

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

TÍTULO:

**Análisis de clasificación por tipo de cliente de la empresa
Labelprintsa S.A. en base a un modelo de clusterización por
Machine Learning.**

AUTORES:

Castillo Moncada, Rogger Francisco

Vera Álava, Andy Jahir

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

TUTOR:

Ing. Carrera Burí Félix Miguel Mgs.

Guayaquil, Ecuador

7 de febrero del 2023



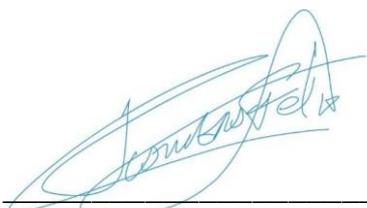
UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

CERTIFICACIÓN

Certificamos que el presente trabajo de integración curricular fue realizado en su totalidad por **Castillo Moncada, Rogger Francisco y Vera Álava, Andy Jahir**, como requerimiento para la obtención del título de **Licenciado en Negocios Internacionales**

TUTOR (A)

f. 
Ing. Carrera Burí, Félix Miguel Mgs.

DIRECTOR DE LA CARRERA


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

Guayaquil, a los 07 del mes de febrero del año 2023



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE ECONOMÍA Y EMPRESA
CARRERA DE NEGOCIOS INTERNACIONALES

DECLARACIÓN DE RESPONSABILIDAD

Nosotros, **Castillo Moncada, Rogger Francisco y Vera Álava, Andy Jahir**

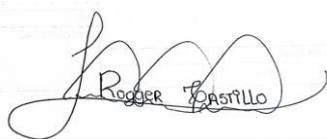
DECLARAMOS QUE:

El Trabajo de Integración Curricular, **Análisis de clasificación por tipo de cliente de la empresa Labelprintsa S.A. en base a un modelo de clusterización por Machine Learning** 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 nuestra total autoría.

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

Guayaquil, a los 07 del mes de febrero del año 2023

LOS AUTORES:

f. 

Castillo Moncada, Rogger Francisco

f. 

Vera Álava, Andy Jahir



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

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

AUTORIZACIÓN

Nosotros, **Castillo Moncada, Rogger Francisco y Vera Álava, Andy Jahir**

Autorizamos a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Integración Curricular, **Análisis de clasificación por tipo de cliente de la empresa Labelprintsa S.A. en base a un modelo de clusterización por Machine Learning**, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y total autoría.

Guayaquil, a los 07 del mes de febrero del año 2023

LOS AUTORES:

f. _____
Castillo Moncada, Rogger Francisco

f. _____
Vera Álava, Andy Jahir



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

REPORTE URKUND

URKUND

Documento [Roger Francisco Castillo Moncada - Andy Jahir Vera Alava.docx](#) (D157556515)

Presentado 2023-02-01 07:55 (-05:00)

Presentado por roggercastillo@gmail.com

Recibido felix.carrera01.ucsg@analysis.orkund.com

Mensaje Tesis - Roger Castillo y Andy Vera [Mostrar el mensaje completo](#)

3% de estas 42 páginas, se componen de texto presente en 8 fuentes.

Lista de fuentes	Bloques
Categoría	Enlace/nombre de archivo
	Universidad Católica de Santiago de Guayaquil / D151134894
	http://201.159.223.180/bitstream/3317/18340/1/T-UCSG-PRE-CEAC-CNI
	https://www.cienciadedatos.net/documentos/37_clustering_y_heatma
	Fundació per a la Universitat Oberta de Catalunya / D90438617
	Fundació per a la Universitat Oberta de Catalunya / D101273222
	Fundació per a la Universitat Oberta de Catalunya / D108782412
	Universidad Técnica Particular de Loja / D123106700

0 Adverten

Ing. Félix Miguel Carrera Buri Mgs.

AGRADECIMIENTO

Quiero comenzar agradeciendo a mi madre por ser padre y madre, por estar siempre presente en los buenos y malos momentos de mi vida, además de nunca dudar de mí y alentarme a siempre ser una mejor persona día a día, incluso cuando yo sentía que no podía más.

Gracias por enseñarme que, a pesar de las caídas, lo que importa es levantarse y seguir adelante.

También quiero agradecer a todos mis amigos que han recorrido conmigo este arduo camino llamado "Universidad". Son personas con las que he podido compartir experiencias únicas y siempre me han estado alentando. En especial, a Fátima y Camila, que desde el inicio han sido mis confidentes y el mejor equipo de trabajo que pude tener, a pesar de nuestras diferencias.

Por último, quiero expresar mi fuerte agradecimiento a todos los profesores de mi carrera, ya que, gracias a sus enseñanzas impartidas en clases, nos han preparado no solo a nivel académico, sino también para el mundo laboral.

- Rogger Castillo

AGRADECIMIENTO

Es con dicha y alegría que en la culminación de este proyecto ofrecerle mi agradecimiento a cada una de las personas que formaron parte de mi crecimiento tanto académica, profesional y personal a lo largo de esta etapa universitaria.

En primer lugar, deseo agradecerle a Dios en quién base mis principios e ideales, además de que me brindo a la mejor madre del mundo Psic. Deisy Alava Patiño quien ha sabido convertirse en mi inspiración y supo darme su apoyo y aliento a lo largo de mi vida forjándome la cual me preparo inculcándome los valores para convertirme en una persona de bien, a su vez a mi padre Abg. Jahir Vera Loor quien también ha sabido aconsejarse y ser un amigo a quien acudir en donde he encontrado un amor incondicional, es gracias a mis padres que me enseñaron que con trabajo duro y el carácter para afrontar la vida puedo lograr cualquier meta que me proponga y que el éxito es exclusivo para aquellos que se arriesgan en salir a buscarlo.

Deseo agradecerles a mis hermanos pues me permite convertirme cada día en una mejor persona tratando de ser alguien en quien puedan ver una guía y apoyo, a mi tía Oty quien supo ser mi segunda madre a lo largo de mi vida forjándome y acercándose continuamente a

Jehová en busca de que pueda llevar mi vida según sus enseñanzas de igual forma a mi tía Deyanira quien me ha aconsejado y se ha preocupado por mí en las diferentes etapas que he afrontado. A mis primos Alanís Vera y Ing. Alvin Vera con quien viví toda esta travesía y quienes han sabido ser mis confidentes y amigos desde mi niñez, junto a los cuales compartí y creé muchos de los recuerdos más importantes y especiales de mi vida universitaria.

A las amistades que conseguí a lo largo del camino y que se han sabido convertir en mi segunda familia siendo Carlos Orlando mi amigo desde inicio de la carrera, Alejandro Navarro casi como un hermano que con sus palabras ocurrentes supo levantarme el ánimo, a Marilyn Zhindon y Anita Crespo quienes supieron darme su apoyo y amistad desde siempre y Carla Sánchez quien se preocupó y me ayudó cuando estuve en dificultades también a mis perros que son mi fortaleza. Finalmente agradezco a mi tutor de tesis el Ing. Félix Carrera, por haberme ofrecido las herramientas y apoyo para poder culminar este proyecto además de mostrarme la rama en la que encontré pasión que es la estadística.

- Andy Vera

DEDICATORIA

El presente trabajo está dedicado a mi familia, especialmente a mi madre Maribel, quien con su esfuerzo ha hecho todo lo posible para proporcionarme todo lo necesario y más. Ella es mi inspiración para seguir adelante y alcanzar todos mis objetivos.

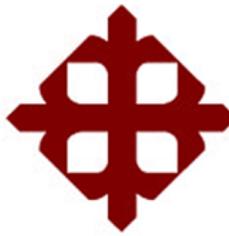
También lo dedico a mi yo del pasado, ya que gracias a su resiliencia y dedicación he logrado completar esta etapa de mi vida, sin dejar atrás a mis amigos que siempre han estado presentes y me han brindado un apoyo incondicional.

- Rogger Castillo

DEDICATORIA

El presente trabajo está dedicado a mi madre, mi más grande fuente de inspiración y aliento en el mundo, de no ser por su sacrificio y esfuerzo no habría llegado a donde estoy ahora y es sin lugar a dudas a quien le debo y le entregó cada uno de los logros que obtenga a lo largo de mi vida al ser quien me crió, me formó, me educo y me amó de forma inmensa e incondicional, quien inclusive en mis momentos más difíciles me extendió la mano y me alentó a no dejarme decaer por las adversidades que encontrará tanto en el exterior como en el interior. Espero pueda sentir el mismo orgullo que siento de ser su hijo

A su vez no puedo más que dedicar todo este esfuerzo a mí mismo en cada una de las fases que sobrelleve y afronte, por no rendirme ni decaer, por haberse levantando en cada momento además de haberme mantenido estoico y perseverante con la meta que me fije manteniendo siempre en mente que tengo la capacidad y la habilidad de lograr y dominar cualquier desafío, no existe reto tan grande que no pueda ser superado. Hoy al fin coseché parte de los frutos que los continuos esfuerzos realizados me están ofreciendo.



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

TRIBUNAL DE SUSTENTACIÓN

f. *Gabriela Hurtado*

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

f. *Cynthia Lizbeth Román Bermeo*

Ing. Cynthia Lizbeth, Román Bermeo Mgs.
COORDINADOR DEL ÁREA

f. *Carmen Padilla*

Ing. Carmen Paola, Padilla Lozano Ph.D
OPONENTE



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

CALIFICACIÓN

Castillo Moncada, Rogger Francisco

Vera Álava, Andy Jahir

Índice General

Capítulo I: Generalidades de la investigación	2
Introducción	2
Problemática	7
Problemática: Empresa	10
Justificación	12
Alcance	14
Objetivos	14
Objetivo General	14
Objetivos Específicos	15
Capítulo II: Fundamentación Teórica	15
Marco Teórico	15
Inteligencia Empresarial	15
ETL (Extract, Transform, Load)	16
<i>Extract</i>	16
<i>Transform</i>	17
<i>Load</i>	17
Inteligencia Artificial	17
Big Data	18
Aprendizaje Automático (ML)	20
Mecanismos de aprendizaje	21
Aprendizaje supervisado.	22
Tipos de algoritmo de supervisión.	24
<i>Algoritmo de clasificación supervisado.</i>	24
<i>Algoritmo de regresión supervisada.</i>	25
Aprendizaje no supervisado.	26
Tipos de aprendizaje no supervisado.	27
<i>Agrupamiento.</i>	27
<i>Asociación.</i>	27
Aprendizaje reforzado	27
Algoritmos	28
<i>Clasificación</i>	29
Clustering por supervisado	30
K-medias por supervisión	31
<i>Funcionamiento de la agrupación por K-medias</i>	32

Centroides	33
Escalado de datos (Estandarización)	33
Marco conceptual	34
Segmentación de clientes	34
Business Intelligence	35
Machine learning	38
Matriz de confusión	39
Outliers	39
Marco Legal	40
Capítulo III	41
Metodología	41
Metodología aplicativa para realización del clúster	44
<i>Planteamiento del problema</i>	44
<i>Estandarización</i>	44
<i>Dibujar el matrixplot y la correlación entre características</i>	45
<i>Calcular el número óptimo de clústeres</i>	45
<i>Método del centroide</i>	46
<i>Método del vecino más próximo</i>	47
<i>Realizar el cálculo con distintos métodos</i>	50
<i>Comparación de los clústeres calculados</i>	50
Operacionalización de variables	50
Metodología en RStudio	52
<i>Carga de data</i>	52
<i>Instalación y carga de paquetes</i>	53
<i>Verificación de variables cuantitativa y transformación a etiqueta</i>	54
<i>Semilla</i>	55
<i>Escalamiento de los datos</i>	55
<i>Estimación del número de clústeres</i>	56
<i>Cálculo de los clústeres</i>	56
<i>Graficamos los clústeres</i>	57
Resultados	58
Conclusiones	70
Recomendaciones	72
Referencias	74
Anexos	86

Índice de Tablas

Tabla 1 Aprendizaje supervisado vs no supervisado	20
Tabla 2 Métodos de distancia entre Clústers	44
Tabla 3 Matriz de distancia - Inicio	46
Tabla 4 Matriz de distancia 2.0	46
Tabla 5 Matriz de distancia 3.0	46
Tabla 6 Matriz de distancia 4.0	47
Tabla 7 Matriz de distancia 5.0	47
Tabla 8 Matriz de distancia Obtenida	47
Tabla 9 Operacionalización de variables	49

Índice de Figuras

Figura 1 Algoritmo de Machine Learning - Clusterización	11
Figura 2 Arquitectura Inteligencia de Negocios	15
Figura 3 Inteligencia Artificial – Usos	17
Figura 4 Big Data Analytics	18
Figura 5 Aplicativos del aprendizaje automático	19
Figura 6 Modelo de Aprendizaje Supervisado	22
Figura 7 Modelo de clasificación supervisado	23
Figura 8 Modelo de regresión supervisado	23
Figura 9 Modelo de Aprendizaje no Supervisado	25
Figura 10 Modelo de Aprendizaje Reforzado	26
Figura 11 Funcionamiento de los algoritmos	27
Figura 12 Formación de un clúster	28
Figura 13 Algoritmo de agrupación por K-medias	30
Figura 14 Visualización de un Centroide	31
Figura 15 Ejemplificación de matriz	42
Figura 16 Dendograma	47
Figura 17 Percentages of Dimension	57
Figura 18 Histogramas	58
Figura 19 Principal component Analysis	59
Figura 20 Variance Explained	59
Figura 21 Estimación de centroides	62
Figura 22 Gráfico de clústers óptimos	62
Figura 23 Creación de K-means	63
Figura 24 Ajuste del modelo	64
Figura 25 Panel de control de los diagramas de cajas por categoría de producto	65
Figura 26 Tendencia de los clusters	66
Figura 27 Clúster por categorías de productos	68
Figura 28 Gráfico elíptico de tipo euclidean	69

RESUMEN

La tecnología y la innovación son fundamentales para el crecimiento y competitividad de las empresas en la actualidad, en donde la automatización de procesos mediante el análisis de datos es una herramienta clave para mejorar la eficiencia y la calidad del servicio ofrecido a los clientes. Sin embargo, el manejo y entendimiento de conceptos de inteligencia artificial, como el aprendizaje automático, sigue siendo un desafío debido a la complejidad de estas tecnologías y a factores prácticos como costos, personal, formación y postura de los empresarios. A pesar de estos desafíos, es esencial que las empresas se adapten a la digitalización de los procesos para poder competir en los mercados globales. La clasificación y segmentación de clientes son técnicas fundamentales para las empresas, ya que permiten adaptar las estrategias y campañas de marketing de manera más personalizada a los diferentes tipos de clientes objetivos, y establecer estrategias eficientes para lograr una ventaja competitiva en el mercado. La presente investigación se enfoca en el desarrollo de un modelo de clasificación de clientes utilizando Machine Learning (ML) y clusterización para la empresa Labelprintsa. El objetivo es establecer estrategias personalizadas para diferentes grupos de clientes debido al aumento de la demanda de métodos de impresión y al crecimiento de la empresa. La aplicación de ML se justifica debido al crecimiento de la empresa y la necesidad de manejar una gran cantidad de datos para aplicar estrategias de retención o adquisición de clientes. El algoritmo propuesto es k means neighbours, el cual se espera logre un mínimo error en la predicción de clasificación de grupos de clientes y un posible aumento en la cuota de mercado.

Palabras Claves: Tecnología, Innovación, Automatización, Análisis de datos, Inteligencia artificial, aprendizaje automático

ABSTRACT

Technology and innovation are fundamental for the growth and competitiveness of companies today, where process automation through data analysis is a key tool to improve efficiency and the quality of service offered to customers. However, the management and understanding of artificial intelligence concepts, such as machine learning, remains a challenge due to the complexity of these technologies and practical factors such as costs, personnel, training, and business stance. Despite these challenges, it is essential for companies to adapt to the digitization of processes to compete in global markets. Customer classification and segmentation are fundamental techniques for companies, as they allow them to adapt marketing strategies and campaigns in a more personalized way to different types of target customers, and to establish efficient strategies to achieve a competitive advantage in the market. This research focuses on the development of a customer classification model using Machine Learning (ML) and clustering for the company Labelprintsa. The objective is to establish personalized strategies for different customer groups due to the increasing demand for printing methods and the growth of the company. The application of ML is justified due to the growth of the company and the need to handle a large amount of data to apply customer retention or acquisition strategies. The proposed algorithm is k means neighbours, which is expected to achieve a minimum error in the prediction of customer group classification and a possible increase in market share.

Keywords : Technology, Innovation, Automation, Data Analysis, Artificial Intelligence, Machine Learning,

RÉSUMÉ

La technologie et l'innovation sont fondamentales pour la croissance et la compétitivité des entreprises d'aujourd'hui, où l'automatisation des processus par l'analyse des données est un outil clé pour améliorer l'efficacité et la qualité du service offert aux clients. Cependant, le traitement et la compréhension des concepts d'intelligence artificielle, tels que l'apprentissage automatique, restent un défi en raison de la complexité de ces technologies et de facteurs pratiques tels que le coût, la dotation en personnel, la formation et la position de l'entreprise. Malgré ces défis, il est essentiel pour les entreprises de s'adapter à la numérisation des processus afin d'être compétitives sur les marchés mondiaux. La classification et la segmentation des clients sont des techniques fondamentales pour les entreprises, car elles leur permettent d'adapter les stratégies et les campagnes de marketing de manière plus personnalisée aux différents types de clients cibles, et d'établir des stratégies efficaces pour obtenir un avantage concurrentiel sur le marché. Cette recherche se concentre sur le développement d'un modèle de classification des clients utilisant le Machine Learning (ML) et le clustering pour la société Labelprintsa. L'objectif est d'établir des stratégies personnalisées pour différents groupes de clients en raison de la demande croissante de méthodes d'impression et de la croissance de l'entreprise. L'application du ML se justifie en raison de la croissance de l'entreprise et de la nécessité de traiter une grande quantité de données pour appliquer des stratégies de fidélisation ou d'acquisition de clients. L'algorithme proposé est celui des k moyens voisins, qui devrait permettre d'obtenir une erreur minimale dans la prédiction de la classification des groupes de clients et une augmentation possible de la part de marché.

Mots-clés : Technologie, Innovation, Automatisation, Analyse des données, Intelligence artificielle, Apprentissage automatique

Capítulo I: Generalidades de la investigación

Introducción

Las empresas alrededor del mundo poseen diferentes procesos dentro de sus diversas áreas y departamentos orientados a la optimización continua de recursos además de la búsqueda de aumentar su capital y clientes para que de esta forma puedan establecer un mejor posicionamiento dentro de los mercados y lograr una ventaja por sobre sus competidores.

Principalmente dentro del área comercial uno de los aspectos más importantes es el establecimiento de estrategias que permitan a la empresa lograr objetivos eficientes. Según la Universidad Latina de Costa Rica “Si una empresa desea alcanzar el éxito entonces la gestión estratégica y un plan, representa una ventaja frente a la competencia.” (2020). Parte fundamental de una buena estrategia consiste en poder localizarla en sectores específicos cosa que solo es posible si previamente se ha establecido una segmentación de mercados, es por lo que se le aporta una gran relevancia a la clasificación de clientes dentro de las empresas

La clasificación de clientes es fundamental principalmente en las áreas comerciales y de marketing dentro de las empresas pues al tener categorizados los diferentes tipos de clientes objetivos pueden adaptar sus estrategias y campañas para que de esta manera alcance al mercado meta de manera más personalizada y sea recibida de una mejor manera. Lastra afirma “Clasificar a los diferentes clientes y potenciales es una labor necesaria para toda organización si se busca mejorar la rentabilidad” (2021). Es por lo que una correcta clasificación y segmentación supone un factor clave y que aporta un valor agregado a cada una de las empresas que lo apliquen, pues dependiendo de los parámetros, métricas y formas en las que se

determine segmentar al público objetivo dependerá el éxito de o fracaso de la campaña o estrategia que se quiera implementar en base a dicha segmentación.

En la actualidad la analítica de datos a través de la estadística se ha convertido en una parte fundamental para las empresas pues consiste en una de las herramientas que tienen al alcance para poder hacer predicción conocer el estatus de la compañía o analizar los diferentes grados de eficiencia de los diferentes procesos e innovaciones que se vayan adquiriendo, no fue hasta hace relativamente poco tiempo que ha tomado más relevancia por parte de los diferentes directivos de las compañías, pues notaron lo fundamental que era su implementación a la par que permitía vislumbrar predicciones y resultados más fidedignos. Esto va más allá a través de un nuevo concepto que se aplica en el contexto de las IA (inteligencias artificiales) el cual es el aprendizaje automático (machine learning). Se pueden identificar muchas opiniones en torno al aprendizaje automático, pero hay una que predomina y es que, el hecho de que las máquinas aprendan “es una habilidad indispensable para hacer sistemas capaces de identificar patrones entre los datos para hacer predicciones” (Alameda, 2022). Esto se ve reflejado en su continuo aumento de su aplicación en las diferentes industrias por su gran polivalencia y versatilidad.

El machine learning es uno de los mejores métodos que las empresas pueden obtener actualmente para realizar una óptima clasificación pues supone la obtención de resultados más precisos de manera rápida y sistematizada, a través de sus diferentes algoritmos aporta diversas opciones dependiendo de los requerimientos de la empresa con respecto al tipo de clasificación, datos y tipo de variables que desee usar gracias a las propiedades adaptativas de esta tecnología, “identificar patrones en estos factores para predecir mejoraría la experiencia del cliente y también reduciría la rotación del

cliente” (Lizardo, 2021). Con lo cual su implementación es primordial principalmente a la par de la búsqueda de segmentaciones de clientes por las múltiples aportaciones que ofrece y el cual debe ser un aspecto imprescindible dentro de cualquier empresa.

El aprendizaje automático actualmente se implementa a lo largo del mundo para múltiples aplicaciones, está a la vanguardia del desarrollo y la innovación con gran tendencia en el incremento de su uso en diferentes sectores e industrias donde sus beneficios y utilidades son requeridos y consolidan un aspecto disruptivo y diferenciador “las empresas están invirtiendo mayores esfuerzos en su implementación, obteniendo mejores resultados que los alcanzados por los humanos” (Management Solutions, 2018). Esto demuestra que alrededor del mundo las empresas han reconocido su importancia y la gran relevancia que ya posee y la evolución que tendrá a través del tiempo.

En países como Estados Unidos o la Unión Europea el machine learning tiene una relevancia indiscutible dentro de cada uno de los sectores y también está viviendo su era dorada por los continuos esfuerzos que se realizan dentro de estas naciones en pro del desarrollo de las tecnologías y sus aplicaciones en diferentes sectores. Otras regiones como Asia también se encuentran liderando el uso y manejo de esta tecnología en conjunto con las IA donde las empresas cada día lo implementan en las diferentes áreas de la empresa donde gracias a la gran interconectividad con la tecnología que tienen los ciudadanos de países como china o Japón las empresas obtienen mayor cantidad de información a través de los pagos móviles realizados que proporcionan los datos necesarios (Micó, 2019). Esto se ve en contraste con regiones como África donde la tecnología se ha ralentizado por las diferentes problemáticas de

la región o en otro aspecto Latinoamérica que se ha visto muy rezagado en la implementación y desarrollo del ML en las empresas.

En Latinoamérica exceptuando ciertos países como México el machine learning y el desarrollo de las IA no es tan impulsado ni implementado tanto en las empresas como los gobiernos sumados a una población que no reconoce aún el alto poder y capacidad de esta herramienta con lo cual se ha ido creando una brecha que se intenta acortar a través de diferentes esfuerzos de que sea reconocido el valor de esta.

Muchas empresas en Latinoamérica aún mantienen un sistema arcaico en la analítica de datos e implementación de estrategias para sus campañas, aunque últimamente se ha ido denotando un cambio en dicha tendencia aún se mantiene como el gran porcentaje de la totalidad de las industrias. Entre los principales problemas que encuentra el machine learning en Latinoamérica para su desarrollo e implementación es la falta de personal capacitado para el manejo de lenguaje de programación pues por el poco acceso al conocimiento que se tiene a esta herramienta y sus implementaciones es necesario la migración a otros países para poder adquirir dichas habilidades con lo cual es complejo para las empresas poder encontrar personal calificado, Rodríguez establece que “la divergencia entre la oferta existente y las necesidades que requieren las empresas se hace especialmente relevante en el campo del aprendizaje automático” (Rodríguez, 2020). Por otro lado se encontró la falta de reconocimiento que se le da, las empresas y directivos no se atreven a realizar innovaciones con lo cual adquieren una falta de comprensión con respecto a la diversidad de usos que tiene el machine learning sumado a la gran relevancia que tiene la clasificación de clientes en la creación de estrategias para las industrias.

Dentro del contexto nacional se puede denotar que el Ecuador se encuentra más rezagado con respecto a los demás países en la región; la implementación de las IA se ve en su mayoría realizado por pocos profesionales y aunque se han desarrollado campañas para poder aumentar el interés por estas nuevas tecnologías su mala implementación, así como la falta de recursos no permite que la población en general pueda continuar adquiriendo conocimientos. Es así también como aun aquellos que puedan solventar la capacidad de poder adquirir cursos o estudios dentro de esta rama, en su gran mayoría aquellos que son ofrecidos, no satisfacen las necesidades mínimas para que se puedan desarrollar eficientemente programas de aprendizaje.

A lo largo del país el enfoque por el aprendizaje automático y su direccionamiento es lento y poco progresivo debido a que aún no existe una cantidad de industrias significativa que lo implemente como por ejemplo en el sector empresarial para el establecimiento de nuevas estrategias o la segmentación de clientes para poder implementar enfoques más adaptados a los requerimientos de los diferentes sectores del mercado en el que se encuentran “debido a que no se solucionan los requerimientos y problemas del cliente en tiempo óptimos acompañados de una mala atención incumpliendo protocolos de satisfacción. Esto genera inconformidad por parte del cliente obligándolos a finalizar la relación que mantengan con la empresa” (Castro, 2019). Por aquello es sin duda fundamental dentro del país y para las empresas la implementación, aprendizaje y extensión de conocimiento del machine learning y la IA así como el manejo de la estadística a la par para una interpretación eficaz puesto que es muy poco el desarrollo y demasiado costoso para las empresas las pocas opciones que se encuentran actualmente en el mercado ofreciendo soluciones no satisfacen por completo la necesidad general.

