	20	1	1	0	0	1	462	Bajo	0.92565	0.03075	0.04360
194	MASCULINO	20	4	0	0	5,25	2	462	Medio	0.01590	0.90785	0.07625
195	FEMENINO	47	4	0	0	3,5	1	33	Medio	0.01370	0.98630	0.00000
196	MASCULINO	35	4	1	0	0	1	6	Bajo	0.99750	0.00165	0.00085
197	MASCULINO	24	4	1	0	0	1	31	Bajo	0.99890	0.00095	0.00015
198	MASCULINO	43	4	1	0	0	1	259	Bajo	0.99895	0.00105	0.00000
199	MASCULINO	48	3	1	0	0	1	149	Bajo	100.000	0.00000	0.00000
200	FEMENINO	23	4	1	0	0	1	120	Bajo	0.99315	0.00635	0.00050
201	MASCULINO	24	4	0	0	13,5	2	7093	Alto	0.00255	0.15215	0.84530
202	FEMENINO	29	3	1	0	0	1	398	Bajo	0.99620	0.00380	0.00000
203	MASCULINO	27	3	1	0	0	1	129	Bajo	0.99955	0.00045	0.00000
204	MASCULINO	45	3	1	0	0	1	269	Bajo	0.99895	0.00105	0.00000
205	MASCULINO	31	3	1	0	0	1	369	Bajo	0.99940	0.00060	0.00000
206	MASCULINO	30	4	1	0	0	1	288	Bajo	0.99540	0.00460	0.00000
207	FEMENINO	42	3	1	0	0	1	179	Bajo	0.99995	0.00005	0.00000
208	FEMENINO	46	4	1	0	0	1	45	Bajo	0.99975	0.00025	0.00000

209	FEMENINO	40	3	1	0	0	3	226	Bajo	0.98990	0.00305	0.00705
210	FEMENINO	35	4	1	0	0	1	299	Bajo	0.99795	0.00120	0.00085
211	FEMENINO	21	4	1	0	0	1	99	Bajo	0.99105	0.00670	0.00225
212	FEMENINO	34	1	0	0	0	1	350	Medio	0.16805	0.83185	0.00010
213	FEMENINO	49	3	1	0	0	1	149	Bajo	100.000	0.00000	0.00000
214	FEMENINO	33	1	1	1	0	1	319	Bajo	0.99355	0.00630	0.00015
215	MASCULINO	36	3	1	0	0	1	259	Bajo	0.99965	0.00035	0.00000
216	FEMENINO	27	3	1	0	0	1	149	Bajo	0.99940	0.00060	0.00000
217	FEMENINO	24	4	1	0	0	1	349	Bajo	0.99615	0.00370	0.00015
218	FEMENINO	25	4	1	0	0	1	148	Bajo	0.99855	0.00140	0.00005
219	MASCULINO	35	1	1	0	0	1	485	Bajo	0.99425	0.00545	0.00030
220	MASCULINO	21	1	0	0	0	1	498	Medio	0.09910	0.89990	0.00100
221	FEMENINO	35	4	1	0	0	1	928	Bajo	0.99380	0.00530	0.00090
222	MASCULINO	27	1	1	0	0	1	134	Bajo	0.99995	0.00005	0.00000
223	MASCULINO	43	4	1	0	0	1	229	Bajo	0.99875	0.00125	0.00000
224	MASCULINO	24	3	1	0	0	1	99	Bajo	0.99930	0.00060	0.00010
225	MASCULINO	36	4	1	0	0	1	69	Bajo	0.99740	0.00255	0.00005