Problemática

En la actualidad, nos encontramos en un entorno empresarial complejo y dinámico; en donde la innovación y la tecnología tienen un rol importante para el crecimiento de las empresas proporcionando una mejora de su desempeño. ¿Qué es lo que una empresa necesita para tener éxito en este ámbito etéreo? necesita considerar los patrones en las necesidades de los clientes, examinar si su producto y/o servicio tendrá un público objetivo, evaluar la innovación de los productos, entre otros. en donde las empresas para examinar estas situaciones recurren al análisis de datos (Diaz et al., 2021).

Agregando a lo anterior la capacidad de una empresa de diferenciarse con respecto a su competencia radica principalmente en la calidad del servicio que ofrezca a sus clientes, en donde se ha reflejado la importancia de la automatización de procesos mediante el uso de herramientas tecnológicas que permitan una mejora en la toma de decisiones en base al análisis de la información (Tejada y Romero, 2020).

Sin embargo, de momento sigue representando un gran desafío la aplicación de estos conceptos ML, principalmente por el nivel de dificultad que conlleva su manejo y entendimiento y factores prácticos como son el coste, el personal, la formación y la postura de los empresarios; donde a pesar de lo citado figuran como una gran oportunidad en las economías globales. Además, en los distintos sectores existe una alta demanda de estos recursos debido a la competitividad en los mercados, en conjunto con el hecho de que la digitalización de los procesos se ha vuelto una necesidad para las empresas mundiales, es decir se deben adaptar a la nueva realidad (Rosero y Muentes, 2016).

Como se mencionó anteriormente el propósito de todas las empresas y la fuente principal de sus ingresos es la venta de sus bienes o servicios, actualmente esto

representa una problemática que enfrentan la mayoría de las compañías en los mercados debido al incremento de la competencia local e internacional, ocasionando un enfrentamiento por el posicionamiento dentro del sector, en donde el pilar fundamental son los clientes debido a que representan los beneficios obtenidos a través de la comercialización de los productos o servicios (Cartagena, 2017).

Es por esta razón que la fiabilidad de los clientes hacia la marca requiere de mucho tiempo y esfuerzo los cuales suponen un gasto para las compañías por lo que cada vez que una estrategia no es efectiva supone una pérdida de dinero, pues los recursos invertidos en intentar atraer a ese cliente no se han transformado en nuevos ingresos, en donde para solucionar este problema hay que analizar de manera detallada todos los datos posibles de la base de clientes, para poder definir las estrategias correctas que vayan acorde a su clasificación y a su vez poder centrarse en los clientes potenciales y más rentables para las empresas (Cartagena, 2017).

Es decir en el entorno actual se tiene que estudiar la mayoría de datos de los clientes, lo cual representa una tarea compleja para ser realizada por una persona por lo que hay la necesidad de automatizar el proceso de clasificación de clientes, en donde las ciencias de la computación y más en concreto la inteligencia artificial tiene mucho que aportar en el ámbito empresarial debido a la presencia de los nuevos paradigmas en el tratamiento de la información y el procesamiento masivo de datos, destacando la tendencia hacia la globalización de los mercados, en donde el riesgo y la incertidumbre en la toma de decisiones ha contribuido al avance progresivo del mismo, en conjunto con la complejidad del mundo actual en donde la exactitud ya no existe y se tiene que ser más competitivo y sustentable a través del tiempo para lograr la longevidad de las empresas (Valverde, 2019).

De modo que dentro de la IA, la rama de la disciplina de aprendizaje automático (Machine Learning) es una posibilidad ante los acontecimiento dentro de las empresas en factor a que este enfoque hace uso de algoritmos para analizar registros en bases de datos internas de los clientes de la compañía, para así descubrir ciertos patrones, interacciones o reglas que pueden explicar o predecir las tendencias futuras que pueden avisar cualquier oportunidad competitiva (Mena,1996). Ayudando a la toma de decisiones o el mejoramiento de la comprensión o conocimientos que se pueda extraer a través de las bases de datos (Gutiérrez et al., 2017).

Igualmente el ML va a permitir la extracción de conclusiones a través del análisis estadístico de los datos introducidos, mediante un proceso de mejora automática conforme se incorporan más datos al algoritmo, de modo que la aplicación de técnicas de ML puede dar lugar a mejoras de eficiencia, reducción de costes, incremento de calidad, aumento del nivel de satisfacción de los clientes, etc., en razón a las posibilidades que ofrecen al automatizar procesos operativos e incrementar las capacidades analíticas de las empresas (Fernández, 2019).

Dicho de otra manera, mediante las técnicas del ML, las entidades pueden también automatizar tareas repetitivas es decir que aportan un menor valor agregado a la compañía. En consecuencia, el nivel de eficacia de las estrategias de los clientes aumenta, al recibir un enfoque para cada clasificación de clientes. Por otra parte, las herramientas de IA permiten a las empresas analizar un gran volumen de datos (tanto estructurados como no estructuradas) y de manera más eficaz, así mismo el aumento en la cantidad de variables conduce a análisis de una mayor calidad dado que se consigue conocer mejor al cliente y obtener resultados más precisos para la aplicación de las estrategias por cada segmento obtenido de la clasificación de clientes (Fernández, 2019).

Por último, el aumento de las capacidades analíticas permite a las empresas explotar toda la información a la que pueden acceder sobre sus clientes, tanto interna como externa, a fin de conocer mejor sus preferencias mediante la capacidad de detectar características que le han llevado a clasificar bien a los clientes y por el contrario excluir las características que generalmente le provocan un error en su clasificación.

Problemática: Empresa

Actualmente en la empresa Labelprintsa S.A. existe una problemática la cual radica en que dentro del mercado de operación ha aumentado la competencia local lo que ha provocado un gran reto y dificultad en una correcta planificación de ventas de los productos y servicios brindados, en donde un factor clave para hacer frente a este problema es tener un conocimiento claro de la segmentación de clientes de la compañía, por eso se decidió realizar una clasificación de clientes como punto de partida en cuanto a una mejora en la implementación de planes de acción exitosos en el departamento comercial (Labelprintsa, 2021).

A pesar de todo la empresa sigue teniendo como visión consolidarse como una compañía líder en el mercado ecuatoriano, elevando la competitividad de sus servicios a nivel internacional, teniendo en consideración sus inicios como un proyecto de inversión de adquisición de máquinas flexográficas y tecnología innovadora en el sector industrial. En resumidas cuentas, por eso es necesario el estudio de los clientes para determinar las estrategias óptimas a implementar, en donde el gerente de la empresa el Ing. Xavier Matamoros vio la necesidad de automatizar los procesos de la compañía, y es aquí donde entran las ciencias de la computación, más concretamente la inteligencia artificial que tiene mucho que aportar en este sentido, sobre todo porque

lleva años ayudando al mundo empresarial en la gestión de sus negocios (Cartagena, 2017).

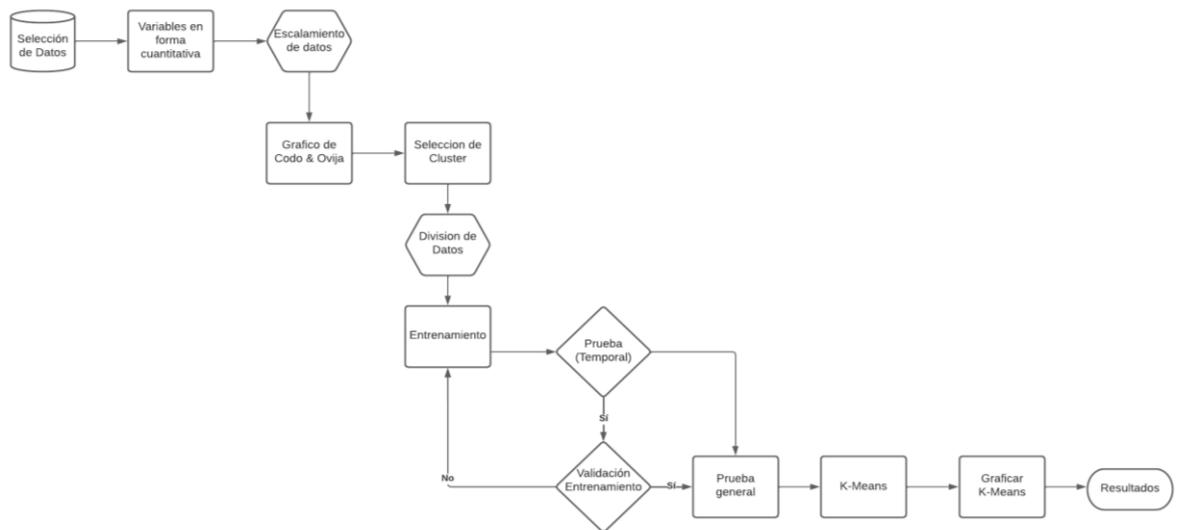
Por otro lado, un punto a resaltar es que la empresa en los últimos años ya ha implementado mejoras tecnológicas en sus procesos internos como son el área contable a través del sistema online “Contifico”, sin embargo para la realización de análisis o pronósticos de la información de las bases de datos siguen haciendo uso del Excel debido a la facilidad del manejo de esta herramienta, sin tener en consideración que se basan en predicciones que tienden a tener incertidumbre considerables para no llegar al resultado esperado.

Por el contrario el uso de la IA con su rama el Machine Learning representa un gran avance tecnológico y de análisis estadístico para la toma de decisiones de la compañía, principalmente con la problemática actual debido a la eficiencia en el análisis de datos a través de los diferentes modelos de clasificación, expresando un análisis efectivo de los datos de forma que se pueda resolver la incertidumbre o la falta de viabilidad cuando se toman decisiones tan explícitas como la definición de estrategias para los clientes. Además, mediante la implementación del MC se puede obtener una ventaja competitiva frente a la competencia en el mercado, debido a una visión más realista de la situación de la empresa en el presente con el mínimo error para asegurar una mejora en la toma de decisiones (Issuu, 2022).

A continuación, se tiene el procedimiento que se maneja en el procesamiento de la base de datos en factor del algoritmo de agrupamiento (Clustering) de Machine Learning.

Figura 1

Algoritmo de Machine Learning - Clusterización



Nota: Diagrama del proceso del algoritmo de Clusterización por Machine Learning

Justificación

La presente investigación se enfoca en el desarrollo de un modelo de clasificación de clientes a través del machine learning por clusterización para la empresa Labelprintsa ya que la empresa ha establecido un requerimiento para poder establecer estrategias acordes a grupos específicos de clientes debido al aumento de la demanda de métodos de impresión para los productos por parte de las empresas. Según García “Con el aumento de la producción digital, las etiquetas y envases son ahora una parte integral de la marca y de toda la experiencia de compra.” (2022). A la par se tiene como objetivo no solo el denotar la importancia de la implementación de este, sino también su funcionalidad en conjunto de la relevancia de una correcta segmentación de mercado, que aporta a las empresas de toda índole con el objetivo de obtener un crecimiento sustancial.

La aplicación del machine learning y el desarrollo de un modelo acorde a las exigencias de Labelprintsa se ve justificado debido al crecimiento de esta la cual en sus últimos aspectos financieros destacados reportó aumento de ingresos netos del 4,35% en 2021 (EMIS, 2022). Dicho crecimiento representa una masiva generación de datos en conjunto con un margen más amplio de clientes a los cuales se les ofrece servicios, así como potenciales nuevos clientes en los cuales deben poder aplicar estrategias de diferentes índoles para la retención o adquisición de los mismos, las cuales deben estar focalizadas y adaptadas a las posibles necesidades de ciertos grupos de clientes “el éxito actual está orientado a quienes invierten sus recursos en su activo digital de mayor importancia: sus datos” (Marketing Directo, 2019). Es así como esta nueva generación de datos requiere un manejo lo más exacto únicamente posible a través del ML a través de un algoritmo de Clúster por k means neighbours (los k vecinos más cercanos) que con el correcto desarrollo y ajuste puede lograr un mínimo error en la predicción de clasificación de grupos de clientes que se traducirá en mejores estrategias de campañas y un posible aumento en la cuota de mercado a la cual tiene disponibilidad la empresa.

La implementación de esta tecnología es vital, con esto se espera poder solventar un argumento válido a través de su aplicación para que el machine learning pueda ampliar su uso y desarrollo en las diferentes compañías del Ecuador, así como de la región estableciendo las diferentes ventajas y facilidades que puede proveer a cualquier empresa con un correcto desarrollo. Los modelos de aprendizaje automático son muy versátiles y útiles en el mundo empresarial. Su capacidad de adaptación al cambio y el aprendizaje le permiten predecir variables futuras que han logrado tasas de éxito superiores al 90%. Por lo tanto, esta tecnología tiene un enorme potencial para el futuro (Coelho, 2022). En este sentido la herramienta propuesta permitirá generar

mayores eficiencias en los procesos de los diferentes departamentos dentro de la empresa al poder focalizar sus objetivos permitiendo de esta manera reducir gastos.

Alcance

La aplicación del Machine Learning en las empresas ha evolucionado en las últimas décadas, con el crecimiento tecnológico en donde las compañías encontraron una nueva forma de automatizar los procesos, con respecto a la clasificación permite la transformación y evaluación de clasificar los datos. Por ejemplo, las empresas dedicadas a la comercialización de servicios utilizando chatbots que no necesitan de la intervención humana con la finalidad de ayudar en la atención y captación de clientes; interacción o fidelización; ventas y comercio online que almacenan datos para uso empresarial, sin embargo, en el Ecuador no muchas compañías han implementado el ML por su falta de conocimientos debido a los beneficios de su uso (Intriago y Villacis, 2022).

El tipo de investigación es descriptiva y referencial y va dirigido a pequeñas, medianas y grandes empresas de Ecuador que no realizan el estudio idóneo con respecto al análisis de datos, además de estudiantes y expertos de BI que necesiten realizar un método más preciso del empleado actualmente, en donde los aportes conceptuales van desde un mejoramiento en el manejo de los datos y aportes a la clasificación de procesos que las empresas realizan continuamente.

Objetivos

Objetivo General

Ejecutar un modelo de machine learning para predecir la clasificación de clientes y tomar decisiones estratégicas sobre el pronóstico del comportamiento futuro de los tipos de clientes.

Objetivos Específicos

- (a) Analizar los conceptos y teorías del uso del machine learning en la clasificación de clientes mediante la revisión de la literatura
- (b) Analizar la metodología de clusterización supervisada para la aplicación de conceptos de Machine Learning
- (c) Evaluar los resultados de la implementación del algoritmo de clusterización y su incidencia en el proceso de clasificación de clientes.

Capítulo II: Fundamentación Teórica

Marco Teórico

Para comprender la utilidad de la implementación del machine learning en las empresas, se debe tener en consideración cuales son las teorías que se encuentran en torno, y que servirán como base para la realización de la investigación, explicando de esta manera los conceptos más relevantes, así como haciendo hincapié en la aplicación del ML enfocado en la clasificación, con el fin de entender el tema en cuestión.

Inteligencia Empresarial

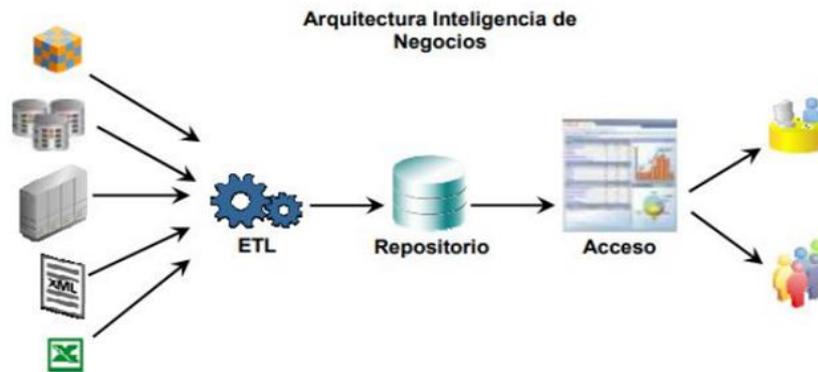
Inteligencia Empresarial o Business Intelligence (BI) se define como la habilidad para la toma de decisiones administrativas, mediante la transformación de los datos en información, y a su vez la información en conocimiento esencial para poder establecer estrategias y herramientas que permitan mejorar u optimizar las actividades y la toma de decisiones dentro de la empresa. Es así como la ejecución de su funcionalidad se combina con las diversas metodologías y tecnologías que permiten reunir, depurar y transformar datos para obtener un análisis con mayor precisión, lo que permite a las empresas tomar mejores decisiones (Gartner, 2011).

Es así como el BI actúa como un factor estratégico para las empresas u organizaciones, ya que genera una potencial ventaja competitiva, que no es otra cosa más que proporcionar información única para responder a los problemas del negocio

ya puedan ser estos: la introducción a nuevos mercados, optimización de procesos, planificación de la producción, análisis de perfiles de clientes, rentabilidad de productos y servicios, etc.

Figura 2

Arquitectura Inteligencia de Negocios



Nota: Recuperado de Inteligencia de Negocios. Aurelio (2021).

ETL (Extract, Transform, Load)

Según International Business Machines Corporation (2020), ETL es un proceso que extrae, transforma y carga datos de múltiples fuentes a un almacén de datos u otro repositorio de datos unificado, en donde proporciona la base para el análisis de datos y los flujos de trabajo de aprendizaje automático.

Extract

Se comienza con la extracción de los datos, en donde los datos sin procesamiento se copian o exportan desde las ubicaciones de origen a un área de preparación. Es aquí donde los equipos de gestión de datos pueden extraer datos de una variedad de fuentes de datos, que pueden ser estructurados o no estructurados.

Transform

En este proceso es donde los datos sin procesar previamente se someten al procesamiento de datos, es decir los datos se transforman y consolidan para su caso de uso analítico previsto

Load

Por último, tenemos en donde los datos transformados se mueven desde el área de procesamiento a un almacén de datos de destino, que por lo general implica una carga inicial de todos los datos, seguido de una carga periódica de cambios de datos con una menor frecuencia, en conjunto con actualizaciones completas para el borrado y reemplazo de datos en el almacén.

Inteligencia Artificial

Según Stubblefield y Luger (1990) La inteligencia artificial forma parte de las ciencias computacionales y se encarga de la automatización de la conducta humana. Es decir, se trata de un sistema informático de interpretación y aprendizaje de datos, centrado en la creación de sistemas capaces de ejecutar tareas de forma automática, ya sea estos máquinas o procesadores de software análogos al cuerpo, el cerebro y mente humanos, los cuales buscan reconocer patrones y tendencias, generar predicciones precisas para mejorar las tareas y lograr objetivos específicos.

Es así como la inteligencia artificial ha destacado debido a su capacidad de adaptación y sus diversas aplicaciones prácticas en la actualidad ideal para automatizar procesos, sobre todo por la facilidad de uso de su programación con excelentes resultados. Según Rusell (2019) las diferentes aproximaciones de la IA se encuentran dada en 4 enfoques aplicados descritos en su libro *Human Compatible. AI and the Problem of Control*:

- (a) Búsqueda de soluciones: Se definen como los enfoques de exploración de árboles (look ahead) (b) Conocimiento y lógica: Se refiere a la lógica proposicional, lenguajes formales y sistemas expertos (c) Incertidumbre y probabilidad: Aquí están los lenguajes probabilísticos de primer orden, redes bayesianas y la combinación de la teoría de la probabilidad con los lenguajes formales (d) Aprendizaje a partir de la experiencia: Es el aprendizaje basado en casos y aquí es donde se encuentra su rama machine learning.

Además, existen diferentes ramas de la Inteligencia Artificial tales como: Machine Learning, Deep Learning, Neural Networks, Computer Vision, Natural Language Processing (NLP), Natural Language Generation (NLG), Chatbot, Virtual Digital Assistants, Recommender Systems y Predictive Analysis.

Figura 3

Inteligencia Artificial – Usos



Nota: Recuperado de Aplicaciones de la Inteligencia Artificial. Liz (2021).

Big Data

Esta disciplina “Big data” hace referencia a la gestión de grandes cantidades de datos, tanto estructurados como no estructurados, cuyas características dificultan su almacenamiento, gestión, procesamiento y análisis, siendo sus particularidades

principales: el volumen (cantidad de datos), la velocidad (procesos necesarios para el tratamiento de los datos de forma ágil), la variedad (tener varias fuentes de recopilación de datos), la veracidad (que tan acertada es la data), el valor y la variabilidad (diferentes interpretaciones que pueden haber en el proceso). En otras palabras, estos datos son tan extensos que los software y herramientas convencionales no pueden gestionarlos (Jiménez, 2014).

Es así como al momento de recoger, procesar y analizar Big data, nos encontramos con datos operativos y analíticos, donde cada uno se almacena según distintos criterios o pautas, en donde los sistemas operativos tratan con grandes bases de datos con múltiples servidores, mientras que, los sistemas analíticos, son capaces de desarrollar análisis de datos más complejos (Casas, 2018).

Del mismo modo tenemos el análisis de big data el cual se da mediante el uso de diferentes técnicas analíticas avanzadas contra la gran cantidad de datos, en donde su comprensión permite impulsar una mejor y más rápida toma de decisiones, modelado y predicción de posibles resultados futuros además de una inteligencia comercial mejorada mediante la extracción de información significativa como pueden ser patrones ocultos, tendencias del mercado y preferencias de los clientes. Es aquí donde su aplicación da apoyo en los diferentes trabajos de investigación con respecto a bases de datos muy grandes, ya que permite analizar el comportamiento del cliente, predecir tendencias futuras para tomar mejores decisiones comerciales, mejorar las campañas de marketing y aumentar la eficiencia operativa, etc (International Business Machines Corporation, 2022).

Figura 4

Big Data Analytics



Nota: Recuperado de Artificial Intelligence & Machine Learning. Brain (2020).

Aprendizaje Automático (ML)

Machine Learning o también llamado aprendizaje automático es un subcampo de las ciencias computacionales y una rama de la Inteligencia Artificial que permite a un sistema aprender de los datos, es decir utiliza grandes cantidades de datos para crear sistemas que aprendan de forma automática y posean la capacidad de resolver o predecir situaciones y comportamientos futuros sin la intervención humana, mediante la creación y alimentación de algoritmos (secuencia de pasos para realizar una tarea) con grandes bases de datos, identificación de patrones y análisis de datos (Cardenas, 2018).

Es así como una característica muy importante de estos algoritmos es la predicción de nuevos patrones basándose en la experiencia continúa aprendida del conjunto de datos utilizados para su entrenamiento, a este suceso se le conoce como generalización. Sin embargo, el machine learning no es un proceso sencillo, ya que conforme el algoritmo recibe datos de entrenamiento, es posible producir modelos más precisos basados en los datos

Figura 5

Aplicativos del aprendizaje automático



Nota: Recuperado de L&D and growth go hand in hand. Dreamlad (2022).

Mecanismos de aprendizaje

La diversidad del Machine Learning radica principalmente en los diferentes procedimientos en las que se puede aplicar, dado a que existen 2 métodos denominados aprendizaje supervisado y el aprendizaje no supervisado, los cuales son escogido en base a los datos a utilizar y en factor a la necesidad a cubrir (Saltos y Villacis, 2022).

En la presente investigación se empleará un algoritmo de k-means neighbourhood para el empleo de un clúster con el objetivo de clasificar tipos de clientes en la empresa LabelprintSA. EL algoritmo mencionado pertenece a los tipos de algoritmos de clasificación supervisada con lo cual se considera indispensable establecer las principales diferenciaciones y de esta forma se pueda reconocer y segregar una de la otra y sea más comprensible los posteriores procesos que se desarrollarán a lo largo de este proyecto.

Tabla 1*Aprendizaje supervisado vs no supervisado*

APRENDIZAJE SUPERVISADO VS NO SUPERVISADO	
SUPERVISADO	No supervisado
Se darán las variables de entrada y salida.	Sólo se proporcionarán los datos de entrada.
Datos de entrada	
Se entrenan usando datos etiquetados.	Se usan contra datos que no están etiquetados.
Complejidad computacional	
Es un método más sencillo.	Es computacionalmente complejo.
Uso de datos	
Utiliza datos de formación para aprender un vínculo entre la entrada y la salida.	No utiliza datos de salida
Aprendizaje en tiempo real	
El método de aprendizaje en tiempo real tiene lugar fuera de línea.	El método de aprendizaje tiene lugar en tiempo real.
Número de clases	
Se conoce el número de clases.	Se desconoce el número de clases.
Inconveniente principal	
La clasificación de grandes datos puede ser un verdadero desafío en el Aprendizaje Supervisado.	No se puede obtener información precisa con respecto a la clasificación de datos y la salida como datos utilizados en el Aprendizaje no Supervisado está etiquetada y es desconocida

Nota: Principales diferencias entre el aprendizaje supervisado y el no supervisado

Aprendizaje supervisado.

Según Aguirre (2019) El aprendizaje supervisado es un método de machine learning que entrena con una data en las que los resultados de salida (outcome, labels) son conocidos. Es decir, trabaja con datos “Etiquetados”, en donde consta de preguntas (características) y respuestas (etiquetas), con la finalidad de que los patrones sean detectados y etiqueten nuevos conjuntos de información. En este caso, la intervención humana es un pilar fundamental dentro de su desempeño ya que etiqueta, clasifica e

introduce los datos en los algoritmos. Este modelo de aprendizaje permite que los algoritmos aprendan y se alimenten de datos históricos, los cuales serán aplicados en entradas desconocidas para obtener la salida correcta, es decir predice el valor de salida (Rojas, 2020).

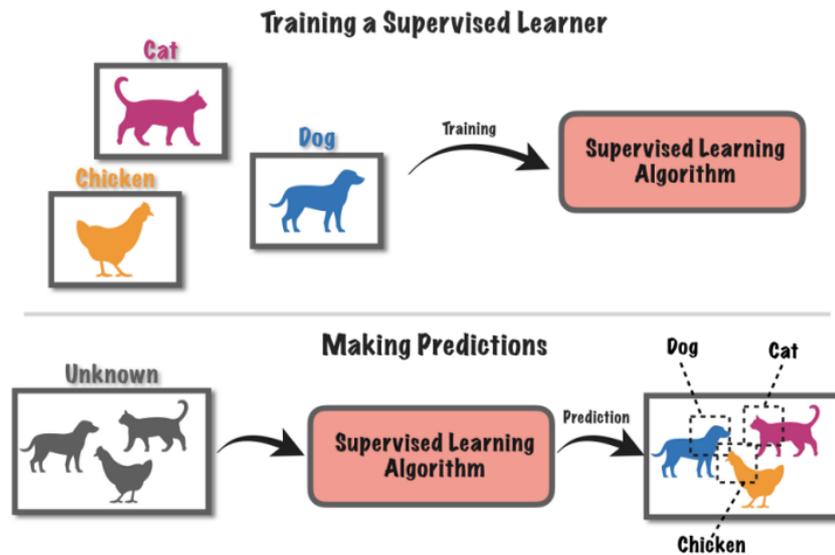
Este modelo de aprendizaje consta de dos pasos esenciales:

- (a) El entrenamiento: Los algoritmos y los datos, permite crear una hipótesis, es decir aquí se calibran los datos de forma que los algoritmos hayan aprendido sobre la información entregada. (b) La prueba: Se usa la hipótesis para generar predicciones o clasificaciones dependiendo del uso que se le dé al modelo

Teniendo como ejemplificación el uso del algoritmo supervisado para entrenar un conjunto de imágenes de animales, además de tener sus etiquetas pertinentes (perro, gato, pollo) en donde el algoritmo va a aprovechar las características de utilidad de las imágenes, como el número de extremidades o el color para encontrar patrones que puedan vincular las imágenes con las etiquetas correctas. Es así como luego de un entrenamiento exitoso, se puede utilizar el algoritmo entrenado para intentar predecir las etiquetas de un nuevo conjunto de imágenes.

Figura 6

Modelo de Aprendizaje Supervisado



Nota: Algoritmos de aprendizaje automático.

(Clark 2020).

Tipos de algoritmo de supervisión.

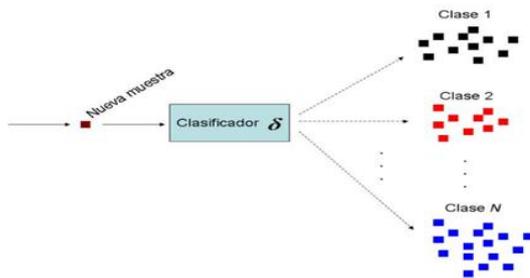
Algoritmo de clasificación supervisado.

Los modelos de clasificación predicen o explican los valores categóricos, en donde para ello necesitan de supervisión. En otras palabras, para clasificar utiliza la información que extrae de un grupo de datos, divididos en clases, en donde los algoritmos aplicados se alimentan y aprenden de la data, para luego clasificar las observaciones en varios grupos (Bautista, 2019).

Además, un dato importante es que cuando el modelo de entrenamiento es entre dos clases, se lo conoce como clasificación binaria, y si predicen más de 2 clases se lo determina clasificación multicategoría.

Figura 7

Modelo de clasificación supervisado



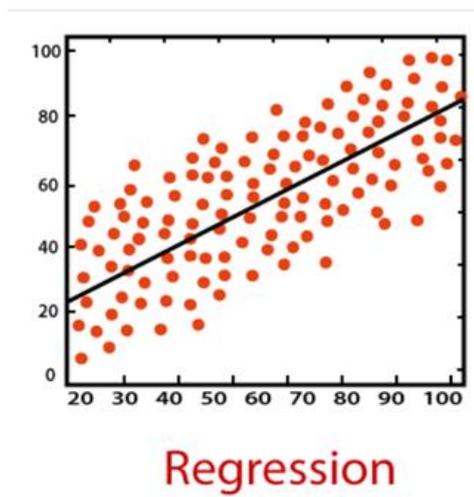
Nota: Tipos de Aprendizaje: Aprendizaje Supervisado de Machine Learning. Greyrat (2022).

Algoritmo de regresión supervisada.

El modelo de regresión supervisada se utiliza para poder predecir variables continuas, en donde la respuesta se puede predecir de manera flexible en función de las entradas del modelo, es decir el valor predicho se puede usar para determinar e identificar la relación lineal entre las cualidades. De este modo, en la regresión el algoritmo identifica una relación funcional entre los parámetros de entrada y salida (Gonzalez, 2018).

Figura 8

Modelo de regresión supervisado



Nota Types of Machine Learning Algorithms. Jones (2021).

Aprendizaje no supervisado.

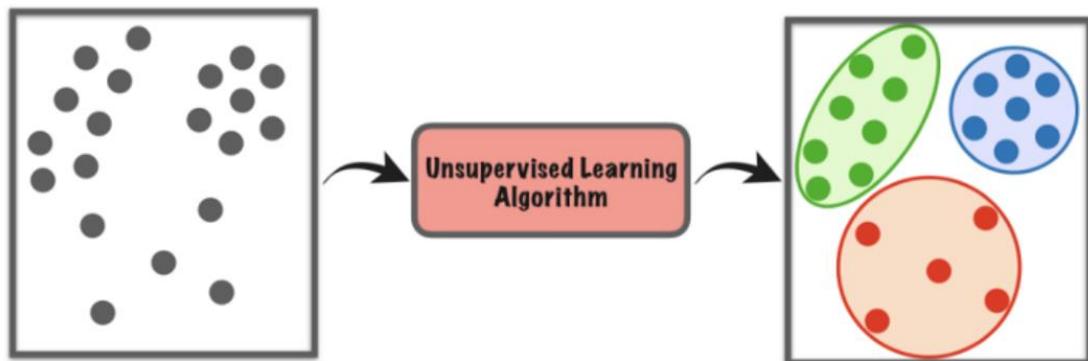
El aprendizaje no supervisado se centra en el aprendizaje en base a la experiencia, es decir detecta patrones de un conjunto de datos, que no posee información válida de los resultados conocidos o etiquetados. Este modelo de aprendizaje es muy útil dado que se aplica cuando la data no está compuesta de datos etiquetados, es decir datos no estructurados y al no tener necesidad de realizar un previo entrenamiento de los datos, este va a intentar descubrir las asociaciones correctas mediante sus igualdades o reconocimiento de patrones ya previos-visualizados, los cuales dará una deducción de los posibles resultados (Álvarez et al., 2020).

En este caso, no es necesaria la intervención humana para proporcionar instrucciones, en cambio la máquina va a determinar las relaciones mediante el análisis de los datos disponibles, de este modo se deja que el algoritmo de aprendizaje automático interprete los datos y los organice de alguna manera para describir su estructura o composición, además a medida que el modelo evalúa mayor cantidad de datos, su capacidad para tomar decisiones sobre los mismos tiene una mejora gradual y se vuelve más precisa (Saltos y Villacis, 2022).

Por otro lado, una de las problemáticas que el aprendizaje no supervisado intenta responder es la agrupación o clúster debido a que tiene como tarea encontrar los diferentes tipos de grupos entre elementos de los datos no estructurados provenientes de la base de datos, buscando la similitud para agruparlos mediante los patrones encontrados en base a su algoritmo el cual está programado para determinar el distanciamiento que se encuentran en los datos.

Figura 9

Modelo de Aprendizaje no Supervisado



Nota: Algoritmos de aprendizaje automático. Modelo no supervisado de Machine Learning (Clark 2020).

Tipos de aprendizaje no supervisado.

Agrupamiento.

El agrupamiento se centra en encontrar los patrones en un conjunto de datos no categorizados, además existen diferentes tipos de agrupamientos: (a) Exclusivo: Un dato puede pertenecer sólo a un clúster (b) Aglomerativo: Cada dato corresponde a un clúster (c) Solapamiento: Se implementan conjuntos difusos, donde cada punto va a más de dos clústeres (d) Probabilístico: Esta técnica de agrupamiento implementa una distribución probabilística para la creación de clústeres (Álvarez et al., 2020).

Asociación.

Esta técnica del modelo de aprendizaje no supervisado pretende encontrar las relaciones entre variables en grandes bases de datos.

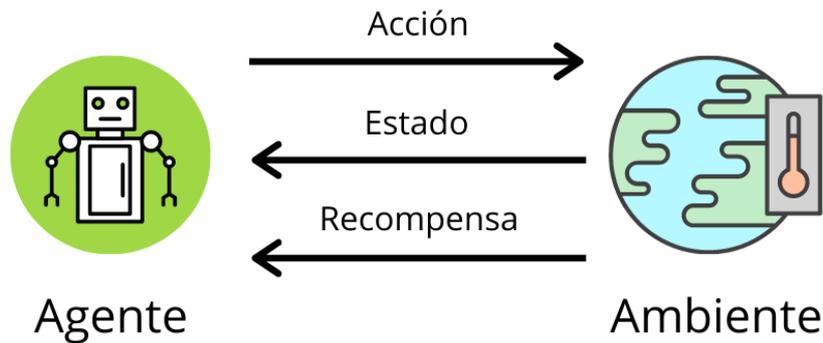
Aprendizaje reforzado

El aprendizaje reforzado se enfoca principalmente en el mejoramiento de la respuesta de los modelos usados mediante la implementación de procesos en base a la retroalimentación, se aplica cuando los datos del sistema no están etiquetados, pero en un periodo de tiempo determinado donde el sistema se irá retroalimentando con

actualizaciones. Este sistema funciona con un mecanismo de “Ensayo-error”, donde los inputs corresponden a los feedbacks que obtiene de los datos del exterior como respuesta a las acciones ejecutadas (Rojas, 2020).

Figura 10

Modelo de Aprendizaje Reforzado



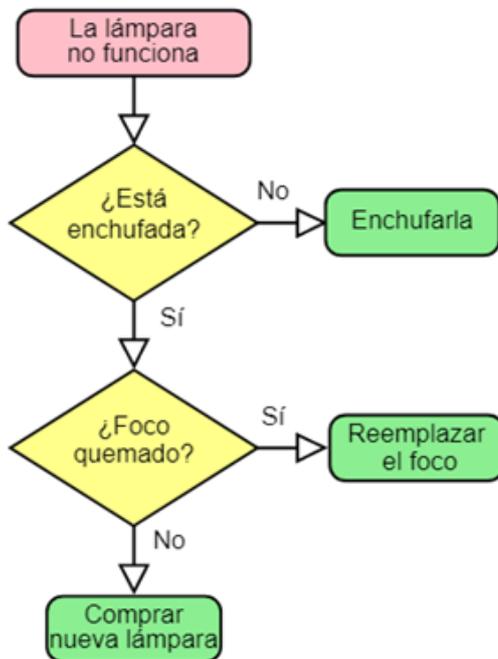
Nota: Aprendizaje por refuerzo. Nato (2020).

Algoritmos

Un algoritmo se puede definir como un conjunto ordenado y finito de instrucciones o reglas concretas y definidas, que deben llevarse a cabo para resolver un determinado problema en un procesador de datos. Los algoritmos constan de 3 partes: Input (Información de entrada), proceso, output (Resultados obtenidos). Así, el algoritmo refleja el análisis de varias variables y conectores que tendrán similitud en un campo, y ese campo será libre de tomar decisiones para habilitar las acciones en una determinada organización cuando la base de datos es muy extensa (Rivera y Villegas, 2022).

Figura 11

Funcionamiento de los algoritmos



Nota: Funcionalidad de los Algoritmos. Hamid (2018)

Clasificación

El algoritmo en cuanto a la clasificación procura establecer una etiqueta a cada uno de los datos culminando en categorías, y es mediante el entrenamiento de los datos más las etiquetas previamente mencionadas que se realizan los procesos de predicción del modelo, además es aquí también cuando al ingresar nueva data al conjunto se procede a realizar las predicciones y el etiquetado correspondiente, además la configuración logarítmica que maneja el método de clasificación es “Binaria”, es decir los datos son entre (0;1) o dicho de otra manera “Si y no”, de esta forma se etiquetan adecuadamente los datos en sus respectivas categorías (Saltos y Villacis, 2022).

La importancia que tiene este método en las aplicaciones son diversas, teniendo como ejemplos: En el momento que se desea realizar un mapeo de clientes que han realizado una compra y mediante el etiquetado que sea otorgado en el conjunto de datos se puede predecir si ese cliente tiene posibilidad de volver a comprar, o también

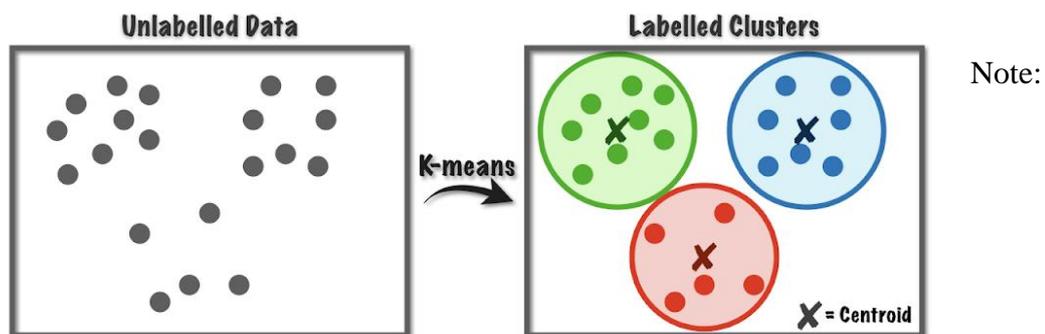
en la migración de clientes de algún servicio de suscripción, en donde con las debidas etiquetas se determinará si ese consumidor puede volver a adquirir el servicio o no (Theng, 2019).

Clustering por supervisado

La clusterización (clustering) es un tipo de método de aprendizaje supervisado del Machine Learning, que consiste en agrupar un conjunto de datos (no etiquetados) en subconjuntos de objetos llamados Clúster, donde la meta principal del proceso de clusterización es encontrar grupos que son diferentes de los otros, y sus objetos sean similares entre sí. Es decir, cada clúster está formado por una recopilación de objetos que son similares (o se consideran similares) entre sí, pero que son distintos respecto a los objetos de otros clústeres (Garcés, 2020).

Figura 12

Formación de un clúster



Supervised Learning with K-means. Sivasami (2021)

Dicho de otro modo, es una técnica de análisis exploratorio de datos el cual nos permite analizar los conjuntos de datos multivariable y tiene como objetivo agrupar los objetos de un dataset según su similitud de forma que los objetos que hay dentro de un grupo (clúster) sean más similares que aquellos que caen en grupos distintos. Además, es importante señalar que no hay criterios para una buena agrupación, más

bien depende del usuario y el criterio a utilizar para la satisfacción de su necesidad (Garcés, 2020).

Por otro lado, en función de cómo se relacionan los clústeres entre sí y con los objetos del conjunto de datos, podemos establecer una primera división entre los algoritmos existentes:

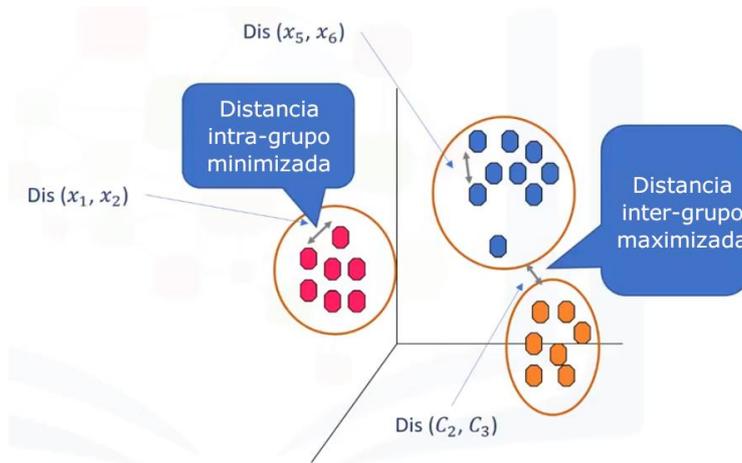
- (a) Clustering Duro: Se da cuando cada objeto pertenece a un solo clúster (por lo cual los clústeres pasarían a ser una partición del dataset)
- (b) Clustering Blando-Difuso: Se da cuando los objetos pertenecen a los clústeres según un grado de confianza o también conocido como de pertenencia.

K-medias por supervisión

K-Medias es un método de agrupamiento, que tiene como objetivo encontrar una partición de un conjunto de n observaciones en k grupos (parámetro del algoritmo), de forma que cada ejemplo pertenezca a una de ellas, concretamente a aquello cuyo centroide está más cerca a fin de minimizar la suma de los cuadrados dentro de cada grupo. De este modo k-medias tiene como dificultad principal la determinación del mejor valor para k (cantidad de clúster) el dónde su valor depende la forma y distribución del conjunto de datos. En síntesis, el algoritmo es relativamente eficiente y normalmente se requieren de pocos pasos para que el proceso se estabilice, pero en contra es necesario la determinación del número de agrupaciones y el sistema es sensible a la posición inicial de los k (clúster) seleccionados.

Figura 13

Algoritmo de agrupación por K-medias



Nota: Clustering. Agrupación por K-medias. Sandoval (2020)

Funcionamiento de la agrupación por K-medias

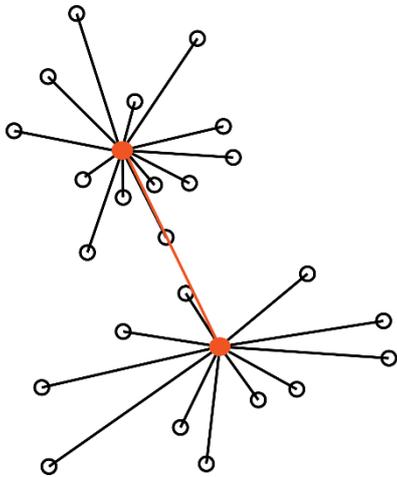
El funcionamiento del algoritmo de K-medias comienza con la determinación del número de particiones que pretendemos, en donde el factor clave de su función es que escoge de manera aleatoria un punto central para cada partición, el cual es denominado centroide de las particiones. Una vez dada la inicialización, que define el centroide se asignan las observaciones al centroide más cercano mediante el cálculo de la distancia de cada punto con respecto a este, en función de la naturaleza de los datos y de la finalidad de las particiones realizadas. Es así como su funcionalidad es la de minimizar la distancia de los puntos dados al centroide de su partición y maximizar la distancia hacia los centroides de las otras particiones. Por último, un punto importante a conocer es que K-medias es un algoritmo iterativo esto quiere decir que se repite los pasos hasta que el algoritmo converja, en base a la búsqueda de particiones con un error mínimo o particiones más consistentes.

Centroides

Un centroide se define como el centro de los clúster o segmentos, en donde la selección del número de centroides k se obtiene de la observación de las muestras o por métodos como del codo (elbow), el cual consiste en la ejecución del algoritmo clustering k -means para determinar un rango de valores de k para posteriormente calcular la varianza en la distancia del clúster con respecto a su centroide para la obtención del valor óptimo de K (Hurwitz y Kirsch, 2019, 16).

Figura 14

Visualización de un Centroide



Nota: Representación visual del centroide. Zelditch (2019).

Escalado de datos (Estandarización)

La escalabilidad de los datos se define como la capacidad de transformar los datos dando como resultado valores distribuidos de forma simétrica, además se debe tener en consideración que para aplicar un escalado se tiene dos enfoques diferentes: el escalado y la normalización.

- (a) Escalado: Transforma los valores de las características de forma que estén confinados en un rango normalmente entre $(0,1)$ o $(-1,1)$. (b) Normalización:

Transforma los valores de las características de forma que todas compartan un mismo valor medio y una misma desviación media

Por otra parte, es importante resaltar que el escalamiento deberá entrenarse solo con los datos de entrenamiento para evitar fugas, y que se tendrá que aplicar de igual manera tanto a los datos de entrenamiento como a los de validación y prueba (Zeilberger, 2020).

Marco conceptual

Segmentación de clientes

La segmentación de clientes corresponde a un factor fundamental en las empresas en el ámbito estratégico ya que, al igual que el mercado la cartera de una empresa tiene un comportamiento similar al no ser homogénea en su totalidad. Esta diversificación hace imposible poder implementar estrategias que abarquen a todo el conjunto, por lo tanto, se implementa una segmentación o clasificación de clientes en grupos donde los integrantes compartan características para lo cual poder focalizar las estrategias y campañas según cada segmento (Thompson, 2005). Esta puede enfocarse en diferentes aspectos que tengan mayor o menor valor para cada individuo del segmento a los cuales se les agrupará según la cantidad de valor que le den a determinado aspecto en común.

Existen múltiples definiciones correspondientes a la segmentación de mercado, según el diccionario de definiciones de la American Marketing Association (AMA) lo define como “el proceso de subdividir un mercado en subconjuntos distintos de clientes que se comportan de la misma manera o que presentan necesidades similares” (2020). A su vez Olamendi lo define como la “División del mercado en grupos más pequeños que comparten ciertas características” (2019). En ambos casos se establece

que la segmentación tiene como objetivo el poder asociar grupos de individuos a criterios específicos los cuales representen a los individuos de cada grupo. Esto aporta un mayor criterio que será de gran aporte hacia las empresas “la principal utilidad será adaptar la estrategia de actuación para cada uno de estos segmentos con el fin de que la empresa tome el mejor posicionamiento posible frente a dicho colectivo” (Díez, 2004).

Business Intelligence

El Business Intelligence (BI) o inteligencia de negocios se puede establecer como “un término general para la tecnología que permite la preparación de datos, la extracción de datos, la gestión de datos y la visualización de datos” (IBM, 2019). Sin embargo, se le aportan múltiples definiciones según cada autor pues a pesar de tener una gran trayectoria que se extiende hasta los años 50, no es hasta las últimas décadas que se le ha proporcionado una mayor relevancia y logrado expandirse alrededor del mundo. Como primera referencia al concepto podemos tomar la definición establecida en 1958 en el artículo “A Business Intelligence System” donde el investigador de IBM Hans Peter Luhn cual acuñó el término, refiriéndose a él como; “la Habilidad de aprehender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada” (Luhn, 1958, citado en Díaz, 2010, p.18). Sin embargo, no fue hasta 1989 cuando Howard Dresner quien es considerado como el padre de la inteligencia de negocios empresarial tomo la definición de Luhn a la cual le dio un tono más moderno definiéndose como “un proceso interactivo para explorar y analizar información estructurada sobre un área, para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones.” (Dresner, 1989, citado en Lluís-Cano, 2007, p.23).

Tomando como referencia las definiciones previamente mencionadas se puede establecer una más formal y generalizada que establece “El Business Intelligence es el conjunto de metodologías, prácticas y capacidades enfocadas al manejo de información que permite tomar mejores decisiones a las empresas” (Silva y Solano, 2017). A partir de esta información podemos establecer al BI como un concepto adaptativo a las necesidades y requerimientos de cada empresa a través de cual se logra el manejo de grandes cantidades de información de forma eficiente y rápida estableciendo patrones que permiten a su vez realizar inferencias al tomarlos como fuentes de información para las decisiones en la empresa.

El BI tiene se ha vuelto indispensable en el desarrollo de las empresas y se consolida como un factor diferenciador que permite obtener mejores resultados por sobre la competencia pues “el objetivo básico del Business Intelligence es apoyar de forma sostenible y continuada a las organizaciones para mejorar su competitividad, facilitando la información necesaria para la toma de decisiones” (Cano, 2007). El cual proporciona un gran aporte, su utilidad radica en la comprensión de las situaciones a través de 3 componentes principales: las fuentes de datos, el Data Warehouse y los Dashboard que permiten una interacción con el usuario idónea y una fácil interpretación sin requerir un conocimiento avanzado de data science (Silva y Solano, 2017).

Inteligencia artificial

Puede parecer que la inteligencia artificial tiene un corto desarrollo que se limita a la actualidad, sin embargo, su trayectoria trasciende más allá de las últimas décadas, remontándose a la segunda guerra mundial con la llamada “máquina de

Turing” se vislumbraron los primeros vestigios de los que ahora se denomina inteligencia artificial (IA) (Pérez, 2018).

Desde sus orígenes múltiples definiciones se le han atribuido a la IA, la primera mención fue establecida por el Dr. Marvin Minsky el cual define a la inteligencia artificial como “la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si fueran hechas por hombres” (Minsky, 1956, citado en Alfonso et al., 2003, p.4). Otra definición más específica que se le puede atribuir a la IA es la detallada por Pérez el cual reconoce a la inteligencia artificial “cuando una máquina es capaz de imitar las funciones cognitivas propias de la mente humana, como: creatividad, sensibilidad, aprendizaje, entendimiento, percepción del ambiente y uso del lenguaje” (2018). Sin embargo, no hay una definición establecida debido todo el rango de usos y aplicaciones que logra abarcar la inteligencia artificial, así como el desconocimiento que aún se tiene de la misma y que se va descubriendo a medida se van desarrollando nuevas metodologías. Aun así, se puede establecer una definición más detallada que abarque las previamente mencionadas y es que “la IA es la capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano” (Rouhiainen, 2018).

Con ello se puede establecer que la inteligencia artificial busca lograr que las máquinas emulen acciones humanas a la par aprendan y desarrollen los conocimientos adquiridos para aumentar dichas habilidades. En otras palabras, desarrollar inteligencia. Sin embargo, la inteligencia también es un término sujeto a la subjetividad, no posee una única dimensión si no que se extrapola a un espacio estructurado de capacidades para el manejo y procesamiento de información que será eventualmente utilizada, de la misma manera la IA emplea multitud de técnicas diferentes para resolver una gran variedad de tareas (Boden, 2018). Esta multitud de

técnicas se conglomeran en un enfoque principal que es el machine learning o aprendizaje automático, la rama de la IA que se encuentra a la vanguardia del estudio y desarrollo.

Machine learning

El machine learning (ML) o aprendizaje automático se remonta a los inicios de la IA, se desarrolló como una rama de la misma y se ha ido impulsando con el paso de los años su estudio e implementación, en la actualidad está viviendo un auge sin precedentes; a pesar de que su uso era minoritario varios factores están provocando que su uso sea más intensivo, con lo que el ML se ha posicionado como uno de los factores primordiales en el avance de diferentes industrias (Management Solutions, 2018).

Se puede llegar a establecer al aprendizaje automático como “una forma de IA que permite que un sistema aprenda de los datos en lugar de a través de una programación explícita” (Hurwitz y kirsch, 2018). Dicho de otra manera, el machine learning permite a las máquinas establecer patrones sin necesidad de una programación específica al aplicar algoritmos con relación a los datos para aprender de ellos continuamente. El mismo puede desglosarse en 3 tipos de algoritmos principales que se desglosan en: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo (Rouhiainen, 2018).

El machine learning ha ido creciendo conforme los años y se han ido desarrollando nuevas metodologías para lograr obtener el mayor potencial de esta subdivisión de la IA, como son las redes neuronales las cuales tiene un funcionamiento aún no comprendido en su totalidad pero que aun así han logrado emular el cerebro humano hasta cierto punto, pudiendo manejar grandes cantidades de datos que le da

nacimiento al aprendizaje profundo, el nuevo hito de la inteligencia artificial y el machine learning (IBM Cloud Education, 2021).