2 2 6	MAS CULI NO	43	4	1	0	0	1	193	Baj o	0.99860	0.00140	0.00000
2 7	MAS CULI NO	20	4	1	0	0	3	4882,5	Alt o	0.12885	0.00685	0.86430
2 8	MAS CULI NO	36	1	1	0	0	2	244	Baj o	0.98475	0.00055	0.01470
2 9	MAS CULI NO	49	3	1	0	0	3	50	Baj o	0.98990	0.00345	0.00665
2 3 0	FEM ENIN O	39	4	1	1	0	1	6	Baj o	0.99530	0.00470	0.00000
2 1	MAS CULI NO	25	3	1	0	0	1	149	Baj o	0.99940	0.00055	0.00005
2 2	MAS CULI NO	35	3	1	0	0	1	139	Baj o	0.99950	0.00020	0.00030
2 3	MAS CULI NO	38	4	1	0	0	1	79	Baj o	0.99095	0.00905	0.00000
2 4	MAS CULI NO	22	1	1	0	0	1	619	Baj o	0.95085	0.04855	0.00060
2 5	MAS CULI NO	29	4	1	1	0	1	474	Baj o	0.97420	0.02580	0.00000
2 6	MAS CULI NO	50	4	0	0	3,5	1	38	Me dio	0.01225	0.98775	0.00000
2 7	MAS CULI NO	46	4	1	0	0	1	36	Baj o	0.99890	0.00110	0.00000
2 8	MAS CULI NO	30	3	1	0	0	1	338	Baj o	0.99685	0.00315	0.00000
2 9	MAS CULI NO	20	1	1	0	10	2	3091,6	Alt o	0.14350	0.08020	0.77630
2 0	MAS CULI NO	41	4	1	0	0	1	269	Baj o	0.99980	0.00020	0.00000
2 1	MAS CULI NO	29	4	1	0	10	7	4793	Alt o	0.05305	0.06140	0.88555
2 2	MAS CULI NO	24	4	1	0	0	1	99	Baj o	0.99900	0.00085	0.00015

2 4 3	MAS CULI NO	26	4	1	0	0	2	4215	Alt o	0.24930	0.00050	0.75020
2 4 4	MAS CULI NO	22	4	1	0	0	2	867	Baj o	0.92970	0.02150	0.04880
2 4 5	MAS CULI NO	34	4	0	0	0	1	129	Me dio	0.10120	0.89880	0.00000
2 4 6	MAS CULI NO	49	4	1	0	0	1	40	Baj o	0.99930	0.00070	0.00000
2 4 7	MAS CULI NO	42	4	1	0	0	1	259	Baj o	0.99975	0.00025	0.00000
2 4 8	MAS CULI NO	47	3	1	0	0	1	42	Baj o	0.99720	0.00280	0.00000
2 4 9	MAS CULI NO	45	1	1	0	0	1	228	Baj o	0.99980	0.00020	0.00000
2 5 0	MAS CULI NO	43	4	0	0	0	1	79	Me dio	0.09115	0.90885	0.00000
2 5 1	MAS CULI NO	47	4	1	0	0	1	25	Baj o	0.99585	0.00415	0.00000
2 5 2	MAS CULI NO	46	4	1	0	0	1	45	Baj o	0.99975	0.00025	0.00000
2 5 3	MAS CULI NO	42	4	1	0	0	1	199	Baj o	0.99940	0.00060	0.00000
2 5 4	MAS CULI NO	36	1	0	0	8	1	277	Me dio	0.01230	0.98765	0.00005
2 5 5	MAS CULI NO	42	4	1	0	0	1	90	Baj o	0.99675	0.00325	0.00000
2 5 6	MAS CULI NO	42	4	1	0	0	2	1791,2 9	Baj o	0.96715	0.00685	0.02600
2 5 7	MAS CULI NO	43	4	0	0	23	1	202	Me dio	0.00800	0.99020	0.00180
2 5 8	MAS CULI NO	30	4	1	0	0	1	124	Baj o	0.99960	0.00040	0.00000
2 5 9	MAS CULI NO	24	4	1	0	0	1	339	Baj o	0.99430	0.00555	0.00015

260	MASCULINO	33	4	1	1	0	1	10	Bajo	0.99680	0.00320	0.00000
261	MASCULINO	26	4	0	0	0	1	329	Medio	0.08790	0.91175	0.00035
262	MASCULINO	25	4	1	0	0	1	458	Bajo	0.99520	0.00475	0.00005
263	MASCULINO	23	4	0	0	0	1	407	Medio	0.08385	0.91375	0.00240
264	FEMENINO	24	4	0	0	0	2	574	Medio	0.23435	0.71825	0.04740
265	FEMENINO	26	3	1	0	0	1	499	Bajo	0.99700	0.00245	0.00055
266	MASCULINO	39	1	1	0	0	1	93	Bajo	0.99910	0.00090	0.00000
267	MASCULINO	39	1	1	0	0	1	67	Bajo	0.99935	0.00065	0.00000
268	MASCULINO	28	3	1	0	0	1	518	Bajo	0.99915	0.00085	0.00000
269	MASCULINO	49	3	1	0	0	1	20	Bajo	0.99765	0.00235	0.00000
270	MASCULINO	50	4	1	0	0	1	22	Bajo	0.98050	0.01950	0.00000
271	MASCULINO	46	4	1	0	0	1	59	Bajo	0.99965	0.00035	0.00000
272	MASCULINO	41	4	0	0	4,5	1	74	Medio	0.00910	0.99090	0.00000
273	FEMENINO	50	4	1	0	0	1	32	Bajo	0.98235	0.01765	0.00000
274	MASCULINO	23	1	0	0	10	3	653	Medio	0.01930	0.92085	0.05985
275	MASCULINO	48	4	0	0	5	1	39	Medio	0.01090	0.98910	0.00000
276	MASCULINO	38	1	1	0	0	2	164	Bajo	0.98570	0.00075	0.01355

277	MAS CULI NO	38	4	1	0	0	1	69	Baj o	0.99715	0.00285	0.00000
278	MAS CULI NO	29	4	0	0	5,26	1	615	Me dio	0.00950	0.98980	0.00070
279	MAS CULI NO	38	3	1	0	0	1	162	Baj o	0.99975	0.00025	0.00000
280	MAS CULI NO	46	4	0	0	0	1	26	Me dio	0.12400	0.87600	0.00000
281	FEM ENIN O	39	4	1	0	0	1	189	Baj o	0.99820	0.00180	0.00000
282	FEM ENIN O	24	1	1	0	0	1	944	Baj o	0.98570	0.01395	0.00035
283	FEM ENIN O	26	4	1	0	0	1	99	Baj o	0.99730	0.00165	0.00105
284	FEM ENIN O	43	4	0	0	15	2	224	Me dio	0.02105	0.91595	0.06300
285	FEM ENIN O	29	1	0	0	23	1	442	Me dio	0.00505	0.99275	0.00220
286	FEM ENIN O	49	4	1	0	0	1	45	Baj o	0.99965	0.00035	0.00000
287	FEM ENIN O	27	1	1	0	0	1	363	Baj o	0.99920	0.00080	0.00000
288	FEM ENIN O	29	4	1	0	0	1	30	Baj o	0.99855	0.00145	0.00000
289	FEM ENIN O	21	1	0	0	3,5	2	339	Me dio	0.04820	0.88675	0.06505
290	FEM ENIN O	35	4	1	0	0	2	102	Baj o	0.98145	0.00160	0.01695
291	MAS CULI NO	30	3	0	0	0	1	284	Me dio	0.12585	0.87405	0.00010
292	FEM ENIN O	40	3	1	0	0	1	84	Baj o	0.99835	0.00165	0.00000
293	FEM ENIN O	35	3	0	0	12	1	119	Me dio	0.03570	0.95735	0.00695

2 9 4	MAS CULI NO	27	4	0	1	0	1	410	Me dio	0.09400	0.90585	0.00015
2 5	MAS CULI NO	21	4	1	0	0	1	139	Baj o	0.99130	0.00645	0.00225
2 6	FEM ENIN O	22	3	1	0	0	1	499	Baj o	0.98295	0.01575	0.00130
2 7	FEM ENIN O	22	3	1	0	0	1	499	Baj o	0.98295	0.01575	0.00130
2 8	FEM ENIN O	34	4	0	0	26	1	354	Me dio	0.00950	0.98735	0.00315
2 9 9	FEM ENIN O	27	1	1	0	0	1	420	Baj o	0.99440	0.00560	0.00000