Matriz de confusión

También denominada matriz de error o tabla de contingencia, dentro del machine learning la matriz de confusión, es una herramienta fundamental con respecto a los algoritmos clasificatorios para poder estimar el error con respecto a la clasificación efectuada, la interpretación de la misma se realiza de forma horizontal, siendo que sobre las columnas se ubican los datos reales y sobre las filas la clasificación, aquellos elementos que se encuentren fuera de la diagonal serán aquellos que fueron mal clasificados y a partir de estos se podrá estimar el error en porcentaje existiendo dos tipo de errores; de omisión cuando se ubican elementos pertenecientes a una variable en otra y de comisión cuando elementos que son ajenos a una determinada clase aparecen en la misma (Sánchez y Muños, 2015).

Outliers

Dentro del análisis de datos, las bases de datos con las cuales se realizan los estudios e inferencias en su gran mayoría son datos que tienen determinada cantidad de datos que pueden afectar en gran medida los resultados estos datos son denominados Outliers. Los Outliers o datos atípicos son denominados dentro de la estadística como observaciones que son numéricamente distante de los datos y no representan de forma idónea al conjunto de data por lo cual llegando a cierto porcentaje deben ser depurados, siendo la mejor manera de identificarlos a través del diagrama de cajas también engloban a los datos faltantes o missing (Tejada, 2021). Estos datos se manejaron a criterio del analista puesto que “no pueden ser caracterizados categóricamente como benéficos o problemáticos, sino que deben ser contemplados

en el contexto del análisis y debe evaluarse el tipo de información que pueden proporcionar” (Ocaña y Peinado, 2018).

Marco Legal

En el ámbito legal, en el Ecuador existen normativas importantes a su vez se han generado varios planes alineados a una normativa legal entre los cuales se destacan: La Ley Orgánica de Telecomunicaciones con El Plan Nacional de Telecomunicaciones y Tecnologías de Información del Ecuador 2016-2021. El Convenio Marco de Cooperación Interinstitucional firmado con los diferentes entes como el Ministerio de Telecomunicaciones y de La Sociedad de la Información, la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación. Los Planes Nacionales para el Desarrollo (2013-2017). La Ley Orgánica de Telecomunicaciones. El Código Orgánico de La Economía Social de los Conocimientos, Creatividad e Innovación.

Así como los artículos que contemplan el desarrollo de la investigación científica y de la innovación tecnológica, de las TIC y diversas áreas del conocimiento que están en la Constitución de la República del Ecuador de los cuales se puede destacar los (Art. 281, Art. 385, Art. 387, Art. 423, Art. 388) 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.

Adicional a esto la Constitución de la República del Ecuador (2008) en su Art. 313 también señala que “El Estado se reserva el derecho de administrar, regular, controlar y gestionar los sectores estratégicos, de conformidad con los principios de

sostenibilidad ambiental, precaución, prevención y eficiencia” (pág. 133). Y por último la Constitución de la República del Ecuador (2008) agrega que en su Art. 321 de la Carta Magna, “el Estado reconoce y garantiza el derecho a la propiedad en sus formas pública, privada, comunitaria, estatal, asociativa, cooperativa, mixta, y que deberá cumplir su función social y ambiental” (pág. 149).

Capítulo III

Metodología

El actual trabajo es de corte transaccional en base a que se recolectarán los datos en un tiempo específico, en el cual la técnica de recolección de datos empleada para la realizar la investigación proviene de la data dura proporcionada por la empresa, que ha otorgado sus datos históricos sobre las ventas realizadas a su cartera de clientes en el 2022, la cual cuenta con 4074 observaciones en total dentro de la base, con el fin del desarrollo de un modelo de Machine Learning (Clusterización por supervisión) que permita clasificar los clientes y para crear mejoras en la toma de decisiones estratégicas por parte de la alta gerencia de la compañía.

Es así que dentro de los alcances pertinentes a la presente investigación, se escogió el alcance correlacional ya que pretende asociar variables mediante un patrón predecible para un grupo muestral o población. También tenemos que el enfoque es mixto (cuantitativo y categórico) en razón de las características de los datos obtenidos, los cuales se obtuvieron a través de una recolección de datos secundarios internos (Almacén de datos - Minería de datos), es decir son datos que están disponibles dentro de la organización que se está llevando a cabo la investigación. Para el desarrollo de este trabajo el análisis de datos a utilizar es la minería de datos, técnica que implica el uso de computadoras poderosas y programas estadísticos avanzados para analizar grandes bases de datos, con la finalidad de descubrir en ellos patrones desconocidos.

El objetivo base del estudio radica en la importancia de la implementación del machine learning en las empresas de los diferentes sectores económicos, para la obtención de una clasificación de clientes en base a las ventas es más eficaz a comparación de modelos tradicionales (modelos estáticos) que se siguen implementando en la actualidad. La investigación posee variables de tipo mixto (cuantitativo y categórico) debido a que se va a usar la recolección de datos con base en la medición numérica-categorica y el análisis estadístico para el establecimiento de patrones de comportamiento, además de que busca obtener respuestas específicas de la información recolectada de los datos, es decir la medición precisa de las variables de estudio con la finalidad de observar la situación actual de la compañía y a su vez comprobar si las decisiones que se han venido tomando en base a su histórico han sido la más eficiente. Para el desarrollo de esta investigación se va a plantear lo siguiente:

- (a) Estipulación de datos de investigación
- (b) Selección de enfoque de estudio
- (c) Selección de variables
- (d) Obtención de histórico de ventas y producción
- (e) Selección de Clientes
- (f) Delimitación de categorías de productos
- (g) Preparación del modelo adecuado
- (h) Valoración de los datos
- (i) Análisis de los resultados

Es así que es fundamental iniciar con la comprensión de la problemática presente y del entorno en que se encuentra, es decir en términos analíticos se aplica Business Understanding “Comprensión del negocio”, en donde dentro de esta fase se evalúan las posibles variables y se verifica su relación directa con el entorno empresarial, con la finalidad de determinar la fuente del desarrollo del problema, más aún cuando no se tiene conocimiento para así convertir este en la definición de un problema de minería de datos y una planificación preliminar para el alcance de los objetivos.

Por consiguiente, tenemos el Data Understanding, que es donde se da lugar a la recolección y preparación de los datos iniciales, aquí es común el realizar análisis exploratorios para poder familiarizarse con los datos obtenidos, y de esta manera detectar los posibles errores dentro de la base. En otras palabras, aquí se aplica estadística descriptiva y análisis de correlación con el fin de determinar la calidad del conjunto de datos y su nivel de relevancia para llevar a cabo el modelo.

Una vez se concluye con la fase de entendimiento de los datos, se procede a lo que se conoce como la Preparación de la data (Data Modeling), en donde se comienza con la elaboración de los datasets, es decir el conjunto de datos que sirven como base del modelo a implementar, su depuración, transformación y el entrenamiento de los algoritmos, etc. Asimismo, dentro de esta fase nos vamos a encontrar con data incompleta, normalización de datos, lo que nos indica que estos sucesos deben rectificarse para contar con cimientos buenos para la implementación del modelo.

Una vez realizada la implementación de las fases previas, se procede a la elección del modelo, dicho de otra manera consiste en la selección y aplicación del modelo de machine learning a realizar, el cual debe relacionarse de manera óptima a la necesidad primaria, además de selección las técnicas a implementar que van a servir de ayuda para llegar a las conclusiones del modelo con la mayor eficiencia posible, además es en este punto donde se emplea el algoritmo el cual aprende a medida que se lo emplea, quiere decir tiene un aprendizaje automático basado en la data histórico disponible con lo cual se vuelve más eficiente a medida que aprende de sus errores.

Por último, tenemos el análisis de los errores el cual es considerado un punto importante dentro de la creación del modelo. Sin embargo, podría ser obviado con una buena depuración de los datos, no obstante, si el error es muy elevado se trata de problemas directamente de aprendizaje del modelo de machine learning, y es así como

una vez satisfechos con los resultados, se procede a integrar el modelo mediante el lenguaje R y expresado en ambiente RStudio.

Metodología aplicada para realización del clúster

Para una correcta clasificación aplicando clúster se debe emplear una metodología que garantice el éxito de la clasificación y el algoritmo creado, la metodología aceptada es la siguiente:

Planteamiento del problema

Punto de partida: Si X es una muestra de n cantidad de individuos sobre los que se miden q variables. Entonces X se determina como un conjunto los cuales pueden ser ordenados en una matriz:

Figura 15

Ejemplificación de matriz

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mp} \end{pmatrix}$$

Nota: Iterative Method of Matrix. Khaladi (2018)

x_{11} : Valor del primer individuo en la primera variable

x_{12} : Valor del primer individuo en la segunda variable

x_{ij} : Valor que presente el individuo i-ésimo en la variable j-ésima

En cada una de las columnas se contienen cada uno de los valores que toman los individuos según cada una de las variables que se estudian.

Estandarización

Se aplica una estandarización para lograr una buena calidad de los clústeres y evitar problemas de unidades buscando crear variables de media 0 y dispersión 1 (Ollé,

2021). Es así como Amat Rodrigo establece “Si una variable tiene una escala mucho mayor que el resto, determinará en gran medida el valor de distancia/similitud” (2017). Esto permite asegurar que ninguna variable tenga un peso mayor con respecto a la realización del clúster.

$$\frac{x_i - \text{mean}(x)}{sd(x)}$$

Dibujar el matrixplot y la correlación entre características

Es importante hacerte una idea de las variables que están más relacionadas entre sí para saber qué variables pueden dominar los clústeres a su vez poder establecer la descripción de las variables cuantitativas a través de gráficos y modelados que permitan vislumbrar cómo podría desarrollarse el modelo.

Calcular el número óptimo de clústeres

Hay varias formas de medir la distancia entre conglomerados que producen diferentes grupos y diferentes dendrogramas. No existe un criterio para elegir cuál de los algoritmos es el mejor. La decisión suele ser subjetiva y depende de qué método refleja mejor los objetivos de un estudio en particular (Rodrigo, s. f.). Cuando se han agrupado dos objetos o grupos P y Q, la distancia del grupo a otro objeto R puede calcularse en base a las distancias entre los tres objetos o grupos de la forma siguiente:

$$d(R, P + Q) = \delta_1 d(R, P) + \delta_2 d(R, Q) + \delta_3 d(P, Q) + \delta_4 |d(R, P) - d(R, Q)|$$

Donde δ_j son constantes de ponderación. Los distintos métodos dependen de los valores que demos a las ponderaciones. En la tabla siguiente mostramos los pesos para algunos de los métodos más comunes.

Tabla 2*Métodos de distancia entre Clústeres*

Método	δ_1	δ_2	δ_3	δ_4
Salto mínimo	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Salto máximo	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
Media	$\frac{n_P}{n_P + n_Q}$	$\frac{n_Q}{n_P + n_Q}$	0	0
Centroide	$\frac{n_P}{n_P + n_Q}$	$\frac{n_Q}{n_P + n_Q}$	$-\frac{n_P n_Q}{(n_P + n_Q)^2}$	0
Mediana	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0
Ward	$\frac{n_R + n_P}{n_R + n_P + n_Q}$	$\frac{n_R + n_Q}{n_R + n_P + n_Q}$	$-\frac{n_R}{n_R + n_P + n_Q}$	0
Método Flexible	$\frac{1-\beta}{2}$	$\frac{1-\beta}{2}$	β	0

Nota: Métodos de Clúster mediante métricas de distancias. Llaguno (2019)

Donde n_R , n_P , n_Q denotan el número de objetos en cada uno de los grupos y β es un valor arbitrario entre 0 y 1

Método del centroide

El método del centroide utiliza como distancia entre grupos la distancia entre los centroides geométricos de cada grupo: (al cuadrado) distancia euclidiana entre ellos. Este método es, también, espacio-conservativo, pero presenta el inconveniente de dejarse influir excesivamente por los grupos de mayor tamaño (ceaces,2018).

A continuación, tenemos la medida de distancia euclidiano: Es la raíz cuadrada de la suma de las diferencias al cuadrado entre los valores de dos casos para cada variable

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ki} - x_{kj})^2}$$

d_{ij} : Distancia entre las observaciones i y k

X_{ki} : Valor de la variable X_k para el caso j

A continuación, tenemos el método de enlace de centroides, en donde la distancia se calcula con la siguiente matriz de distancia:

$$d_{mj} = (N_k d_{kj} + N_I d_{Ij})/N_m - (N_k N_I d_{kl})/(N_m^2)$$

d_{mj} : Distancia entre los conglomerados m y j

m : Conglomerado combinado que consta de los conglomerados k y I, con $m = (k, i)$

d_{kj} : Distancia entre los conglomerados k y j

d_{Ij} : Distancia entre los conglomerados I y j

N_k : Número de observaciones en el conglomerado k

N_I : Número de observaciones en el conglomerado I

N_m : Número de observaciones en el conglomerado m

Una vez obtenidos los centroides, encontramos la función de probabilidad que mejor se ajusta a los datos. Es decir, el algoritmo itera nuestros datos, modificando los parámetros vistos anteriormente (mu, sigma y la media) hasta encontrar el que mejor se ajusta (Alberto, 2021).

Método del vecino más próximo

k-means y Gaussian mixture models son los métodos más empleados para el cálculo de la optimización con respecto a la cantidad de clúster a implementar lo cual es fundamental para realizar una buena segmentación (Universidad Autónoma de Madrid, s. f.).

En el método del vecino más próximo la distancia entre dos clústeres es el mínimo de las distancias entre un objeto de un clúster y un objeto del otro.

$$d(R, P + Q) = \min (d(R, P), d(R, Q))$$

Sobre una matriz de distancias

Tabla 3*Matriz de distancia - Inicio*

		1	2	3	4	5	
Nota:	1	0.0					
Ejemplificación	2	1.0	0.0				
	3	5.0	4.5	0.0			
	4	8.5	7.8	3.6	0.0		del
inicio de una	5	7.2	6.7	2.2	2.0	0.0	matriz

de distancia. Todorova (2018)

Después de agrupar el 1 y el 2 en el clúster A, calculamos las distancias de A 3, 4 y 5

Tabla 4*Matriz de distancia 2.0*

		1	2		distan
3	5.0	4.5	$\min(5; 4.5)$		4,5
4	8.5	7.8	$\min(8.5; 7.8)$		7,8
5	7.2	6.7	$\min(7.2; 6.7)$		6,7

Nota: Mínimos y distancia de la matriz. Todorova (2018)

La matriz de distancias es entonces

Tabla 5*Matriz de distancia 3.0*

	A	3	4	5
A	0.0			
3	4.5	0.0		
4	7,8	3.6	0.0	
5	6,7	2.2	2.0	0.0

Nota: Fase preliminar de la obtención de la matriz de distancia. Todorova (2018).

La distancia más pequeña es entre 4 y 5, por lo que los fusionamos en un clúster que denominamos B, Calculamos la distancia entre B y el resto, es decir, Ay

3. Entre A y B, buscamos las distancias entre todos los pares de puntos y calculamos el mínimo

Tabla 6

Matriz de distancia 4.0

		B	
A		4	5
	1	8,5	7,2
	2	7,8	6,7

Nota: Obtención del valor mínimo en la fase de la matriz de distancia

El mínimo de los 4 valores es 7,2. La distancia entre B y 3 es 2,2

Tabla 7

Matriz de distancia 5.0

	A	B	3
A	0.0		
B	7.2	0.0	
3	4.5	2.2	0.0

Nota: Obtención de la unificación de los valores en la matriz de distancia

El valor más pequeño es 2.2, luego juntamos B con 3 en C, la distancia es la siguiente

Tabla 8

Matriz de distancia Obtenida

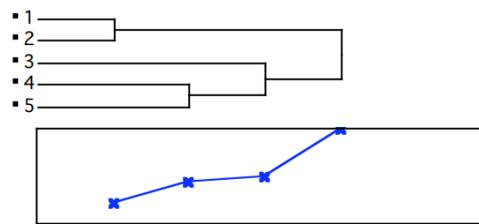
	A	C
A	0	
C	4,5	0

Nota: Resultante de la matriz de distancia. Todorova (2018).

Posteriormente se obtendrá el dendrograma

Figura 16

Dendograma



Nota: Gráfico de los resultados obtenidos de la matriz de distancia

Realizar el cálculo con distintos métodos

Es esencial realizar un cálculo variado de los clústeres a través de distintos modelos tales como k-means, GMM o hierarchical para poder realizar una validación y complementar el modelo y así garantizar una correcta partición evitando una superposición de los distintos clústeres, a la par poder determinar la técnica que mejor se adapte al caso

Comparación de los clústeres calculados

Por último, se debe comparar las características de los grupos que se han creado con la técnica seleccionada determinar si existen diferencias significativas entre grupos según las características y en qué variables se ven esas diferencias. Consiste en aportar una interpretación real a los clústeres que se han encontrado (Ollé, 2021).

Operacionalización de variables

La operacionalización de variables no es otra cosa que la separación de estas, con el propósito de definir la función que va a cumplir cada una en el modelo a utilizar en la investigación. Además, se identificaron las variables más relevantes para la creación del modelo de clusterización. Como complemento, se ha realizado data cleaning (limpieza de datos) debido a que existen datos atípicos que afectan la robustez del modelo, el resultado de esto son las 501 observaciones divididas en 9 variables:

(a) Mes (b) Clientes (c) Etiquetas (d) Impresoras (e) Lectores de barra (f) Materia Prima (g) Repuestos de máquinas (h) Suministros (i) Tinta Ribbons

El desglose para las variables es la siguiente:

A continuación, se tiene la imagen que muestra la operacionalización de variables en la base de datos de clientes de la empresa Labelprintsa S.A. en 2022:

Tabla 9

Operacionalización de variables

VARIABLES	Definición conceptual	Métrica	Indicador	Idoneo para el procesamiento?
Mes	Periodo de tiempo conformado por 30 o 31 días	Enero, Febrero, Marzo, Abril, Mayo, Junio, Julio, Agosto, Septiembre, Octubre, Noviembre, Diciembre	Momento en el que se llevo a cabo la compra	No
Clientes	Los clientes que compraron en 2022	Numerico	Es el numero de la cartera de clientes que han comprado	Si
Etiquetas	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si
Impresoras	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si
Lectores de barra	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si
Materia Prima	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si
Repuestos de maquinas	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si
Suministros	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si
Tinta Ribbons	Es una línea de la categoría de productos de la empresa	Numerico: sin decimal	Cantidad de compra por los clientes en el año	Si

Nota: Elaborado con información de la base de datos de compra de clientes de la empresa Labelprintsa S.A. del año 2022.

En la tabla se identificó 8 variables que aportarán valor al modelo, cliente y (Etiquetas, Impresoras, Lectores de barra, Materia Prima, Repuestos de máquinas, Suministros y Tinta Ribbons), que representan la categorización de los productos que maneja la compañía. La primera variable cliente es cualitativa, sin embargo, para su procesamiento en RStudio, se ha asignado el uso del código rownames para su

establecimiento como nombre de la fila del marco de datos en R debido a que es la variable de estudio del modelo, es decir la dependiente. Las demás variables son cuantitativas las cuales indican la cantidad de producto que compraron los clientes en el periodo 2022, esto significa que el total de cantidad comprada determina los clientes involucrados en cada línea de producto.

Metodología en RStudio

Para la implementación correcta de una clasificación mediante clúster en RStudio se debe emplear una metodología detallada que garantice el éxito en los resultados mediante el algoritmo planteado que va a permitir clasificar los datos, la metodología a seguir es la siguiente:

Carga de data

```
read.csv2(file, header = TRUE, sep = ";", quote = "\"\\\"\\\"", "dec  
= "\"", " , fill = TRUE, comment.char = "", ...
```

El código permite leer un archivo en formato de tabla y crea un marco de datos a partir de él, con casos correspondientes a líneas y variables a campos en el archivo.

file: Es el nombre del archivo del que se van a leer los datos

header: Es un valor lógico que indica si el archivo contiene los nombres de las variables en su primera línea.

sep: Se conoce como el carácter separador de campo

quote: Es el conjunto de caracteres de comillas, se usa para deshabilitar las comillas por completo

dec: Es el carácter utilizado en el archivo para los puntos decimales

fill: Función lógica, si es verdadero, en caso de que las filas tengan una longitud diferente, los campos en blanco se agregan implícitamente

... : Indica que se pasarán más argumentos a read. Table

En este punto también se utiliza lo que es el código:

```
attach (what, pos = 2L, name = deparse1 (substitute (what), backtick  
= FALSE), warn.conflicts = TRUE
```

Este comando hace que la base de datos se adjunte a la ruta de búsqueda de R, es decir que R busca en la base de datos cuando evalúa una variable, por lo que se puede acceder a los objetos en la base simplemente dando sus nombres.

what: “Base de datos” – En la cual puede ser un marco de datos, una lista o un archivo de datos R creado con guardar o NULL.

pos: Es el número entero que especifica la posición en la búsqueda () donde adjuntar

name: El nombre por utilizar para la base de datos adjunta

warn.conflicts: Función lógica, en donde si es verdadero, se imprimen advertencias sobre conflictos al adjuntar la base de datos, a menos que esa base contenga un objeto.

Instalación y carga de paquetes

```
Install.packages = ("name of package")
```

Este comando realiza la función de descargar e instalar paquetes desde repositorios similares a CRAN o desde archivos locales.

```
install.packages  
= (fpc, tidyverse, cluster, factoextra, NbClust, tidyr, writexl, dataexplorer)
```

Son un atajo para instalar fácilmente las colecciones de paquetes principales de cada código, el cual abarca tareas de ciencia de datos como: importación de datos, limpieza, manipulación, visualización y programación. A continuación, tenemos la explicación detallada:

Fpc package overview: Varios métodos para la agrupación y validación de clústeres.

Tidyverse: Es un conjunto de paquetes que funcionan en armonía porque comparten representaciones de datos comunes y un diseño de “API”, el paquete está diseñado para facilitar la instalación y carga de múltiples paquetes “tidyverse”

Factoextra: Es un método de impresión para un objeto de clase factoextra

NbClust: El paquete proporciona 30 índices para determinar el número de conglomerados y propone al usuario el mejor esquema de conglomerado a partir de los diferentes resultados obtenidos al variar todas las combinaciones de número de conglomerados, medidas de distancia y métodos.

Tidyr: Es un paquete que contiene herramientas para cambiar la forma (girar) y la jerarquía (anidar y desanidar) de un conjunto de datos y extraer valores de columnas de cadenas, además también incluye herramientas para trabajar con valores perdidos (tanto implícitos como explícitos) principalmente se encarga de realizar descriptivos.

Kmeans: Determina el número K de clusters que se van a generar y nstart, que determina el número de veces que se va a repetir el proceso, cada vez con una asignación aleatoria inicial distinta

DataExplorer: Simplifica y automatiza el proceso EDA y la generación de informes

Writexl: Crea un archivo Excel según una base determinada en la dirección establecida por el setwd

Ggplot2: Un sistema para crear gráficos 'declarativamente', basado en "La gramática de los gráficos".

```
library("fpc, tidyverse, cluster, factoextra, NbClust, tidyr, writexl")
```

Por último, tenemos la carga de las paqueterías instaladas, el código se encarga de verificar y actualizar la lista de paquetes adjuntos actualmente para su uso, y no se recargan un espacio de nombres que ya está cargado.

Verificación de variables cuantitativa y transformación a etiqueta

```
rownames(x, do.NULL = TRUE, prefix = now)rownames(x) < -value
```

El código se encarga de recuperar o establecer los nombres de fila o columna de un objeto similar a una matriz, a su vez tenemos sus argumentos:

x: Un objeto R similar a una matriz, con al menos dos dimensiones para los nombres de columna

do.NULL: Es una función lógica, en donde si es falso y los nombres son NULL, se crean los nombres

prefix: Es para la creación de los nombres *value*: Es un valor válido para el componente de *dinmnames* (función que opera tanto en filas como en columnas a la vez para establecer los nombres de objetos R).

Un punto importante es el uso de *row.names = NULL* fuerza la numeración de filas, es decir los nombres de fila faltantes o NULL generan nombres de fila que se consideran “automáticos”

Semilla

Permite predecir y reproducir la secuencia de los datos generados a través de un generador de número pseudoaleatorios, hace el código reproducible (Coder, 2020).

La semilla se instala mediante la función:

Setseed()

Escalamiento de los datos

scale (x, center = TRUE, scale = TRUE)

El uso del escalamiento se da con la función genérica *scale*, cuyo método por defecto centra y/o escala las columnas de una matriz numérica, en donde sus argumentos son los siguientes:

x: Es una matriz numérica (como objeto)

center: Explica que ya sea un valor lógico o un vector numérico similar de longitud igual al número de columnas de *x*

scale: Explica que ya sea un valor lógico o un vector numérico de longitud igual al número de columnas de *x*.

En conjunto tenemos el comando **head** el cual devuelve la primera o la última parte de un vector, matriz, tabla, marco de datos o función.

Estimación del número de clústeres

En este punto pasamos a hacer uso del código NbClust para poder determinar el número de conglomerados o clústeres

```
NbClust (data = NULL, diss = NULL, distance = "euclidean", min.nc = 2, max.nc  
        = 15, method = NULL, index = "all", alphaBeale = 0.1
```

data: Es la matriz o el conjunto de datos

diss: Es para definir la matriz de disimilitud a utilizar, de forma predeterminada
diss=NULL

distance: Es la medida de distancia que se va a utilizar para calcular la matriz de disimilitud, en este caso se aplica “euclidean”

min.nc: Es el número mínimo de clústeres, entre 1 y (número de objetos -1)

max.nc: Es el número máximo de clústers, entre 2 y (número de objetos -1), mayor o igual a min.nc. Por defecto, max.nc=15

method: Es el método de análisis de conglomerados a utilizar, en este caso “kmeans”

index: Es el índice por calcular, en este caso de clústeres “all”

AlfaBeale: Es el valor de significación para el índice de Beale

Cálculo de los clústeres

Mediante la función **Kmeans** del paquete stats se realiza la partición de los clústeres considerando los parámetros establecidos previamente con el K óptimo siendo la función:

```
kmeans(x, centers, iter.max = 10, nstart = 1, algorithm  
       = c("Hartigan – Wong", "Lloyd", "Forgy",  
          "MacQueen"), trace = FALSE
```

x: Matriz de datos escalados al cual se desea realizar la segmentación

centers: Cantidad de centroides determinados se por los K o un conjunto de centros de conglomerados iniciales distintos.

iter.max: Número máximo de iteraciones

nstart: si los centroides son un número, cuántos conjuntos aleatorios deben ser elegidos

algorithm: carácter: puede ser abreviado. Tenga en cuenta que "Lloyd" y "Forgy" son nombres alternativos para un algoritmo.

object: un objeto R de clase "kmeans", típicamente el resultado ob de ob = kmeans(..).

method: carácter: puede ser abreviado. "centers" causas ajustadas para devolver centros de clúster (uno para cada punto de entrada) y "classes" causas ajustadas para devolver un vector de asignaciones de clase.

trace: número lógico o entero, actualmente solo se usa en el método predeterminado

Por último, de igual forma los métodos de partición, como la agrupación en clústeres de k-means, requieren que los usuarios especifiquen la cantidad de clústeres que se generarán y aquí ingresa el código **fviz_bsclust ()** el cual determina y visualiza el número óptimo de clúster usando diferentes métodos: dentro de los clústeres, sumas de cuadrados y estadísticas de brecha.

Graficamos los clústeres

fviz_cluster ()

Para concluir tenemos la visualización de los resultados de agrupamiento, basado en ggplot2 de los métodos de partición, incluido kmeans y Mclust, donde las observaciones se representan mediante puntos en el gráfico y se dibuja una elipse alrededor de cada grupo.

Resultados

En los resultados arrojados por el modelo se determinan los siguientes aspectos claves que validan y respaldan la correcta ejecución de este y permiten establecer que direccionamiento tiene cada clúster y por lo tanto la preferencia del cliente perteneciente al conglomerado. Como primer aspecto se realiza un análisis descriptivo de la data a través del paquete “DataExplorer” el cual mediante el siguiente código:

```
DataExplorer::create_report(Clientes)
```

Crea un reporte automatizado HTML con información descriptiva general de las diferentes variables pertenecientes a la base de datos en donde se pudo obtener la siguiente información con respecto al dataset

Figura 17

Percentages of Dimension

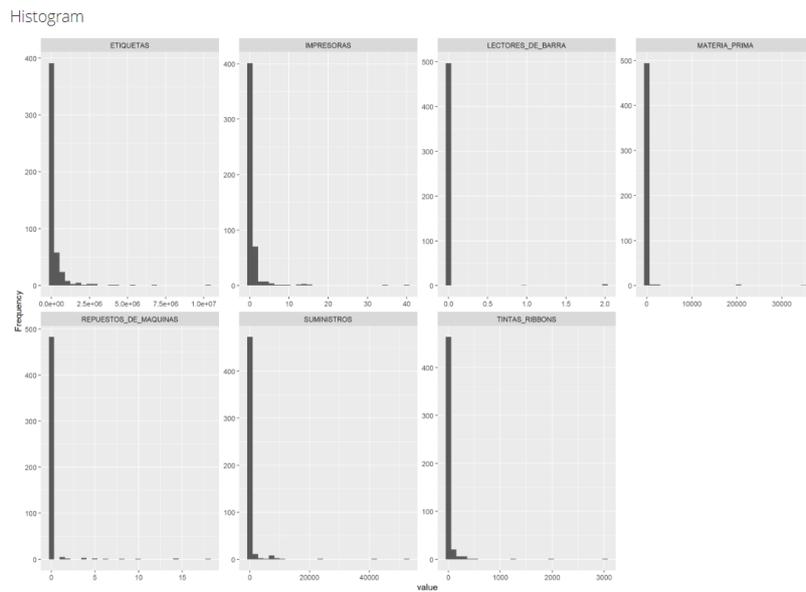


Nota: Reporte obtenido de la modulación en RStudio

Se determina que la base de datos está en condiciones para su empleo en el estudio al no tener datos perdidos y todas las observaciones con respecto a sus variables se encuentran completas por lo que no requiere un tratamiento adicional de depuración.

Figura 18

Histogramas

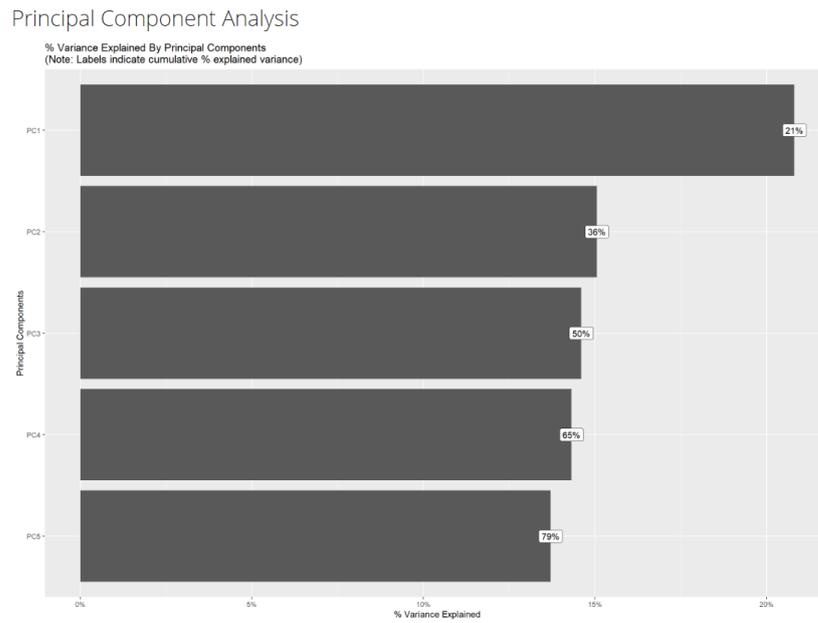


Nota: Reporte obtenido de la modulación en RStudio

El histograma con respecto a cada variable permite vislumbrar las tendencias de compras por categoría que tienen los clientes de la empresa siendo que en la mayoría de las categorías existen clientes que tienen bajas tendencias de compras, así como en contraste determinadas empresas que predominan en cada una de las variables.

Figura 19

Principal component Analysis

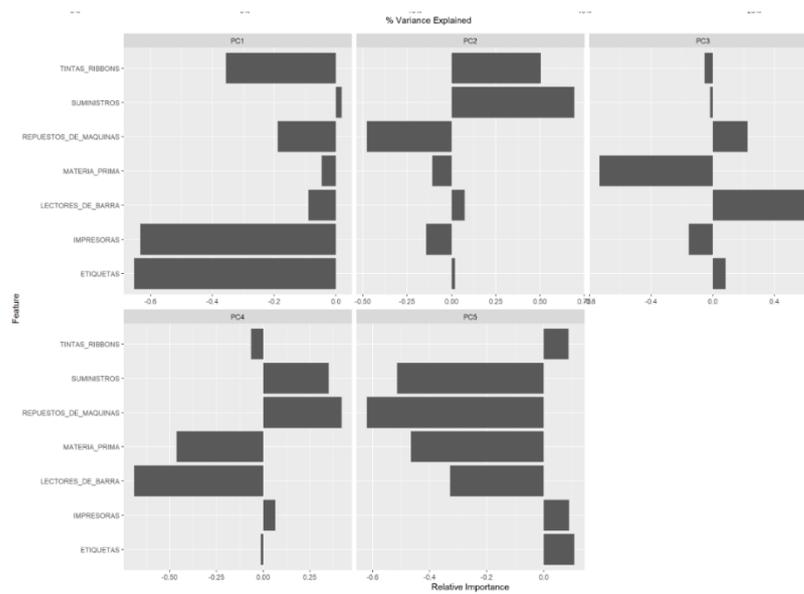


Nota:

Reporte obtenido de la modulación en RStudio

Figura 20

Variance Explained



Nota: Reporte obtenido de la modulación en RStudio

Con respecto a los PCA donde se busca a través de los componentes principales explica gran parte de la varianza de los datos en el cual se determinaron 5 componentes principales donde a través de la acumulativa solo se requirió el 79% para poder representar y explicar la data. Con el ACP se busca ver cuál de los

componentes tienen mayor afinidad con respecto a la lectura de datos en la clasificación de clientes, donde la varianza está explicada de forma positiva siendo los componentes 2 y 4 aquellos que tienen mejor resultado donde se puede observar la mejor reducción de dimensiones y se complementan entre sí.

Seguido se realiza una estimación de la cantidad de k centroides que se van a establecer para el modelo aplicando el siguiente código

```
NCluster = NbClust(DataESC, distance = "euclidean", min.nc = 2, max.nc  
= 15, method = "kmeans", index = "alllong")
```

Donde a través de este, según el dataset escalado; mediante la distancia euclidean se logra estimar las diferentes posibles cantidades óptimas de centroides por todos los métodos aceptados, estableciendo como mínimo 2 clúster y un máximo de 15. De esta forma se podrá estimar el número óptimo de clústeres para el modelo según el valor que obtenga la mayor frecuencia de cada método realizado por el algoritmo.

El resultado arrojado determinó lo siguiente:

Figura 21

Estimación de centroides

```
Among all indices:  
* 3 proposed 2 as the best number of clusters  
* 1 proposed 3 as the best number of clusters  
* 1 proposed 4 as the best number of clusters  
* 1 proposed 5 as the best number of clusters  
* 1 proposed 6 as the best number of clusters  
* 5 proposed 7 as the best number of clusters  
* 1 proposed 8 as the best number of clusters  
* 2 proposed 9 as the best number of clusters  
* 3 proposed 13 as the best number of clusters  
* 5 proposed 15 as the best number of clusters  
***** Conclusion *****  
* According to the majority rule, the best number of clusters is 7
```

Nota: Obtenido de la modulación del código en RStudio

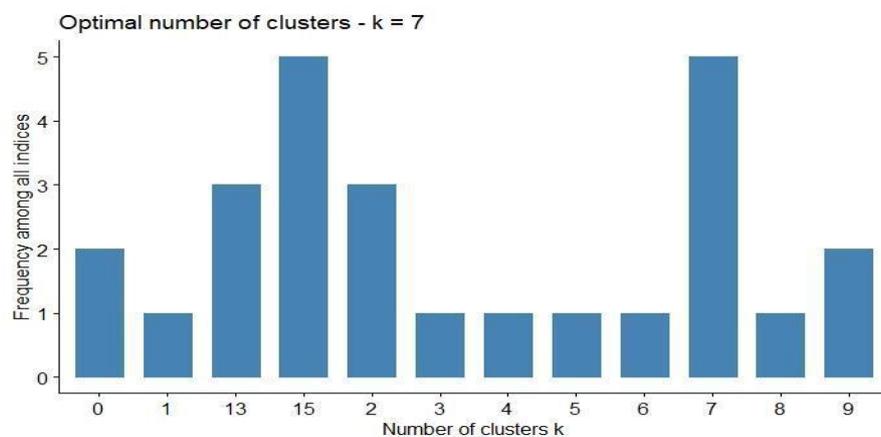
Para una mejor visualización de la cantidad óptima de clústeres a emplear se usa el siguiente código

```
fviz_nbclust(NCluster)
```

El mismo permitirá visualizar de mejor manera las diferentes cantidades con sus respectivas frecuencias en conjunto con el óptimo de K. Según el mayor número de frecuencias se va a establecer como número óptimo de clústeres 7 siento que es la cantidad con la cual la mayoría de los métodos coincidió.

Figura 22

Gráfico de clústeres óptimos



Nota: Obtenido de la modulación del código en RStudio

Se realiza el cálculo de clúster utilizando los datos de los cuales se les había realizado un escalamiento donde se consolidó la cantidad de centroides basándonos en el cálculo del número óptimo de K que es 7, a dicha partición de conglomerados se le asignó como referencia el nombre de “kn”.

```
kn = kmeans(ClientesESC, centers = 7, nstart = 25)
```

El mismo que al ser llamado arroja la información correspondiente a la creación de los k-means, donde se puede observar los diferentes clústeres del 1 al 7 con sus tamaños siendo de 3, 3, 5, 7, 3, 477, 3 respectivamente. A la vez se encuentran las

medias de cada clúster con respecto a cada una de las variables representadas por la categoría de productos.

Figura 23

Creación de K-means

K-means clustering with 7 clusters of sizes 3, 3, 5, 7, 3, 477, 3

Cluster means:

	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS
1	-0.29756357	-0.23537993	-0.08697882	-0.08199926	-0.14207964
2	-0.22190575	0.43275234	-0.08697882	12.40083007	-0.14207964
3	-0.07879695	0.03187298	-0.08697882	-0.08199926	8.87899250
4	6.02615552	4.68016461	0.80233523	-0.08199926	0.46201002
5	-0.28421219	-0.23537993	12.36341781	-0.08199926	-0.14207964
6	-0.09082550	-0.07429982	-0.08697882	-0.07438261	-0.09627689
7	1.31523405	0.87817385	-0.08697882	-0.08199926	-0.14207964

	SUMINISTROS	TINTAS_RIBBONS
1	11.44102815	-0.03187699
2	-0.13593432	0.05716304
3	-0.13402180	-0.14478733
4	-0.12445638	0.39221337
5	-0.14253942	-0.15501746

Nota: Obtenido de la modulación del código en RStudio

Se puede visualizar la eficiencia del modelo determinado por $(\text{between_SS} / \text{total_SS})$ el cual es una medida de calidad que mide la fracción de la variabilidad de los datos que explican los clústeres e indica que tanto están separados los grupos de manera inter-clúster en relación con el agrupamiento intra-clúster.

Debido a que se busca lograr conglomerados que tengan propiedades de cohesión interna con respecto a una separación externa entre los mismos lo ideal es un valor lo más cercano a 1. Ya que el modelo arroja un valor del 73.0% indica un buen ajuste que se traduce en una buena segmentación.

Figura 24

Ajuste del modelo

within cluster sum of squares by cluster:

```
[1] 3.920503e+01 3.582683e+01 3.139670e+01 3.282031e+02 4.278224e-03 4.410603e+02
```

```
[7] 6.816076e+01
```

```
(between_SS / total_SS = 73.0 %)
```

Nota: Obtenido de la modulación del código en RStudio

Debido a la gran cantidad de conglomerados que se creó se opta por realizar un diagrama de caja para poder visualizar las tendencias de cada uno de los clústeres con respecto a las variables para poder observar las tendencias que poseen con respecto a sus hábitos de compras y preferencias de categorías de productos, a su vez de poder interpretar como pueden asociarse para la creación de campañas en conjunto orientados a las necesidades de cada uno de los conglomerados

Se crea un objeto y se le asocia un nombre a cada uno de los gráficos de cajas correspondientes a cada categoría para posteriormente crear un panel de control que permita visualizar todos los gráficos en conjunto para una mejor interpretación a través del código

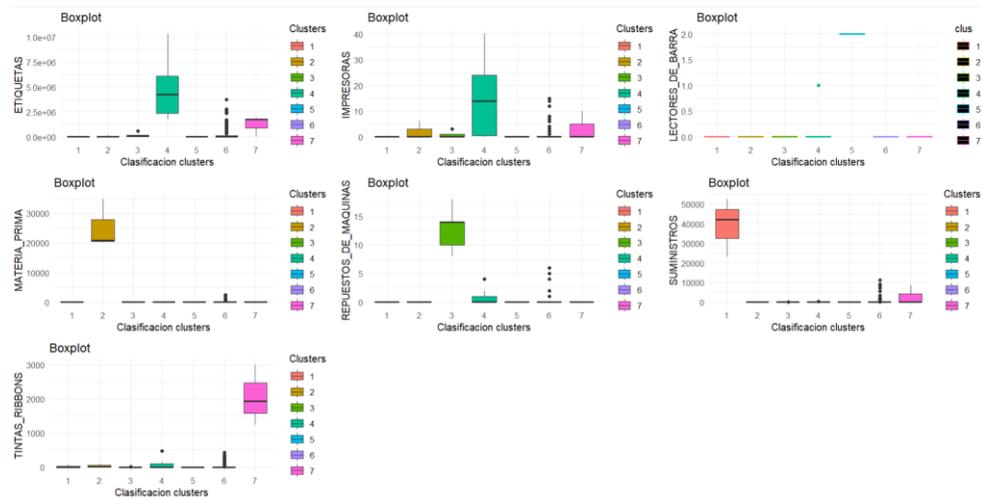
```
gridExtra::grid.arrange(fig1, fig2, fig3, fig4, fig5, fig6, fig7)
```

Se logra apreciar como en el segmento de etiquetas hay una tendencia creciente por cada uno de los clústeres, siendo el clúster 4 el más predominante, esto con relación a que dicha categoría es aquella con la cual la compañía inició sus actividades y por la que es reconocida por lo tanto es en la que la mayor cantidad de conglomerados interactuara. Con respecto a la categoría de impresoras notamos también una predominancia con respecto al clúster 4 a la par que se denota un gran potencial en los clústeres 2, 3 y 7 lo que indica que se debe potenciar el marketing y ofrecimiento de productos relacionados a dichas categorías a aquellos clientes que pertenecen a dichos clústeres que pueden aumentar su nivel de compra. En la categoría de lectores de barras

notamos una dominancia en el clúster 5 siendo los que mayormente comprar dichos productos. Con respecto a la categoría de materia prima domina el clúster 2 que en conjunto con las impresoras nos da un indicativo que las empresas pertenecientes a dicho conglomerado sean industrias publicitarias o se dediquen a la producción de área grafica. En repuestos de maquinas encontramos que el clúster 3 es el que tiene mayor tendencia y un potencial de aumento con respecto al clúster 4 que se relación con respecto a su preferencia por la categoría de impresoras el ofrecerles productos de refacciones para dichas impresoras que eventualmente necesitaran mantenimiento. En suministros domina el clúster 1 en conjunto con un crecimiento del clúster 7 segmento a cuál se le debe prestar atención. Por ultimo en Tintas Ribbons notamos que el clúster 7 tiene gran dominancia lo cual se relaciona con su creciente en la compra de etiquetas y impresoras las cuales también incluyen etiquetadoras con lo cual dichas empresas sean industrias de gran tamaño y una amplia cartera de productos ofrecidos al publico que requieren se etiquetados, hay un potencial en el clúster 4 que tiene una tendencia similar en las otras categorías con respecto al segmento 7 por lo que se les puede orientar a una campaña en conjunto que abarque las necesidades de los clientes pertenecientes a estas agrupaciones.

Figura 25

Panel de control de los diagramas de cajas por categoría de producto

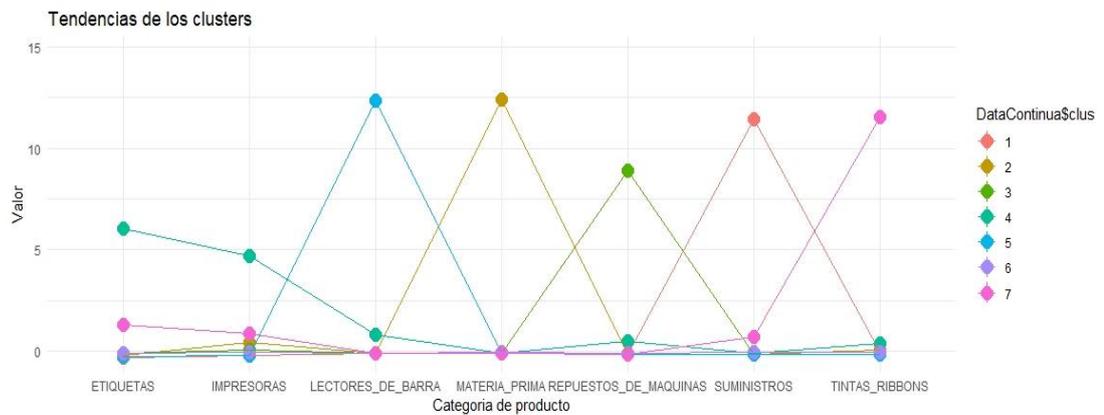


Nota: Obtenido de la modulación del código en RStudio

Después de una transformación en la data para poder visualizar gráficamente de una forma conjunta cada uno de los clústeres y sus preferencias, mediante una puntuación determinada por las categorías de productos con respecto a cada clúster, se puede apreciar en qué variables ciertos clústeres poseen similitud con respecto a su preferencia por la compra de dicha categoría de producto, así como la posible tendencia a crecer la demanda de otras categorías.

Figura 26

Tendencia de los clústeres



Nota: Obtenido de la modulación del código en RStudio

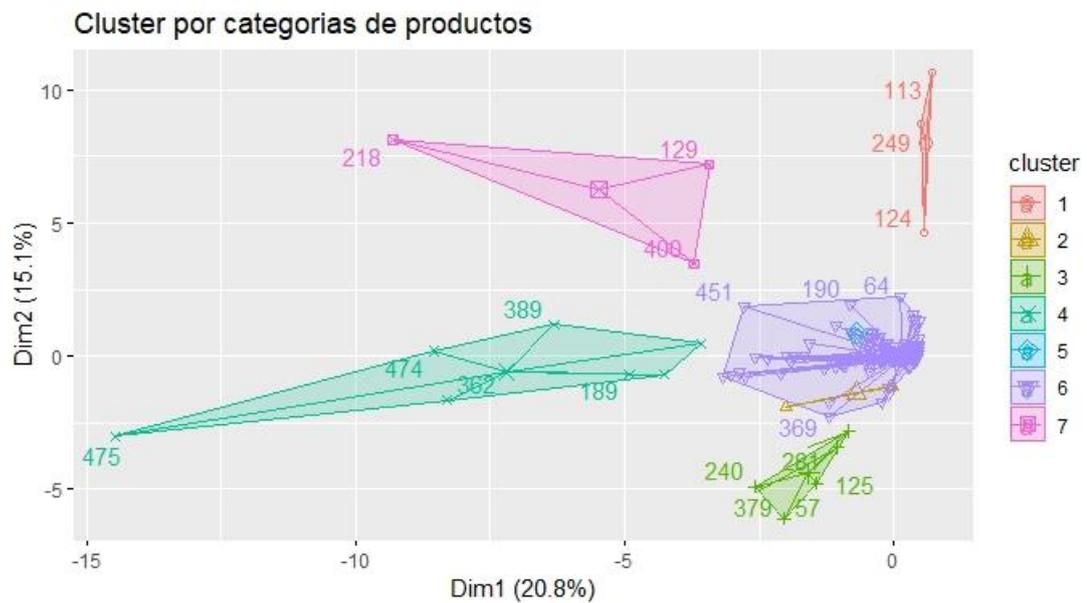
En el mismo se puede evidenciar las tendencias en conjunto de cada uno de los segmentos, se aprecia que únicamente el conglomerado 6 no tiene una predisposición hacia ninguna categoría esto se puede explicar debido a que los individuos pertenecientes a dicho segmento tengan preferencia que se ajusten a los demás clústeres el mismo que podrá ser apreciado en un gráfico donde se pueda determinar el comportamiento de los individuos de cada clúster. Se observa que los individuos pertenecientes al 4 conjunto con los que tiene mayor potencial en la mayoría de las categorías por lo que son clientes a los cuales la compañía debe buscar retener sin descuidar a los pertenecientes a las demás agrupaciones que tiene un gran potencial. Es en todas estas agrupaciones que tienen una tendencia al alza donde existe mayor potencial de que aumenten la cantidad de compra si se realiza un enfoque correcto para la proposición de dichos productos de las categorías ya determinadas.

Por último la mejor forma de hacer un correcto análisis de un algoritmo de clusterización es a través de la forma gráfica con lo cual se graficaron los clústeres para poder determinar cómo interactúan los conglomerados entre sí con respecto a los individuos que pertenecen a cada uno.

```
fviz_cluster(kn, data = ClientesESC, ellipse.type = "convex",  
             repel = TRUE, star.plot = TRUE,  
             main = "Cluster por categorías de productos")
```

Figura 27

Clúster por categorías de productos



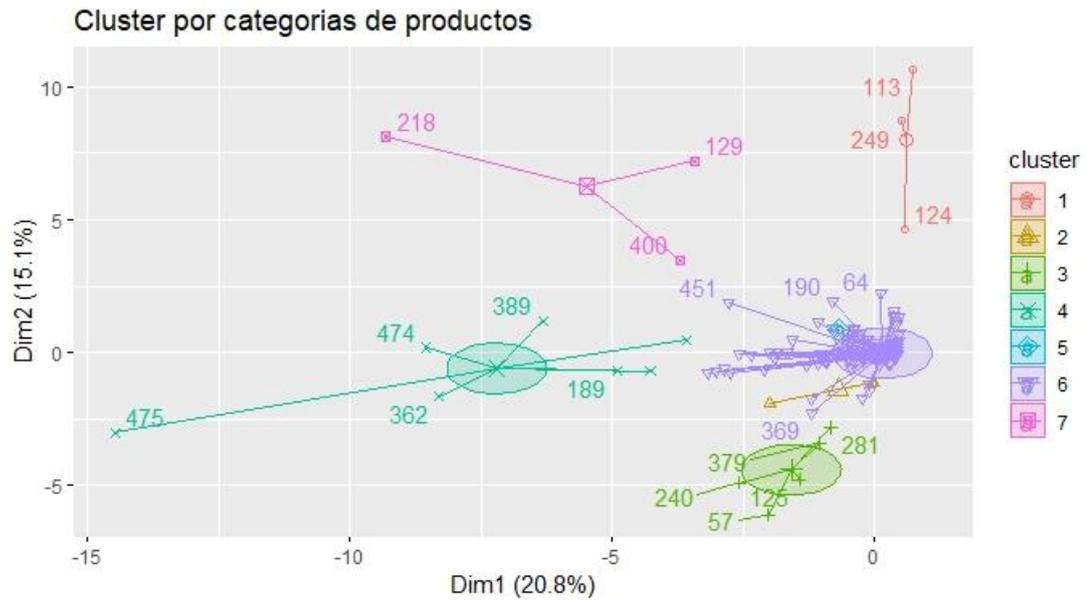
Nota: Obtenido de la modulación del código en RStudio

En el mismo se puede observar como el clúster 6 es el de mayor tamaño y contiene al clúster 5 y 2 lo que permite inferir que, aunque tienen ciertas preferencias similares cada clúster tiene determinada tendencia que lo separa del 6 a la par se pueden observar los clústeres que tienen más distancias como puede ser el 4 en donde ciertas observaciones se están alejando y posiblemente en una k mayor de conglomerados se cree un nuevo clúster o agrupación.