**Presidencia
de la República
del Ecuador**



**Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes**



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **Muñoz Franco Milena Gardenia**, con C.C: # 0943863423 y **Suárez Aragonés Genaro José**, con C.C: # 0930366521, autores del trabajo de integración curricular: **nálisis de clusterización basado en el algoritmo K- Means Neighbourhood para la clasificación de tipos de clientes web en una empresa de Retail E-Commerce**, previo a la obtención del título de **Licenciado en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de integración curricular para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

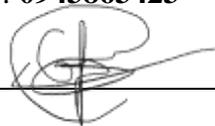
2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de integración curricular, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, **4 de septiembre de 2023**.

f. 

Nombre: **Muñoz Franco Milena Gardenia**

C.C: **0943863423**

f. 

Nombre: **Suárez Aragonés Genaro José**

C.C: **0930366521**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TRABAJO DE INTEGRACIÓN CURRICULAR

TEMA Y SUBTEMA:	Análisis de clusterización basado en el algoritmo K- Means Neighbourhood para la clasificación de tipos de clientes web en una empresa de Retail E-Commerce	
AUTOR(ES)	Muñoz Franco, Milena Gardenia; Suárez Aragonés, Genaro José	
REVISOR(ES)/TUTOR(ES)	Félix Miguel Carrera Buri	
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil	
FACULTAD:	Facultad de Economía y Empresa	
CARRERA:	Negocios Internacionales	
TÍTULO OBTENIDO:	Licenciado en Negocios Internacionales	
FECHA DE PUBLICACIÓN:	4 de septiembre de 2023	No. DE PÁGINAS: 90
ÁREAS TEMÁTICAS:	Clasificación de clientes, Aprendizaje automático, Inteligencia de Negocios	
PALABRAS CLAVES/ KEYWORDS:	E-commerce, Retail, K-means, Inteligencia de Negocios, Clusterización, Machine Learning.	
RESUMEN/ABSTRACT:	<p>En el mundo moderno y con la digitalización como pilar clave de nuestro día a día, saber e implementar las nuevas tecnologías que se desarrollan es de vital importancia pues son capaces de agilizar o mejorar los procesos que llevamos a cabo en nuestras actividades cotidianas. Sin lugar a duda el apogeo de la inteligencia artificial y sus derivados es la tendencia que hoy en día se desarrolla, y en el campo empresarial no es la excepción siendo infinita sus aplicaciones. Conocer las preferencias de tus clientes sin necesidad de tenerlos presentes representa una gran ventaja competitiva con respecto a la competencia, años atrás esta tarea se vería como algo imposible mas hoy en día con la implementación de nuevas tecnologías paso de ser un sueño a una realidad. En este trabajo se pretende demostrar como la implementación de un sistema de algoritmo de Machine Learning permitiría a las empresas Retail que están presentes en el E-commerce facilitar su clasificación de sus clientes basándose en características similares, esto mediante el modelo de K- means Neighbor y para su validación mediante el modelo de árbol de decisión, dando consigo así agrupaciones de clústeres similares basados en variables como sería el tipo de pago, el tipo de envío, su uso no de cupones, el valor del envío y su número de transacciones y con ello poder posteriormente brindar descuentos, promociones o campañas dirigidas de mejor manera a los clientes que se les podría hacer más atractivo.</p>	
ADJUNTO PDF:	SI <input checked="" type="checkbox"/>	NO <input type="checkbox"/>
CONTACTO CON AUTOR/ES:	Teléfono: +593 999889374 Teléfono: +593 985109442	E-mail: milenagmf@hotmail.com E-mail: genarojosesuarez@gmail.com
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UIC):	Nombre: Román Bermeo, Cynthia Lizbeth Teléfono: +593-984228698 E-mail: cynthia.roman@cu.ucsg.edu.ec	
SECCIÓN PARA USO DE BIBLIOTECA		
Nº. DE REGISTRO (en base a datos):		
Nº. DE CLASIFICACIÓN:		
DIRECCIÓN URL (tesis en la web):		