En el gráfico como tal no se consigue distinguir de forma óptima alguno de los clústeres que son opacados por el clúster 6 el cual posee la mayor cantidad de clientes por lo tanto se opta por un gráfico elíptico estrella del tipo euclidean. Donde se consigue una mejor visualización de todos los clústeres o conglomerados, así como algunos individuos y su tendencia en el plano.

Figura 28

Gráfico elíptico de tipo euclidean



Nota: Obtenido de la modulación del código en RStudio

En el mismo se puede evidenciar como algunos clientes tienen tendencia a cambiar de conglomerado esto debido a que, aunque pertenecen al clúster 6 sus hábitos de compra se orientan hacia las predisposiciones que tienen otros segmentos con respecto a determinada categoría por lo que podrían ser asociados en una campaña en conjunto con respecto a dichas preferencias a los clientes pertenecientes a dichos clústeres.

Conclusiones

La implementación del machine learning ha demostrado ser fundamental en los distintos ámbitos y segmentos en los que es aplicado yendo a través de las distintas ramas científicas, económicas y sociales consolidándose como un requerimiento clave de innovación y progreso. Dentro del ámbito empresarial es donde se ha visto una gran diversificación en las últimas décadas con respecto a los distintos métodos de implementación de este en aspectos como la clasificación de clientes, para lo cual la creación de conglomerados por K-means clustering ha demostrado ser sumamente eficaz.

El conjunto de técnicas empleadas para la segmentación por el modelo de machine learning K-means permiten realizar una partición fiel a los requerimientos de cada uno de las observaciones pertenecientes a cada centroide validando dicha agrupación por medio de distintos parámetros que consolidan a cada uno de los conglomerados buscando que exista una gran semejanza intra-cluster y garantizando la mayor disparidad inter-cluster con lo que cada una de las agrupaciones tendrán una tendencia en específico con respecto a sus necesidades según la base de datos proporcionada y segmentada para las distintas categorías de productos ofrecidas a las distintas empresas conforme sus histórico de compras.

Con el empleo de este modelo se permite vislumbrar diversos datos con respecto a datos cuantitativos y que al tener un fundamento gráfico para su interpretación también garantiza la realización de análisis más rápidos y menos subjetivos que pueden ser derivados distintos departamentos de las empresas y de esta manera ser empleada dicha información en distintas campañas por lo que su implementación y desarrollo en conjunto con otros modelos que permitan un mayor análisis es necesaria en las diversas

empresas donde podrán visualizar las preferencias de compras y productos con respecto a cada cliente para mejores campañas y estrategias.

Se establece que la ventaja del modelo de clasificación de cliente por clusterización es que su interactividad es esencialmente gráfica, entonces a través del gráfico de segmentos de clúster en conjunto con el de tendencias por líneas y puntos se podrá explicar al directorio que debe crearse una campaña en conjunto focalizándose en las categorías de repuestos de máquinas, suministros y tintas ribbons donde se denota que ciertos clústeres tienen tendencia al alza y deben ser tratados en conjunto. A su vez también se puede percibir que, aunque un clúster tiene tendencia hacia una categoría también se ve una predisposición hacia el crecimiento con respecto a otras categorías que posiblemente complementen a la categoría principal de dicho clúster.

Recomendaciones

Para la realización de cualquier investigación posterior que tenga que realizar un estudio por medio de la aplicación del machine learning tienen que considerar una indagación más adecuada de acuerdo con la problemática que se va a analizar en cuestión. Del mismo modo, el obtener información de los datos que se van a emplear de forma rápida y actualizada en factor a que se requieren data de calidad para un estudio más eficiente.

Con el propósito de mejorar el modelo de clusterización es pertinente realizar un análisis y una escalabilidad en las bases de datos de la incidencia de sus variables. Si bien es cierto, la investigación tiene un enfoque hacia la clasificación de clientes, sin embargo, la estandarización es importante dentro de la aplicación de algoritmos de clasificación e influencia en el proceso de transformar los valores de las características de las variables consideradas aptas para el procesamiento en el modelo. De esta forma, el análisis se complementa con conceptos en el área comercial y de ventas.

En el procesamiento de información, sólo se utilizaron datos de una empresa de etiquetas en la ciudad de Guayaquil. A pesar de estar bien posicionada en el mercado, existe una gran brecha de clientes potenciales en la competencia local, por lo tanto, el análisis sirve para tener una noción acerca del funcionamiento del modelo junto con las variables más relevantes. Para que refleje de una forma acertada la realidad de la empresa en el mercado, en base a su cartera de clientes.

Una vez analizados los resultados obtenidos del modelo, se recomienda el establecimiento de estrategias viables como es la diversificación de las campañas comerciales que maneja la compañía, en base a los clústeres obtenidos (grupo de clientes) teniendo en cuenta la incidencia de compra de las categorías de productos por parte de la cartera de clientes, lo cual permitirá optimizar el uso de recursos por

parte de la empresa. De igual manera, otra estrategia que se puede implementar es campañas de marketing especializada en las categorías de productos, considerando 2 perspectivas: (a) Las categorías de productos con mayor volumen de compra (b) Las categorías de productos con menor incidencia de compra. Con lo cual la empresa puede potenciar las ventas de los productos más comprados por su cartera de clientes y a su vez potenciar las categorías con poco alcance.

Tener en cuenta si las variables a utilizar para la creación del modelo son relevantes para el enfoque de la investigación. Con el fin de mejorar la toma de decisiones, se deben incluir periódicamente más data histórica, debido a que los datos van a indicar la evolución en el comportamiento de la cartera de clientes de las empresas y con esta información de segmentación, la alta gerencia puede beneficiarse a través de la creación de estrategias personalizadas para cada segmento de cliente.

Considerar el realizar los procesos de aplicación de manera adecuada y rápida en el programa R Studio debido a que hay errores que pueden suceder en el modelo y podrían alterar toda la investigación siendo causante de no conseguir los resultados óptimos a la clasificación que se plantea, ya que gracias a esto se garantiza una eficiencia en cuanto al manejo de la data y el rendimiento de los resultados.

Referencias

- Aguirre. (n.d.). *Machine Learning*. Pueblo Continente. Retrieved November 20, 2022, from <http://200.62.226.189/PuebloContinente/article/view/1363/1171>
- Alameda, T. (2022, 17 octubre). «Machine learning»: ¿qué es y cómo funciona? BBVA NOTICIAS. <https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/>
- Alberto, R. (2021, 16 diciembre). Clustering con Gaussian Mixture Model en SKlearn y sus parámetros aplicados al Marketing. Medium. <https://rubialesalberto.medium.com/clustering-con-gaussian-mixture-model-en-sklearn-y-sus-par%C3%A1metros-aplicado-al-marketing-74b9d8454b86>
- Alfonso, M., Cazorla, M., Colomina, O., Escolano, F. y Lozano, M. (2003). *Inteligencia artificial: Modelos, técnicas y áreas de aplicación [Digital]*. En *Modelos, técnicas y áreas de aplicación* (1.a ed.). Paraninfo. https://books.google.com.ec/books?id=_spC6S7UfZgC&dq=inteligencia+artificial+pdf&lr=&hl=es&source=gbs_navlinks_s
- Álvarez, M., Quirós, L., & Cortés, M. (2020). *Inteligencia artificial y aprendizaje automatico*. Revista Medica Sinergia. Retrieved November 20, 2022, from <https://www.revistamedicasinergia.com/index.php/rms/article/view/557/923>

- Amat Rodrigo, J. (2017, 21 septiembre). RPubs - Clustering y heatmaps: aprendizaje no supervisado con R. RPubs. Recuperado 1 de febrero de 2023, de https://rpubs.com/Joaquin_AR/310338
- American Marketing Association. (s. f.). AMA. AMA Online Portal. Recuperado 22 de noviembre de 2022, de <https://myama.my.site.com/s/article/Marketing-Dictionary>
- Asencio Diaz, L. d. l. A., Chiang Cornejo, R., Crisóstomo Fernández, F., Hernández Quiroz, G., & Lajo Aurazo, A. (2021, 12). *Técnicas de Machine Learning para la clasificación automática de clientes en una empresa de seguros*. Repositorio ESAN. Retrieved October 26, 2022, from <https://repositorio.esan.edu.pe/handle/20.500.12640/2933>
- Bautista, A. (2019). *Análisis, implementación y evaluación de modelos de aprendizaje automático relacional*. Repositorio Digital UCE. Retrieved November 20, 2022, from <http://www.dspace.uce.edu.ec/bitstream/25000/19631/1/T-UCE-0011-ICF-173.pdf>
- Boden, M. A. (2018). *Inteligencia artificial (Edición estándar)* [Digital]. Turner. https://books.google.com.ec/books?id=LCnYDwAAQBAJ&dq=inteligencia+artificial+pdf&lr=&hl=es&source=gbs_navlinks_s
- Camacho Rosero, M. P., y Reyna Muentes, C. A. (2016, February 19). *TheOfficialHalie - Represent J.A ft. Chris Money*. YouTube. Retrieved

October 17, 2022, from

<http://repositorio.ucsg.edu.ec/handle/3317/18340>

Cardenas, J. (2018). EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD DENNIYE HINESTROZA RAMÍREZ JUAN MANUEL CÁRDENAS. Retrieved November 20, 2022, from

<https://repository.unilibre.edu.co/bitstream/handle/10901/17289/EL%20MACHINE%20LEARNING.pdf?sequence=1&isAllowed=y>

Cartagena, P. (2017, 09 21). *Predicción de la probabilidad de éxito en la adquisición de clientes*. RUA. Retrieved October 23, 2022, from <https://rua.ua.es/dspace/handle/10045/69432>

Casas, J. (2018). *Introducción al Big Data*. Universitat Oberta de Catalunya. Retrieved November 21, 2022, from

http://openaccess.uoc.edu/webapps/o2/bitstream/10609/141647/4/Análisis%20de%20datos%20en%20entornos%20big%20data_Introduccion%20al%20big%20data.pdf

Castro Triviño, A. (2019). *Perspectiva para la mejora del servicio al cliente en las empresas públicas de la ciudad de Guayaquil [Tesis]*. Universidad Politécnica Salesiana.

ceaces. (s. f.). *Análisis Clúster*. CEACES. Recuperado 1 de febrero de 2023, de <https://www.uv.es/ceaces/multivari/cluster/CLUSTER2.htm>

Coder, R. (2020, 7 julio). *Fijar semilla en R*. R CODER. <https://r-coder.com/fijar-semilla-r/>

- Coelho, F. (2022, 4 octubre). Qué es Machine Learning: casos de éxito en empresas. DIGITAL55. <https://digital55.com/que-es-machine-learning-casos-exito-empresas/>
- Díaz, J. (2010). Introducción al Business Intelligence [Digital]. En *Introducción al Business Intelligence* (1., Vol. 1). Editorial Uoc. https://books.google.es/books?id=iU3RAXYQXMkC&hl=es&source=gbs_navlinks_s
- Díez, V. I. I. (2004). Cómo conquistar el mercado con una estrategia CRM [Digital]. FC EDITORIAL. https://books.google.com.ec/books/about/C%C3%B3mo_conquistar_el_mercado_con_una_estr.html?id=XLEVkuXa99UC&source=kp_book_description&redir_esc=y
- Domenech, J. (2020). *Diagrama de correlación - dispersión*. DIAGRAMA DE CORRELACIÓN-DISPERSIÓN. Retrieved November 23, 2022, from http://jomaneliga.es/PDF/Administrativo/Calidad/Diagrama_Correlacion_Dispersion.pdf
- EMIS. (2022, 14 enero). Labelprintsa S.A. Perfil de Compañía - Ecuador | Finanzas y ejecutivos clave | EMIS. https://www.emis.com/php/company-profile/EC/Labelprintsa_SA_es_5463924.html

- Estadística, P. Y. (2022, 8 marzo). ▷ Valores atípicos (outliers): qué son, ejemplos, calculadora, . . . Probabilidad y Estadística.
<https://www.probabilidadyestadistica.net/valores-atipicos-outliers/>
- Fernández, A. (2019, March 29). *Inteligencia artificial en los servicios financieros. Artículos Analíticos. Boletín Económico 2/2019*. Banco de España. Retrieved October 26, 2022, from
<https://core.ac.uk/download/pdf/322617455.pdf>
- Garcés, H. (2020, February 27). *Aplicación de Machine Learning en la demanda de productos permisibles*. Repositorio Digital UCSG. Retrieved November 22, 2022, from
<http://repositorio.ucsg.edu.ec/handle/3317/14264>
- García, G. (2022, 8 abril). Estas son las tendencias de la industria de etiquetas y envases para 2022. The Food Tech.
<https://thefoodtech.com/maquinaria-para-ensado-y-procesamiento/estas-son-las-tendencias-de-la-industria-de-etiquetas-y-envases-para-2022/>
- Gartner. (2011, February 11). *La cruda realidad del Business Intelligence* (Gartner, 2011). Business Intelligence fácil. Retrieved November 20, 2022, from <https://www.businessintelligence.info/mercado/gartner-business-intelligence-2011.html>
- Gonzales, L. (2018, 06 22). *Diferencia entre algoritmos de clasificación y regresión - Aprende IA*. Aprende IA. Retrieved November 20, 2022,

from <https://aprendeia.com/diferencia-entre-algoritmos-de-clasificacion-y-regresion/>

Gutiérrez Cordero, M. L., Segovia Vargas, M. J., y Ramos Escamilla, M.

(2017, 12). *Análisis del Riesgo de Caída de Cartera en Seguros*.

Science Direct. Retrieved October 23, 2022, from

<https://reader.elsevier.com/reader/sd/pii/S0185084917300592?token=375DCA004CFFF21E94D2B753562B1470A8D292F64B1EF94382E6CF7C8C64AAB1657B694B0BE40E8EA04BD37E8EC7BB8A&originRegion=us-east-1&originCreation=20221023230509>

Hurwitz, J. & kirsch, D. (2018). *Machine Learning: For Dummies* (Vol. 1)

[Digital]. John Wiley & Sons, Inc.

<https://www.ibm.com/downloads/cas/GB8ZMQZ3>

Hurwitz, J., & Kirsch, D. (2019). *Machine Learning for Dummies* (Limited

Edition ed.). IBM. <https://www.ibm.com/downloads/cas/GB8ZMQZ3>

IBM Cloud Education. (2021, 14 septiembre). *Redes neuronales*. IBM.

<https://www.ibm.com/cl-es/cloud/learn/neural-networks>

IBM. (s. f.). *What is business intelligence?* Recuperado 22 de noviembre de

2022, de <https://www.ibm.com/analytics/business-intelligence>

International Business Machines Corporation (IBM). (2020, April 28). *What is*

ETL (Extract, Transform, Load)? IBM. Retrieved November 20,

2022, from <https://www.ibm.com/cloud/learn/etl>

- Issuu. (2022, January 31). *Revista Industrias enero 2022 by Cámara de Industrias de Guayaquil*. Issuu. Retrieved October 19, 2022, from https://issuu.com/industrias/docs/revista_industrias_enero_2022
- Jiménez, C. (2014, December 1). *Big data. Un nuevo paradigma de análisis de datos*. Repositorio Comillas. Retrieved November 21, 2022, from <https://repositorio.comillas.edu/xmlui/handle/11531/4873>
- Labelprintsa S.A. (2021). *Nuestra Historia – Sitio LabelPrint ECUADOR*. Sitio LabelPrint ECUADOR. Retrieved October 19, 2022, from <https://labelprintec.com/nuestrahistoria/>
- Lastra, E. F. (2021, 16 julio). *Cómo clasificar y segmentar a tus clientes*. artyco | the data driven company. <https://artyco.com/como-clasificar-segmentar-clientes/>
- Lizardo, M. A. (2021, 20 octubre). *5 aplicaciones de Machine Learning en la analítica de clientes | Inflow*. Inflow | Laboratorio de Crecimiento con Datos. <https://inflow.cl/2021/10/20/5-aplicaciones-de-machine-learning-en-la-analitica-de-clientes/>
- Lluís Cano, J. (2007). *Business intelligence: competir con información [Digital]*. En *Business intelligence: competir con información* (1.a ed., Vol. 1). Banesto, Fundación Cultur. <https://fhcevirtual.umsa.bo/btecavirtual/?q=Business%20intelligence%3A%20competir%20con%20informaci%C3%B3n>
- Management Solutions. (2018). *Machine Learning, una pieza clave en la transformación de los modelos de negocio*. © Management Solutions,

1(1).

<https://www.managementsolutions.com/sites/default/files/publicaciones/esp/machine-learning.pdf>

Management Solutions. (2018). Machine Learning: Una pieza clave en la transformación de los modelos de negocio. Management Solutions, 1, 6.

<https://www.managementsolutions.com/sites/default/files/publicaciones/esp/machine-learning.pdf>

MD. (2019, 23 septiembre). La relación entre el Big Data y el crecimiento empresarial. Marketing Directo.

<https://www.marketingdirecto.com/marketing-general/marketing/la-relacion-entre-el-big-data-y-el-crecimiento-empresarial>

Micó, J. L. (2019, 29 mayo). China lidera la carrera para recopilar datos para la inteligencia artificial. La Vanguardia.

<https://www.lavanguardia.com/tecnologia/20190528/462546452120/inteligencia-artificial-chinac-carrera-datos.html>

Ocaña Peinado, F. M. (s. f.). Tratamiento estadístico de outliers y datos faltantes. Técnicas estadísticas en Nutrición y Salud, 1.

<https://www.ugr.es/~fmocan/MATERIALES%20DOCTORADO/Tratamiento%20de%20outliers%20y%20missing.pdf>

Olamendi, G. (s. f.). Diccionario de marketing [Digital]. En Diccionario de marketing (1.a ed., Vol. 1).

<https://esgerencia.com/pdf/blogmkt/diccionario.pdf>

- Ollé, J. (2021, 28 diciembre). Qué es y para qué sirve el clustering – un ejemplo de aplicación práctica. Conceptos claros.
<https://conceptosclaros.com/que-es-clustering/>
- Pérez, B. (2018). Inteligencia artificial. INCyTU, 012, 1.
https://www.foroconsultivo.org.mx/INCyTU/documentos/Completa/INCYTU_18-012.pdf
- Rivera, J., y Villegas, C. (2022). *Business Ethics and Artificial Intelligence*. DSpace Principal. Retrieved November 21, 2022, from
<https://repositorio.comillas.edu/xmlui/handle/11531/57390>
- Rodrigo, J. A. (s. f.). Clustering y heatmaps: aprendizaje no supervisado.
https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps
- Rodríguez, P. (2020, 31 agosto).» La falta de profesionales estrangula la expansión de la inteligencia artificial.
<https://telos.fundaciontelefonica.com/la-cofa/la-falta-de-profesionales-estrangula-la-expansion-de-la-inteligencia-artificial/>
- Rojas, E. (2020). *Machine Learning: análisis de lenguajes de programación*. Proquest. Retrieved November 20, 2022, from
<https://www.proquest.com/openview/c7e24c997199215aa26a39107dd2fe98/1?pq-origsite=gscholar&cbl=1006393>
- Rouhiainen, L. (2018). Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro [Digital]. Alianza Editorial.

https://www.planetadelibros.com/libros_contenido_extra/40/39307_In teligencia_artificial.pdf

Saltos Intriago, D. A., y Villacis Breilh, O. J. (2022, February 21).

Implementación de Machine Learning en el área de ventas de la empresa Zapec S.A. Repositorio Digital UCSG. Retrieved October 16, 2022, from <http://repositorio.ucsg.edu.ec/handle/3317/18337>

Saltos, D., y Villacis, O. (2022, February 21). *Implementación de Machine Learning en el área de ventas de la empresa Zapec S.A.* Repositorio Digital UCSG. Retrieved November 20, 2022, from <http://repositorio.ucsg.edu.ec/handle/3317/18337>

Sánchez Muños, J. M. (2015). Análisis de Calidad Cartográfica mediante el estudio de la Matriz de Confusión. *Pensamiento Matemático*, 6(2), 9-26.

Silva Solano, L. (2017). Business Intelligence: un balance para su implementación. *INNOVAG*, 3, 27-36.

<https://revistas.pucp.edu.pe/index.php/innovag/article/view/19742>

Stubblefield, & Luger. (1990, November 1). *[PDF] Artificial Intelligence and the Design of Expert Systems*. Semantic Scholar. Retrieved November 20, 2022, from <https://www.semanticscholar.org/paper/Artificial-Intelligence-and-the-Design-of-Expert-Luger-Stubblefield/826af5087384f59ae540068eaf6ed04a8a38a7ee>

Tejada Romero, C. (2020, February 27). *Análisis del uso de la Inteligencia Artificial en la atención presencial de los clientes de la Empresa de*

Telecomunicaciones Región Sur en el 2018. Repositorio Continental.

Retrieved October 23, 2022, from

<https://repositorio.continental.edu.pe/handle/20.500.12394/7203>

Tejada, P. S. de. (2021, 19 mayo). ¿Cómo identificar y filtrar outliers (valores atípicos) en Tableau? The Information Lab.

<https://theinformationlab.es/como-identificar-y-filtrar-outliers-valores-atipicos-en-tableau/>

Theng, M. (2019, January 17). *Machine Learning Algorithms for Predictive*

Analytics. Research Gate. Retrieved November 21, 2022, from

https://www.researchgate.net/publication/342976767_Machine_Learning_Algorithms_for_Predictive_Analytics_A_Review_and_New_Perspectives

Thompson, I. (2005). La Segmentación del Mercado. *La Segmentación del Mercado*, 1(1), 1.

https://www.accioneduca.org/admin/archivos/clases/material/segmentacion-y-mercado_1563828020.pdf

Universidad Autónoma de Madrid. (s. f.). Análisis conglomerados Cluster.

Portal Estadística Aplicada, 1(1). https://www.estadistica.net/Master-Econometria/Analisis_Cluster.pdf

Universidad Latina de Costa Rica. (2020, 22 julio). Importancia de la estrategia empresarial y su aplicación. Ulatina.

<https://www.ulatina.ac.cr/articulos/importancia-de-la-estrategia-empresarial-y-su-aplicacion>

Valverde, S. (2019, 07). *Aplicaciones de la inteligencia artificial en la empresa*. UCrea. Retrieved October 23, 2022, from <https://repositorio.unican.es/xmlui/handle/10902/17521>

Zeilberger, D. (2020, December 20). *Algoritmos de Clustering - Fernando Sancho Caparrini*. Dpto. Ciencias de la Computación e Inteligencia Artificial. Retrieved November 22, 2022, from <https://www.cs.us.es/~fsancho/?e=230>

Anexos

Anexo 1 Script Completo Rstudio

```
1
2 setwd("C:/Users/andyv/OneDrive/Esritorio/TESIS/Script")
3
4 #carga de datos####
5 Clientes=read.csv2("../Data/Clientes_por_categoria.csv")
6 View(Clientes)
7 attach(Clientes)
8
9 #Paquetes####
10 install.packages(c("fpc","tidyverse","cluster","factoextra",
11                   "NbClust","tidyr","writex1","DataExplorer"))
12
13 library(tidyverse)
14 library(fpc)
15 library(cluster)
16 library(factoextra)
17 library(NbClust)
18 library(tidyr)
19 library(ggplot2)
20 library(writex1)
21 library(DataExplorer)
22
23
24 #Optimizacion####
25 Clientes$CLIENTES=NULL
26
27 #Análisis Descriptivo####
28 DataExplorer::create_report(Clientes)
29
30 #Semilla####
31 set.seed(123)
32
33 #Escalamiento de los datos####
34 ClientesESC=data.frame(scale(Clientes))
--
```

```

35
36 head(ClientesESC)
37
38 #Estimacion del numero de clusters####
39 NCluster=NbClust(ClientesESC,distance = "euclidean",
40                 min.nc=2,max.nc=15,method="kmeans",index = "all")
41
42 fviz_nbclust(NCluster)
43
44 #Calculó de los clusters####
45 kn=kmeans(ClientesESC, centers = 7, nstart = 25)
46 kn
47 $tr(kn)
48
49 #Establecer la clasificacion en la base principal####
50 ClientesTratados=read.csv2("../Data/Clientes_por_categoria.csv")
51 ClientesTratados$clus=as.factor(kn$cluster)
52 ClientesTratados
53 view(ClientesTratados)
54
55 ClientesESC2=scale(Clientes)
56 ClientesESC2=as.data.frame(ClientesESC2)
57 ClientesESC2$clus=as.factor(kn$cluster)
58 ClientesESC2
59 view(ClientesESC2)
60
61 #Descriptivos de la clasificacion####
62 fig1=ggplot(ClientesTratados) +
63   aes(x = clus, y = ETIQUETAS, fill = clus) +
64   geom_boxplot() +
65   scale_fill_hue(direction = 1) +
66   labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters") +
67   theme_minimal()
68

```

```

69 fig2=ggplot(ClientesTratados) +
70   aes(x = clus, y = IMPRESORAS, fill = clus) +
71   geom_boxplot() +
72   scale_fill_hue(direction = 1) +
73   labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters") +
74   theme_minimal()
75
76 fig3=ggplot(ClientesTratados)+
77   aes(x = clus, y = LECTORES_DE_BARRA, colour = clus)+
78   geom_boxplot(fill = "black")+
79   scale_color_hue(direction = 1)+
80   labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters")+
81   theme_minimal()
82
83 fig4=ggplot(ClientesTratados) +
84   aes(x = clus, y = MATERIA_PRIMA, fill = clus) +
85   geom_boxplot() +
86   scale_fill_hue(direction = 1) +
87   labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters") +
88   theme_minimal()
89
90 fig5=ggplot(ClientesTratados) +
91   aes(x = clus, y = REPUESTOS_DE_MAQUINAS, fill = clus) +
92   geom_boxplot() +
93   scale_fill_hue(direction = 1) +
94   labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters") +
95   theme_minimal()
96
97 fig6=ggplot(ClientesTratados) +
98   aes(x = clus, y = SUMINISTROS, fill = clus) +
99   geom_boxplot() +
100  scale_fill_hue(direction = 1) +
101  labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters") +
102  theme_minimal()

```

```

103
104 fig7=ggplot(ClientesTratados) +
105   aes(x = clus, y = TINTAS_RIBBONS, fill = clus) +
106   geom_boxplot() +
107   scale_fill_hue(direction = 1) +
108   labs(x = "Clasificacion clusters", title = "Boxplot", fill = "clusters") +
109   theme_minimal()
110
111 gridExtra::grid.arrange(fig1,fig2,fig3,fig4,fig5,fig6,fig7)
112
113 #Grafico de lineas y punto por categoria de productos####
114 DataContinua=gather(ClientesESC2, Categoria, Valor,
115                    ETIQUETAS:TINTAS_RIBBONS,
116                    factor_key = TRUE)
117 DataContinua
118
119 ggplot(DataContinua, aes(as.factor(x=DataContinua$Categoria),
120                          y=DataContinua$Valor, group= DataContinua$clus,
121                          colour=DataContinua$clus))+
122   stat_summary(fun=mean,geom="pointrange",size=1)+
123   labs(x = "Categoria de producto",y="Valor",title = "Tendencias de los clusters
124   stat_summary(geom = "line")+theme_minimal()
125
126 #Grafico de los clusteres####
127 fviz_cluster(kn, data = ClientesESC,
128              ellipse.type = "convex",repel = TRUE,
129              star.plot = TRUE,main = "Cluster por categorias de productos")
130
131 fviz_cluster(kn, data = ClientesESC,
132              ellipse.type = "euclid", repel = TRUE,
133              star.plot = TRUE,main = "Cluster por categorias de productos")
134
135 #Creacion de documento Excel con la clasificacion####
136 write_xlsx(ClientesTratados, "ClientesSegm.xlsx")

```

Anexo 2 base de datos con la clasificación según el respectivo conglomerado

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
1	DIGITALCONNECTION_SA	5000	0	0	0.00	0	0	2 6
2	ABAD_CALVA_LUIS_FRANCISCO	56705	0	0	0.00	0	1	0 6
3	ADITIVOS_Y_ALIMENTOS_S_A_ADILISA	212894	0	0	0.00	0	0	0 6
4	AEREOSERV_SA	60000	0	0	0.00	0	0	0 6
5	AGILIDER_SA	120000	0	0	0.00	0	0	0 6
6	AGRAIN_D_SA	153638	0	0	0.00	0	8	0 6
7	AGRICOLA_AGROSERVICIO_CAMPOVERDE_AGRICAMPOVER...	17000	0	0	0.00	0	0	28 6
8	AGRICOLA_ARVECOR_SA	205456	0	0	0.00	0	8	0 6
9	AGRICOLA_BANANERA_ECUATORIANA_-AGBE_SA	30030	0	0	0.00	0	0	0 6
10	AGRICOLA_BIOAGRO_SA	292216	1	0	0.00	0	13	0 6
11	AGRICOLA_CARFEDI_SA	213084	0	0	0.00	0	0	0 6
12	AGRICOLA_DESESU_SA	372672	0	0	0.00	0	10	0 6
13	AGRICOLA_FREDNERS_SA	240710	0	0	0.00	0	0	0 6
14	AGRICOLA_JADKEN_SA	0	0	0	0.00	0	7000	0 6
15	AGRICOLA_KRASNAYA_SA	0	0	0	0.00	0	9	0 6
16	AGRICOLA_LONGDAY_SA	314390	0	0	0.00	0	11	0 6
17	AGRICOLA_MARIA_ISABEL_MARISAGRIC_SA	0	1	0	0.00	0	0	0 6
18	AGRICOLA_PESTIG_SA	440290	0	0	0.00	0	10	0 6
19	AGRICOLA_SOLEDAD_COLSDAD_SA	0	1	0	0.00	0	0	0 6
20	AGRICOMINSA_AGRICOLA_COMERCIAL_INDUSTRIAL_SA_A...	326800	0	0	0.00	0	0	0 6
21	AGROBANSUR_S_A	0	1	0	0.00	0	0	0 6
22	AGROBERRUZ_SA	135395	0	0	0.00	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
22 AGROBERRUZ_SA	135395	0	0	0.00	0	0	0	6
23 AGROBONUS_SA	0	0	0	0.00	0	230	0	6
24 AGROCARIBE_SA	177200	0	0	0.00	0	0	0	6
25 AGROFYRA_SAS	0	0	0	0.00	0	13	0	6
26 AGRONEGOCIOS_LOS_ANGELES_SAC	2638746	0	0	0.00	0	0	0	6
27 AGROORGANICA_SA	0	0	0	0.00	0	10	0	6
28 AGROVEGETALES_SA	18400	0	0	0.00	0	0	0	6
29 AGROVICTORIA_SA	81000	6	0	0.00	0	74	0	6
30 AGZULASA_CIA_LTDA	181060	6	0	0.00	0	10	285	6
31 ALESSA_SA	364100	0	0	0.00	0	0	255	6
32 ALFAFRUITS_SA	233880	0	0	0.00	0	0	0	6
33 ALMACEN_PROMOCIONES_SA_ALMAPROM	94400	0	0	0.00	0	0	0	6
34 ALMACENES_BUENHOGAR_EW_CIALTDA	1000000	1	0	0.00	0	0	30	6
35 ALMACENES_FIGUEROA_SA_FIGUESA	0	0	0	0.00	0	0	12	6
36 ALPHACELL_SA	0	0	0	0.00	0	88	0	6
37 ALVAREZ_INTRIAGO_CHRISTOPHER_OSVALDO	0	0	0	0.00	0	0	16	6
38 AMERICAN_CALL_CENTER_SA_AMERICALL	5000	0	0	0.00	0	0	2	6
39 AMPYCARL_SA	0	0	0	0.00	0	3500	0	6
40 ANDEAN_NATURAL_PRODUCTS_EXPORT_IMPORT_SAC	1656690	0	0	0.00	0	0	0	6
41 APRACOM_SA	1000	0	0	0.00	0	0	0	6
42 APROFE	0	0	0	0.00	0	1	0	6
43 ARBOLEDA_LLIGUIN_EDISON_GONZALO	6000	0	0	0.00	0	0	0	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
44 ARCINIEGA_ACOSTA_JOHANNA_PAULETTE	10000	0	0	0.00	0	0	2	6
45 ARTES_GRAFICAS_SENEFELDER_CA	0	0	0	20913.09	0	0	9	2
46 ASESORES_SERVICIOS_Y_COMERCIO_VINDELPO_CIA_LTDA	0	0	0	2500.00	0	0	0	6
47 ASOCIACIÓN_DE_PRODUCCIÓN_AGRÍCOLA_DE_BANANO_Y...	587190	2	0	0.00	0	14	0	6
48 ASOCIACIÓN_DE_PRODUCCIÓN_AGRÍCOLA_TIERRA_VERDE...	1052010	0	0	0.00	0	30	0	6
49 ASOCIACION_DE_PRODUCTORES_NUEVA_ESPERANZA	48048	0	0	0.00	0	0	0	6
50 AULESTIA_QUEVEDO_WILLIAM_IVAN	10000	0	0	0.00	0	0	0	6
51 AVICOLA_AVIFIP_SA	0	0	0	0.00	0	7000	0	6
52 AXIONLOG_ECUADOR_SA	78695	0	0	0.00	0	0	30	6
53 AYALA_MUNOZ_TERESA_DE_JESUS	5000	0	0	0.00	0	0	0	6
54 AZUCARDA_SA	0	2	0	0.00	0	0	0	6
55 BALANCEADOS_NOVA_SA_BALNOVA	40000	0	0	0.00	0	0	0	6
56 BAMAREX_SA	0	0	0	0.00	0	5	0	6
57 BANABAYCORP_SA	100000	0	0	0.00	18	100	0	3
58 BANABIO_SA	182200	0	0	0.00	0	0	20	6
59 BANALCAR_SA_BANALCARSA	111594	0	0	0.00	0	21	0	6
60 BANANA_REAL_REALBANANA_COMERCIALIZADORA_EXPO...	774250	0	0	0.00	0	30	0	6
61 BANANEC_SA	122500	0	0	0.00	0	0	67	6
62 BANASOFT_C_LTDA	0	1	0	0.00	0	0	0	6
63 BANEXCEL_SA	41720	2	0	0.00	0	5	0	6
64 BARROS_VALLEJO_JORGE_ENRIQUE	426000	0	0	0.00	0	11200	3	6
65 BASSIL_HAND_TAILORED_SA	50000	0	0	0.00	0	0	0	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
66 BAZURTO_NARANJO_BETTY_ESPERANZA	15000	0	0	0.00	0	1	0	6
67 BERNAL_MARIA_EULALIA	30000	0	0	0.00	0	0	0	6
68 BIOFACTOR_SA	1215000	2	0	0.00	0	300	0	6
69 BIOFEEDER_SAS	2000	0	0	0.00	0	0	0	6
70 BODEGAS_DEPORTIVAS_ZONE_OUTLET_ZONEOUTLET_SA	122970	0	0	0.00	0	0	24	6
71 BONILLA_MURILLO_LILIANA_DEL_PILAR	50000	0	0	0.00	0	0	0	6
72 BRAVO_VALAREZO_MARIA_AUGUSTA	5000	0	0	0.00	0	0	0	6
73 BRELDYNG_SA	21000	0	0	0.00	0	0	0	6
74 BRPLAST_SA	0	1	0	0.00	0	1	0	6
75 BUENANO_LEON_PEDRO_SAMUEL	0	0	0	0.00	0	0	3	6
76 BURBANO_SANCHEZ_SIXTO_WALFERDIN	0	0	0	0.00	5	0	0	6
77 CALDERON_NOBLECILLA_JULIO_CESAR_ENRIQUE	12000	0	0	0.00	0	0	0	6
78 CALLINGA_JAVIER_HUMBERTO	1372500	2	0	0.00	0	0	12	6
79 CALVA_RODRIGUEZ_FRANCISCO_XAVIER	57500	1	0	0.00	0	0	0	6
80 CAMPELLO_SOSA_DOUGLAS_GABRIEL	3000	0	0	0.00	0	0	0	6
81 CARCHI_ALVARADO_VERONICA_ELIZABETH	12500	0	0	0.00	0	0	8	6
82 CARVAJAL_AVALA_MARIA_DEL_CARMEN	12000	0	0	0.00	0	0	0	6
83 CASA_LUIS_PAZMIÑO_IMPORT_&EXPORT_SA	0	1	0	0.00	0	0	0	6
84 CASTRO_PICO_ISABEL_ELENA	0	0	0	0.00	0	2380	0	6
85 CEGA_INTERNATIONAL_TRADERS_SA	35000	0	0	0.00	0	0	6	6
86 CEGAPACK_SA	3000	0	0	0.00	0	0	1	6
87 CEGASUPPLY_SA	7000	0	0	0.00	0	0	20	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
88 CENTRO_CRISTIANO_DE_GUAYAQUIL	60000	0	0	0.00	0	0	0	6
89 CERVANTES_ANZULES_RUBEN_DARIO	201000	0	0	0.00	0	0	0	6
90 CEVALLOS_FUENTES_AMANDA_MARIELA	7930	0	0	0.00	0	0	0	6
91 CHAN_VERA_MILIN_STEFANY	2000	0	0	0.00	0	0	1	6
92 CHANGKUON_CARRERA_VERONICA_CECILIA	172470	0	0	0.00	0	0	12	6
93 CHANGO_MASABANDA_CLAUDIO_GIOVANNI	70000	0	0	0.00	0	0	0	6
94 CHARCUTERIA_LA_SEGOVIANA_SAS	240	0	0	0.00	0	0	0	6
95 CHAVEZ_TERAN_JAIME_ERNESTO	0	1	0	0.00	0	7	0	6
96 CHEMCROP_SA	0	0	0	0.00	0	8	0	6
97 CHIQUITA_BANANA_ECUADOR_CBBRANDS_SA	878458	12	0	0.00	0	20	0	6
98 CHOEZ_MENDOZA_GILBERTH_ERNESTO	1521020	0	0	0.00	0	0	0	6
99 CHONG_QUI_MOREANO_JOSE_ALEJANDRO	0	1	0	0.00	0	0	0	6
100 CHONG_QUI_MOREANO_MARIO_MARTIN	0	1	0	0.00	0	0	0	6
101 CHUYA_MENDEZ_NELSON_EFRAIN	36000	0	0	0.00	0	0	0	6
102 CILIO_SA	490640	0	0	0.00	0	0	0	6
103 CISNEROS_FLORES_ISRAEL_ALEJANDRO	1000	0	0	0.00	0	0	0	6
104 CLEARVISION_CENTER_SAS	2000	0	0	0.00	0	0	2	6
105 CLINICA_PANAMERICANA_CLUMESA_SA	51000	0	0	0.00	0	2	40	6
106 COELLO_BAJANA_FRANCISCO_HERNAN	68540	0	0	0.00	0	0	1	6
107 COELLO_BARBERAN_RUTH_STEFANY	0	0	0	0.00	0	2000	0	6
108 COMERCIAL_FRANCISCO_COFRANCIS_SA	25000	0	0	0.00	0	0	16	6
109 COMERCIAL_JAPONESA_AUTOMOTRIZ_C_LTDA_(COJAPAN)	130000	0	0	0.00	0	0	0	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
110 COMERCIALIZADORA_GALARZA_&ASOCIADOS_IMPAKAR...	10000	0	0	0.00	0	2	0	6
111 COMERCIALIZADORA_ITALY_COSMETIC_ITALCOSIM_SA	50000	0	0	0.00	0	0	0	6
112 COMERCIALIZADORA_Y_CONSERVADORA_DE_PESCADO_B...	0	1	0	0.00	0	0	0	6
113 COMISARIATOS_VELEZ_BONILLA_SOCIEDAD_ANONIMA	0	0	0	0.00	0	52500	0	1
114 COMPANIA_AGRICOLA_ANGELA_MARIA_SA	0	4	0	0.00	0	0	0	6
115 COMPANIA_AGRICOLA_RIO_VENTANAS_SA_CARIVESA	0	2	0	0.00	0	0	0	6
116 COMPANIA_COMERCIALIZADORA_CARNESLAFINCA_SA	280000	0	0	0.00	0	0	0	6
117 COMPANIA_DE_DESARROLLO_BANANERO_DEL_ECUADOR_...	1162703	0	0	0.00	0	0	0	6
118 COMPANIA_ECUATORIANA_PRODUCTORA_DE_DERIVADOS_...	3000	0	0	0.00	0	0	2	6
119 COMPANIA_ANONIMA_CLINICA_GUAYAQUIL_SERVICIOS_M...	45000	0	0	0.00	0	3	0	6
120 COMPANIA_DE_TRANSPORTE_DE_CARGA_PESADA_MEGALO...	6000	0	0	0.00	0	0	2	6
121 COMPUTRONSA_SA	50000	0	0	0.00	0	0	0	6
122 CONTECON_GUAYAQUIL_SA	100000	0	0	0.00	0	0	0	6
123 COOPERATIVA_DE_PRODUCCION_AGRICOLA_BANANEROS_...	1400	0	0	0.00	0	0	1	6
124 CORPORACION_INDUSTRIAL_ARREAGA&ASOCIADOS_CO...	0	0	0	0.00	0	23000	0	1
125 CORDOVA_PINZA_MAURA_EUIDA	10000	0	0	0.00	14	2	0	3
126 CORPMUNDO_SA	0	1	0	0.00	0	28	0	6
127 CORPORACION_CORNEJO_MONTALVAN_CORMONCOM_SA	6073	0	0	0.00	0	0	0	6
128 CORPORACION_DE_TURISMO_JOUN_VEP_CIA_LTDA	14280	0	0	0.00	0	0	0	6
129 CORPORACION_FAVORITA_CA	0	0	0	0.00	0	8500	1935	7
130 CORREA_ARIAS_PAOLA_ANDREA	40000	0	0	0.00	0	2100	0	6
131 COSAABS_CORPORACION_SAAB_SA	254165	0	0	0.00	0	0	0	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus	
132 CRESPO_BUSTAMANTE_IVAN_ANDRES	750	0	0	0.00	0	0	0	6	
133 CULTIVATION_SOLUTIONS_SA	40000	0	0	0.00	0	0	60	6	
134 CUOREFARMACIAS_DEL_ECUADOR_CIALTDA	100000	0	0	0.00	0	1	0	6	
135 DAFITSA_SA	0	0	0	0.00	0	0	3	6	
136 DAMASCOSWETT_SA	451280	0	0	0.00	0	0	0	6	
137 DE_SAEGER_JOSE_GERMAINE_J_P	100000	0	0	0.00	0	10	0	6	
138 DE_SOUSA_JARAMILLO_ALDO_ARAMIS	0	0	0	0.00	0	2100	0	6	
139 DEALFRUIT_SA	51720	0	0	0.00	0	0	0	6	
140 DELGADO_PONCE_CARMEN_GUADALUPE	30000	0	0	0.00	0	0	5	6	
141 DELIFRUIT_SA	675716	4	0	0.00	0	0	0	6	
142 DELIVERY_EXPRESS_S_A_DELIEXPRESS	0	0	0	0.00	0	7000	0	6	
143 DEPOSITO_DENTAL_NEO-TEC_CIA_LTDA	547512	0	0	0.00	0	0	0	6	
144 DESIGN_LABORATORY_SA_DISIGLAB	10000	0	0	0.00	0	0	0	6	
145 DEVIES_CORP_SA	216000	0	0	0.00	0	0	0	6	
146 DIAZ_ROSA_MARIA	47990	0	0	0.00	0	2100	4	6	
147 DIORVETT_C_LTDA	8080	0	0	0.00	0	0	7	6	
148 DIPASO_SA	75000	0	0	0.00	0	0	0	6	
149 DISTRIBUCIONES_LORTIZ_DIST_LORTIZ_DLORTIZ_SA	34000	0	0	0.00	0	0	0	6	
150 DISTRIBUIDORA_COMERCIAL_LA_RAIZ_DISCOMRAIZ_SA	36000	0	0	0.00	0	0	0	6	
151 DISTRIBUIDORA_DE_ALIMENTOS_CARNICOS_(DIALCA)_CA	3210	0	0	0.00	0	0	0	6	
152 DISTRIBUIDORA_DESCALZI_SA	30020	0	0	0.00	0	0	0	45	6
153 DISTRIBUIDORA_FLEXOGRAFICA_DE_ETIQUETAS_LABELFLEX...	0	0	0	545.00	0	0	0	6	

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
154 DISTRIBUIDORA_JORGE_ROMERO_ROSERO_JRR_CIA_LTDA	70000	0	0	0	0	0	0	6
155 DISTRIBUIDORA_SA	90000	0	0	0	0	0	0	6
156 DOLUPA_CA	1028570	0	0	0	0	0	45	6
157 DORALSTORAGE_SA	80000	2	0	0	0	10	0	6
158 DUAGUI_SA	190493	0	0	0	0	0	269	6
159 DURALUM_SA	40000	0	0	0	0	0	31	6
160 DVS_FRUTAS_SA	51700	0	0	0	0	16	0	6
161 EASYLENS_SA	2000	1	0	0	0	0	2	6
162 ECOBANE_CIA_LTDA	2381460	4	0	0	0	0	45	6
163 ECONOPRINT_SA	0	1	0	0	0	0	0	6
164 ECUALIQUIDOS_SA	60000	1	0	0	0	40	14	6
165 EGAS_VEINTIMILLA_EMILIO_ANDRES	0	0	0	0	0	36	0	6
166 ELECTRONICA_SIGLO_XXI_ELECTROSIGLO_SA	124000	0	0	0	0	0	36	6
167 ELITE_CARGO_SERVICE_ELITECARGOSER_SA	35820	0	0	0	0	0	0	6
168 EMBUTIDOS_EL_COSTENO_EMBUCOSTE_SA	120000	0	0	0	0	0	51	6
169 EMPRESA_PESQUERA_POLAR_SA	597950	0	0	0	0	2	48	6
170 ENCALADA_MORA_CELIA_MARIA	0	0	0	0	0	2	0	6
171 ENDERICA_BRITO_VALERIA_STEPHANIE	5914	0	0	0	0	0	0	6
172 ENVASES_DEL_LITORAL_SA	60000	0	2	0	0	0	0	5
173 ERAZO_MOISES_VICENTE	40000	8	0	0	0	3	0	6
174 ESCALANTE_GARCIA_CHRISTIAN_EDUARDO_	3000	0	0	0	0	0	0	6
175 ESKORURAL_SA	5000	1	0	0	0	0	5	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
176 ESTRADA_GONZALEZ_CRISTHIAN_JAVIER	90000	0	0	0	0	0	0	6
177 ETYECU_SA	29100	0	0	0	0	12	0	6
178 EVERMORCARGO_SA	3500	0	0	0	0	4	0	6
179 EXAMASA_SA	0	2	0	0	0	25	0	6
180 EXPOGLOBAL_SA	78929	0	0	0	0	0	0	6
181 EXPORTACIONES_DUREXPORTA_SA	459210	0	0	0	0	40	0	6
182 EXPORTADORA_AGRICULTORES_BANANEROS_DEL_SUR_SA...	383800	0	0	0	0	20	0	6
183 EXPORTADORA_BANANERA_INDEPENDIENTE_EXPORBAN_SA	570700	14	0	0	0	30	0	6
184 EXPORTADORA_DE_BANANO_ECUATORIANO_DE_CALIDAD...	50640	0	0	0	0	0	0	6
185 EXPORTADORA_DE_FRUTAS_EXPORTVAL_SA	39580	0	0	0	0	4	0	6
186 EXPORTADORA_HOYOS_COBO_EXPORHOYOS_SA	205900	1	0	0	0	10	0	6
187 EXPORTADORA_Y_EMPACADORA_FEI_YANG_FEIYANG_SA	42000	0	0	0	0	0	0	6
188 EXPORTGRID_SA	46650	0	0	0	0	0	0	6
189 EXPORTJAIME_SA	2740000	14	0	0	0	60	0	4
190 FERREMUNDO_SA	1006620	0	0	0	0	7000	198	6
191 FERRETERIA_ESPINOZA	100000	0	0	0	0	0	0	6
192 FERTILIZANTES_DEL_PACIFICO_FERPACIFIC_SA	0	0	0	0	0	10	0	6
193 FERTISA_AGIF_CL	50000	0	0	0	0	0	75	6
194 FERVACARGO_SA	0	0	0	0	0	0	20	6
195 FERVADELIVERY_SA	49000	0	0	0	0	0	8	6
196 FIRESKY_SA	66000	3	0	0	0	0	204	6
197 FRAMAEXPORT_SA	371040	0	0	0	0	0	0	6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
198 FRANCO_MORAN_WALTER_MIGUEL	200000	0	0	0	0	0	0	0 6
199 FRESKBANA_SA	146080	0	0	0	0	0	24	0 6
200 FRUIT_DELIVERY_ECUADOR_EDEFRUITSA_SA	40390	0	0	0	0	0	5	0 6
201 FRULAN_SA	10000	0	0	0	0	0	0	6 6
202 FRUMORE_SA	300000	1	0	0	0	0	100	0 6
203 FRUTADELI_SA	0	0	0	0	0	0	70	0 6
204 FRUTAGRANDE_SA	27860	0	0	0	0	0	10	0 6
205 FRUTERA_DON_MARCOS_FRUDONMAR_SA	100000	0	0	0	0	0	20	0 6
206 FRUTOS_BELLOS_FRUBELL_SA	724462	1	0	0	0	0	10	0 6
207 FUNDACION_DAMAS_DEL_HONORABLE_CUERPO_CONSUL...	100000	0	0	0	0	0	0	0 6
208 GAGLAY_QUINGAGUANO_ANGEL_GEOVANY	300000	0	0	0	0	0	0	24 6
209 GALARZA_GUAMAN_GEORGE_DE_JESUS	50000	0	0	0	0	0	2	0 6
210 GALARZA_VALAREZO_MARCOS_ANTONIO	6500	0	0	0	0	0	0	2 6
211 GALERIA_DE_MUEBLES_GAL_MUEBLES_SA	11070	0	0	0	0	0	0	0 6
212 GAMMA_SALUD_SALUDGA_SA	80500	0	0	0	0	0	0	1 6
213 GARCES_PARRA_HON_DARIO	0	0	0	0	0	0	0	5 6
214 GARCES_ZULUAGA_JULIAN_ESTEBAN	2500	1	0	0	0	0	0	0 6
215 GARCIA_MUNOZ_DEVORA_GALLUTH	0	0	0	0	0	0	0	5 6
216 GIL_MEDINA_CARLOS_JAVIER	0	0	0	0	0	0	0	27 6
217 GINTRACOM_SAS	123288	0	0	0	0	0	0	0 6
218 GISIS_SA	1745000	10	0	0	0	0	0	3019 7
219 GLICE_SA	300000	0	0	0	0	0	8	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
220 GLOBALCOMPU_SA	10000	0	0	0	0	0	0	0 6
221 GLOBECUADOR_SA	23870	0	0	0	0	0	1	10 6
222 GODDCORP_SA	42000	7	0	0	0	0	15	0 6
223 GODOY_USIÑA_HENRY_VINICIO	6000	0	0	0	0	0	0	0 6
224 GOLDKING_SA	38880	1	0	0	0	0	0	12 6
225 GONZALEZ_BONILLA_LUIS_ADRIAN	0	0	0	0	0	0	3	0 6
226 GRAN_COLOMBIANA_DEL_MAR_GRANCOMAR_SA	1000	0	0	0	0	0	0	1 6
227 GRANDA_MENDOZA_STALIN_DAVID	0	2	0	0	0	0	0	0 6
228 GRANOTEC_ECUADOR_SA_GRANOTECUA	6288	0	0	0	0	0	2	48 6
229 GRUPASA_GRUPO_PAPELERO_CIA_LTDA	182000	0	0	0	0	0	0	200 6
230 GUAIGUA_SANDOVAL_GLADYS_ROCIO	0	0	0	0	0	0	4	0 6
231 GUERRERO_JARA_VERONICA_JANETH	2790	0	0	0	0	0	0	0 6
232 GUERRERO_QUIMI_PETER_ADRIAN	0	0	0	0	0	0	2	0 6
233 GUEVARA_TOMASELLY_PEGGY_ANDREA	0	0	0	0	0	0	2	0 6
234 GUIJARRO_BRITO_ANA_LUCRECIA	40000	0	0	0	0	0	0	18 6
235 GUZMAN_PAUCAR_MIGUEL_IVAN	113780	0	0	0	0	0	0	40 6
236 HACIENDA_GUAYABO_GUAYABOSA_SA	0	4	0	0	0	0	0	0 6
237 HACIENDA_LA_GORDITA_HALAGOR_SA	0	1	0	0	0	0	0	0 6
238 HAFREIT_SA	0	2	0	0	0	0	20	0 6
239 HARUKA_SA	0	0	0	0	0	0	0	1 6
240 HILUES_SA	600000	3	0	0	0	14	40	0 3
241 HOVOS_DIAZ_MARIA_DEL_PILAR	629000	0	0	0	0	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	clus
242 ICOLTAG_SA	0	0	0	0	0	0	9	13 6
243 IMDISVAL_SA	11200	0	0	0	0	0	0	3 6
244 IMPORTADORA_BOHORQUEZ_C_LTDA	615000	0	0	0	0	0	88	200 6
245 IMPORTADORA_DE_FERRETERIA_Y_GASES_INFEGAS_SA	15000	0	0	0	0	0	0	2 6
246 IMPORTADORA_MEDIMPO_SA	4500	0	0	0	0	0	0	0 6
247 IMPUBLI_CIA_LTDA	0	0	0	1875	0	0	0	0 6
248 INDUMAD_INDUSTRIA_MADERERA_SA	14620	0	0	0	0	0	0	0 6
249 INDURA_ECUADOR_SA	30000	0	0	0	0	0	42000	65 1
250 INDUSTRIA_ECUATORIANA_DE_CABLES_INCABLE_SA	19000	0	0	0	0	0	0	120 6
251 INDUSTRIA_GRAFICA_FLEXOGRAFIA_ECUATORIANA_FLEXO...	200000	6	0	34813	0	0	0	103 2
252 INDUSTRIAL_LINGESA_SA	322225	0	0	0	0	0	0	72 6
253 INDUSTRIAL_Y_AGRICOLA_CANAS_CA	575562	3	0	0	0	0	60	14 6
254 INDUSTRIAS_PLASTICAS_JOSA_CIALTDA	250000	0	0	0	0	0	0	70 6
255 INMOBILIARIA_INTELCA_CIA_LTDA	20000	0	0	0	0	0	0	0 6
256 INNOVA-MKT_SA	0	0	0	0	0	0	48	0 6
257 INSERMADERA_SA	0	0	0	0	0	0	700	0 6
258 INSTITUTO_SECCULAR_DE_SCHOENSTATT_HERMANAS_DE_M...	0	0	0	0	0	0	0	2 6
259 INSUPLAT_SA	0	0	0	0	0	0	0	30 6
260 INTACO_ECUADOR_SA	100000	0	0	0	0	0	0	100 6
261 INTERBANANA_EXPORT_INTERBANANAEXPORT_SA	790878	0	0	0	0	0	30	0 6
262 INTERNATIONAL_APPAREL_LALANCORD_LALANCORDSA_SA	34680	0	0	0	0	0	5600	4 6
263 INTEROC_SA	175000	0	0	0	0	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
264	INVERSIONES_AGRICOLAS_INKA_BANANA_SAC	373920	0	0	0	0	0	0 6
265	IRENECUADOR_SA	192230	0	0	0	0	0	0 6
266	IZQUIERDO_BASTIDAS_RENE_MAURICIO	0	0	0	0	3500	0	0 6
267	JARA_MOLINA_LIDIA_AZUCENA	0	0	0	0	0	4	0 6
268	JARAMILLO_BALCAZAR_DENNIS	14000	0	0	0	0	0	0 6
269	JARAMILLO_CAICEDO_MARCOS_FABRIZIO	320	0	0	0	0	0	0 6
270	JARAMILLO_LUZON_EDI_GUSTAVO	0	0	0	1280	0	0	0 6
271	JAROMA_SA	0	2	0	0	0	12	0 6
272	JASAFRUT_SA	0	12	0	0	0	70	0 6
273	JEDESCO_SA	0	0	0	0	0	20	0 6
274	JOUN-YEP_SA_S	10080	0	0	0	0	0	0 6
275	JUAN_RAMIREZ_VALAREZO	750	0	0	0	0	0	0 6
276	JUNTA_PROVINCIAL_DE_CRUZ_ROJA_DEL_CANAR	92000	0	0	0	0	0	0 6
277	JUNTA_PROVINCIAL_DE_LA_CRUZ_ROJA_DEL_GUAYAS	751000	0	0	0	0	0	425 6
278	KARCHER_ECUADOR_SA	9000	0	0	0	0	0	0 6
279	KITTON_SA	249510	0	0	0	0	2451	12 6
280	LA_FABRIL_SA	1721000	4	0	0	0	100	0 6
281	LA_MARAVILLA	0	1	0	0	0	8	0 3
282	LABMAC_SA	80000	0	0	0	0	0	0 6
283	LABORATORIO_CLINICO_ALCIVAR_SA_ALCLINIC	425000	0	0	0	0	0	0 6
284	LABORATORIO_CLINICO_ARRIAGA_CA	250000	0	0	0	0	0	0 6
285	LABORATORIO_CLINICO_HCKA_SA	146000	0	0	0	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
286	LABORATORIO_GENESIS_LABGENESIS_CIA_LTDA	0	0	0	0	0	0	10 6
287	LABORATORIOS_GUERRERO_SA	0	0	0	0	0	2	0 6
288	LABORATORIOS_ROCNARF_SA	341900	1	0	0	0	1	362 6
289	LAFATTORIA_SA	20000	0	0	0	0	0	20 6
290	LAM_CHEANG_WILIAM_DAVID	45000	0	0	0	0	0	0 6
291	LASCANO_DURAZNO_ARTURO_EMILIANO	744000	0	0	0	0	0	0 6
292	LATAMFIBERHOME_CABLE_C_LTDA	0	0	0	0	0	210	0 6
293	LATINCOSMETIC_SA	75000	0	0	0	0	0	0 6
294	LAURISABEL_SA	9000	2	0	0	0	1400	4 6
295	LAYDES_HUIÑAPI_JOB	70000	0	0	0	0	0	0 6
296	LAZO_GARCES_SUSAN_GABRIELA	1000	0	0	0	0	0	1 6
297	LEMA_LEON_REYNA_CAROLINA	150500	0	0	0	0	0	0 6
298	LEMA_LEON_SILVIA_VERONICA	55000	0	0	0	0	0	5 6
299	LENOR_ECUADOR_CIA_LTDA	31500	0	0	0	0	0	7 6
300	LEON_YUCTA MARIA_BEATRIZ	280000	0	0	0	0	0	0 6
301	LEON_YUCTA MARTHA_PATRICIA	14000	0	0	0	0	0	0 6
302	LINDE_ECUADOR_SA	500000	0	0	0	0	60	0 6
303	LLANGARI_CACOANGO_BENIAMIN_WASHINGTON	0	0	0	0	0	1	0 6
304	LLANTICENTRO_L&M_SA	0	0	0	0	0	4	0 6
305	LLUGSHA_FERNANDEZ_ARTURO_FREDY	720000	0	0	0	0	0	0 6
306	LOGYTECHMOBILE_SA	200265	0	0	0	0	0	10 6
307	LOPEZ_BURITICA_NICOLAS_EMILIO	120000	0	0	0	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
308	LOPEZ_TOSCANO_MARIO_FRANCISCO	50000	0	0	0	0	0	18 6
309	LUCARZ_SA	5400	0	0	0	0	0	0 6
310	LUCERO_YEPEZ_ORLANDO_XAVIER	0	0	0	690	0	0	0 6
311	LUDERSON_SA	2702699	2	0	0	0	8	0 6
312	LUZURIAGA_MERINO_MARCO_ANTONIO	93340	0	0	0	0	0	0 6
313	MACHARE_MORLAS_KHRISELL_DEVANIRA	2401000	0	0	0	0	0	0 6
314	MACIAS_CHOEZ_GONZALO_SIDNEY	0	1	0	0	1	0	0 6
315	MACOMON_SA	8000	0	0	0	0	0	13 6
316	MAGALLANES_PAREDES_NARCISA_DE_JESUS	0	0	0	0	0	0	2 6
317	MAMUT_ANDINO_CA	15000	0	0	0	0	0	2 6
318	MANUFACTURAS_CONFECCIONES_Y_DISTRIBUCIONES_BUE...	10000	0	0	0	0	0	8 6
319	MARCA_TIGRE_JORGE_MAURICIO	74250	1	0	0	0	0	0 6
320	MARITIME_SERVICIOS_LINE_DEL_ECUADOR_MSL_DEL_ECUAD...	2000	0	0	0	0	0	0 6
321	MARPLANTIS_SA	1228140	0	0	0	0	30	0 6
322	MARRIOTT_SA	5720	0	0	0	0	0	0 6
323	MARTINEZ_CORONEL_EINSTEIN_ECUADOR	68000	0	0	0	0	0	0 6
324	MARULANDA_CAICEDO RICARDO_ANDRES	16000	0	0	0	0	0	0 6
325	MAURA_LOOR_STALYN_ADOLFO	60000	0	0	0	0	7000	21 6
326	MAXIRROZ_SA	51500	0	0	0	0	0	0 6
327	MDX_ECUADOR_SA	30000	0	0	0	0	0	0 6
328	MEDISFARMACO_SA	100000	0	0	0	0	0	0 6
329	MEGAMETALES_SA	300000	0	0	0	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
330 MENACORP_SAS	0	0	0	0	0	0	0	1 6
331 MESIAS_PABON_MARCELO_ROBERTO	0	2	0	0	0	0	0	0 6
332 MICHILENA_VEGA_ALEXANDRA_ELIZABETH	6000	0	0	0	0	0	0	0 6
333 MINANGO_NARVAEZ_MARTHA_SUSANA	109450	0	0	0	0	0	0	0 6
334 MINUCHE_CASTRO_SANDRA_MARIA_LORENA	0	1	0	0	0	0	0	0 6
335 MIRANDA_ERAZO_MIRIAM_YOLANDA	150000	0	0	0	0	0	0	0 6
336 MIRANDA_FLORES_MARIA_TERESA	20000	0	0	0	0	0	2	0 6
337 MIRANDA_PEREZ_PEDRO_FELIX	0	0	0	0	0	0	0	5 6
338 MOLESTINA_PINO_MARCELA_MARIA	45500	0	0	0	0	0	0	0 6
339 MOLINA_VIVANCO_DIEGO_VICENTE	425460	0	0	0	0	0	0	12 6
340 MONSERRATE_LLAMUCA_DOUGLAS_ISAAC	7110	0	0	0	0	0	1	0 6
341 MONTERO_FLORES_LUIS_ESTUARDO	623000	0	0	0	0	0	0	0 6
342 MONTIEL_LOPEZ_SEGUNDO_ISMAEL	83680	0	0	0	0	0	0	48 6
343 MORA_RIVAS_SANDY_ALBERTO	0	1	0	0	0	0	1	0 6
344 MORA_RODRIGUEZ_DIEGO_NICOLAS	204100	0	0	0	0	0	0	0 6
345 MORALES_MONTANO_JUAN_CARLOS	8500	1	0	0	0	0	0	4 6
346 MORALTORR_SA	67480	0	0	0	0	0	0	0 6
347 MOREANO_SANCHEZ_ROSA_MARIA	0	3	0	0	0	0	0	0 6
348 MOUNTAIN_KING_SAS	20000	1	0	0	0	0	0	16 6
349 MUNOZ_SOLORZANO_JESSICA_JOHANA_MONSERRATE	12500	0	0	0	0	0	0	6 6
350 MUNOZ_VELE_JOSE_ALCIBIADES	97500	0	0	0	0	0	0	2 6
351 MUÑOZ_MERA_JONATAN_FABRICIO	75000	0	0	0	0	0	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
352 MURILLO_ALVARADO_ROBERT_DIDI	12500	0	0	0	0	0	0	1 6
353 NAPORTEC_SA	140000	0	0	0	0	8000	0	0 6
354 NAVARRETE_SANCHEZ_JOSE_LUIS	60000	0	0	0	0	0	0	0 6
355 NEGOCIOS_Y_LOGISTICA_NEGOLOGIC_SA	245903	0	0	0	0	0	0	0 6
356 NEW_YORK_DISCOUNT_DISNEW_SA	658000	0	0	0	0	0	0	108 6
357 NIKOTRON_SA	127302	0	0	0	0	1400	0	0 6
358 NOBOA_TRADING_CO_TCN_SA	2785634	0	0	0	0	125	0	0 6
359 NORVIDA_SA	137500	0	0	0	0	0	0	0 6
360 NOVAMODA_SA	81620	0	0	0	0	0	0	8 6
361 NUNEZ_RAMON_PEDRO_SIMON	3400	0	0	0	0	0	0	15 6
362 OBSA_ORO_BANANA_S_A	1777200	34	0	0	0	50	0	0 4
363 OCAMPO_MACAY_EDISON_ANDRES	459700	0	0	0	0	0	0	0 6
364 OCHOA_CHACON_EDISON_CAVETANO	0	0	0	0	0	3	0	0 6
365 OLEAS_SANCHEZ_INVERSIONES_OLEXSAIMPORT_SA	180000	0	0	0	0	0	0	0 6
366 ORDOÑEZ_CARRION_GABRIEL_EMILIO	4200	0	0	0	0	0	0	0 6
367 ORELLANA_BERNAL_FABRICIO_ALBERTO	61800	0	0	0	0	1	0	0 6
368 ORMAZA_ANCHUNDIA_ELSA_GUADALUPE	0	2	0	0	0	0	0	0 6
369 ORODELT_SA	0	4	0	0	0	6	60	0 6
370 ORTIZ_AGUIRRE_ALEX_JAVIER	0	1	0	0	0	0	2	0 6
371 OTISGRAF_SA	121620	6	0	0	0	6	0	0 6
372 PACHAR_PLAZA_JULIO_CESAR	30980	0	0	0	0	0	0	15 6
373 PANIFICADORA_INDUSTRIAL_CIA_LTDA	0	0	0	0	0	15	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
374 PAPELEK_CIALTDA	0	0	0	0	0	0	0	22 6
375 PAPELERA_NACIONAL_SA	888000	0	0	0	0	0	0	54 6
376 PAPELESA_C_LTDA	4241000	0	0	0	0	0	0	154 4
377 PAREDES_CAMPOVERDE_GUILLERMO_ENRIQUE	130000	0	0	0	0	7000	0	0 6
378 PAREDES_ROMERO_MICHAEL_JAMIL	70000	0	0	0	0	0	0	0 6
379 PASTEBE_SA	159266	0	0	0	0	10	2	9 3
380 PAUL_DE_MAR_SA	0	3	0	0	0	200	0	0 6
381 PAZMINO_ALVARADO_ANGEL_RAMIRO	18000	0	0	0	0	0	1	0 6
382 PAZMIÑO_BARCENES_JOSELYNE_ADRIANA	0	0	0	0	0	4	0	0 6
383 PECES_Y_CRUSTACEOS_DEL_PACIFICO_PEZCRUP_SA	30870	0	0	0	0	1	0	5 6
384 PESANTEZ_LOPEZ_MIGUEL_ALBERTO	90000	0	0	0	0	0	0	6 6
385 PIN_BAQUE_GRECIA_MARIBEL	30000	0	0	0	0	0	0	0 6
386 PINAS_RICAS_DEL_ECUADOR_SA_PIRECUASA	30000	0	0	0	0	0	0	0 6
387 PINEDA_CORAL_JAIRO_DARIO	0	0	2	0	0	0	0	0 5
388 PINEDA_GUALOTO_MELANY_GEANELLA	6000	0	0	0	0	0	0	11 6
389 PINTURAS_UNIDAS_SA	5480570	1	1	0	2	288	470	4 6
390 PLASTICOS_DEL_LITORAL_PLASTLIT_SA	0	0	0	0	0	154	100	6 6
391 PLASTIVILL_SA	105000	1	0	0	0	0	0	149 6
392 POFIDEL_SA	111680	0	0	0	0	0	0	0 6
393 POLIFIBRAS_DEL_ECUADOR_SA_POLIFECSA	82000	0	0	0	0	0	0	0 6
394 PONCE_LEON_TOMAS_ANDRES	2500	0	0	0	0	0	0	2 6
395 PREDIOS_RUSTICOS_LA_RURAL_COMPANIA_ANONIMA	933062	0	0	0	0	15	0	0 6

CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
395 PREDIOS_RUSTICOS_LA_RURAL_COMPANIA_ANONIMA	933062	0	0	0	0	15	0	6
396 PROALUM_SA	30000	0	0	0	0	0	15	6
397 PRODEX_CIA_LTDA	0	0	0	0	1	0	0	6
398 PRODUCTORA_Y_COMERCIALIZADORA_EMAUS_SCC	108400	0	0	0	0	0	0	6
399 PRODUCTOS_DEL_AGRO_SYLVIA_MARIA_SA_AGROSYLMA	200000	0	0	0	0	0	197	6
400 PRODUCTOS_METALURGICOS_SA_PROMESA	1908890	0	0	0	0	0	1229	7
401 PRODUCTOS_Y_SERVICIOS_GRAFICOS_PROSERGRAF_SC&C	0	0	0	0	0	0	26	6
402 PROLUMIC_SA	32550	1	0	0	0	0	0	6
403 PROMOTICK_SA	60000	0	0	0	0	0	3	6
404 PROPALCOL_SA	95900	1	0	0	0	1	40	6
405 PUCO_BUENAÑO_MARCO_ANTONIO	10000	1	0	0	0	0	0	6
406 PUCO_TOAPANTA_MARCO_ANTONIO	200000	0	0	0	0	10	0	6
407 QUIMIAVIL_SA	562000	0	0	0	0	0	0	6
408 QUINDE_LAVAYEN_MARCO_ANTONIO	875000	0	0	0	0	8	0	6
409 QUINDE_PIZA_FERNANDO_VICENTE	54000	0	0	0	0	0	0	6
410 QUINTO_RONQUILLO_FREDDY_LUIS	0	0	0	20240	0	67	0	2
411 RED_DE_FARMACIAS_RED FARMA_SAS	0	0	2	0	0	0	0	5
412 REPREMARVA_CIA_LTDA	50000	0	0	0	0	0	0	6
413 REPRESENTACIONES_ADHESIVOS_&_FILM_SA_REPFILSA	0	0	0	0	0	15	0	6
414 REPUBLICNEG_SA	50000	3	0	0	0	50	0	6
415 RESULCONSUL_SA	107800	0	0	0	0	0	0	6
416 REVERSE_OSMOSIS_SYSTEM_SA_REVOSYS	4000	0	0	0	0	0	0	6

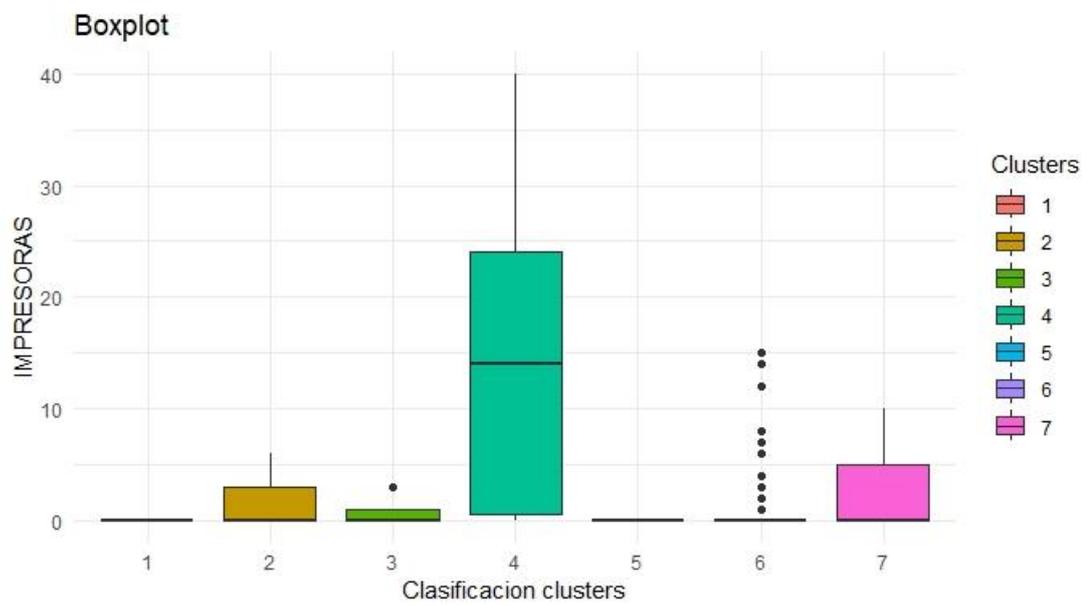
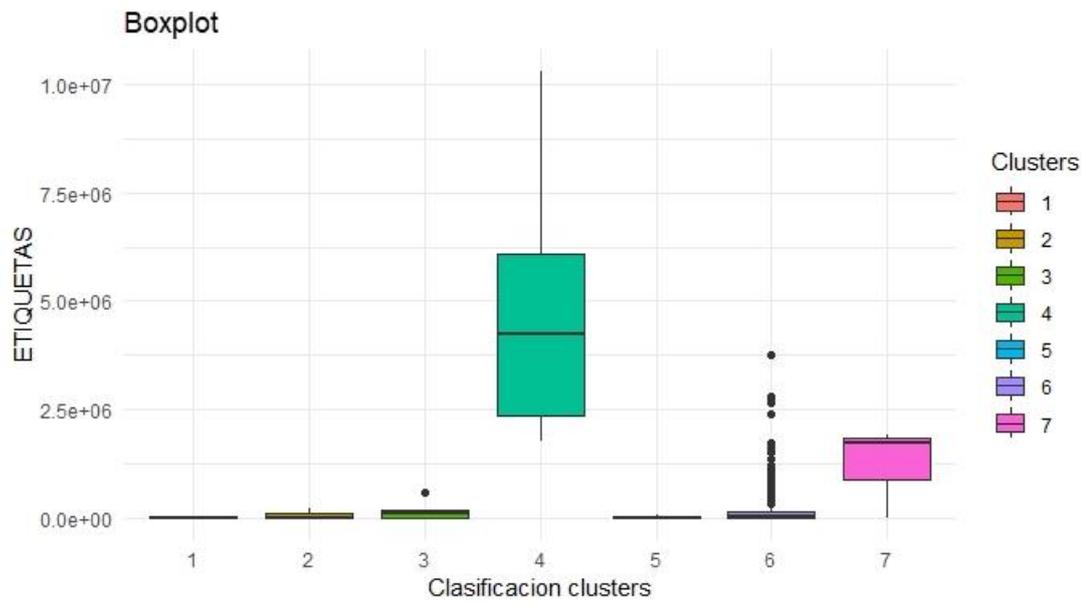
CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
417 REYBANPAC_REY_BANANO_DEL_PACIFICO_CL	2004723	14	0	0	0	30	0	4
418 RICOMARSA_SA	200000	0	0	0	0	10	0	6
419 RISK_PROTECTION_RISKPRO_SA	13000	0	0	0	0	0	4	6
420 RODRIGUEZ_CASTELLANO_JESSICA_JUDITH	0	0	0	480	0	4	72	6
421 RODRIGUEZ_ROMERO_KARINA_SALOME	15000	0	0	0	0	0	2	6
422 Rojas_Mendez_Wendy_Beatriz	30000	0	0	0	0	0	3	6
423 ROJAS_NAVAS_MARGARITA_DEL_CARMEN	7125	0	0	0	0	0	2	6
424 ROJAS_ORDOÑEZ_MARLENY_EMPERATRIZ	25000	0	0	0	0	0	0	6
425 ROMAN_LINO_DIXON_ARMANDO	0	1	0	0	0	0	0	6
426 ROMERO_NUNEZ_MILTON_DANIEL	0	0	0	0	0	4	0	6
427 RON_TOLEDO_MONICA_XIMENA	1000	0	0	0	0	0	0	6
428 ROSERO_BALSECA_SILVIA_TATIANA	0	0	0	0	0	100	0	6
429 RUBIO_CELLERI_YESENNIA_ELIZABETH	1000	0	0	0	0	0	0	6
430 RUFORCORP_SA	20000	15	0	0	0	20	0	6
431 RUJANO_SALAZAR_JOSE_MANUEL	0	0	0	0	0	8	0	6
432 SABANDO_CHERREZ_MATIAS_ISRAEL	2500	1	0	0	0	0	1	6
433 SABROSTAR_FRUIT_COMPANY_SA	501000	15	0	0	0	30	0	6
434 SACOS_DURAN_REYSAC_SA	223590	0	0	0	0	1	315	6
435 SAILORPAINT_SA	70000	0	0	0	0	0	25	6
436 SALAPISA_SA	0	1	0	0	0	0	0	6
437 SALAZAR_CUEVA_GABY_MARIUXI	5000	0	0	0	0	0	6	6
438 SALCEDO_FAYTONG_CAROLINA_ESTEFANIA	0	1	0	0	0	1	0	6

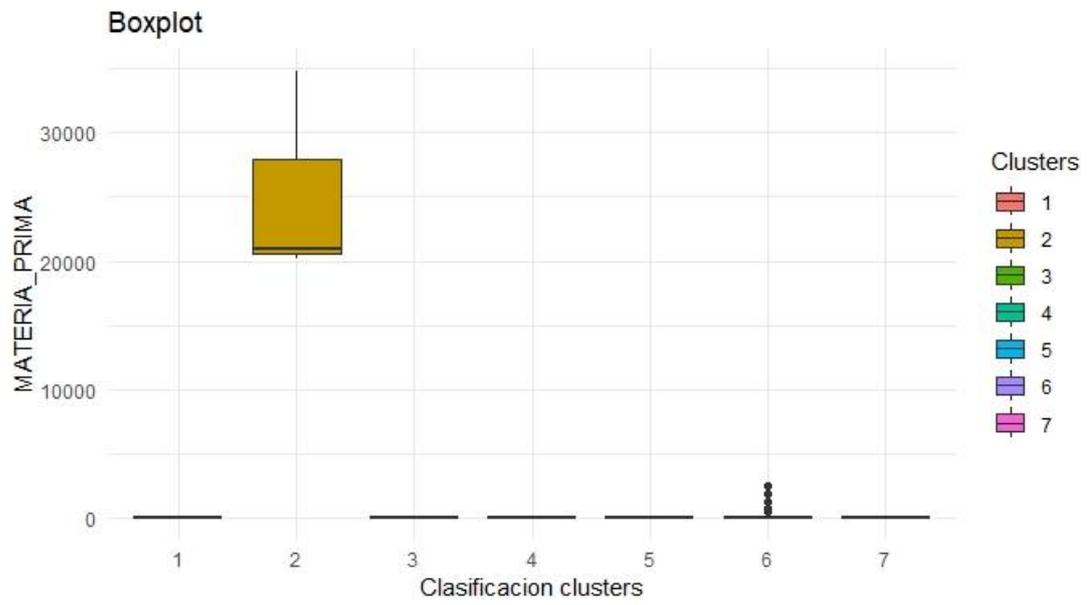
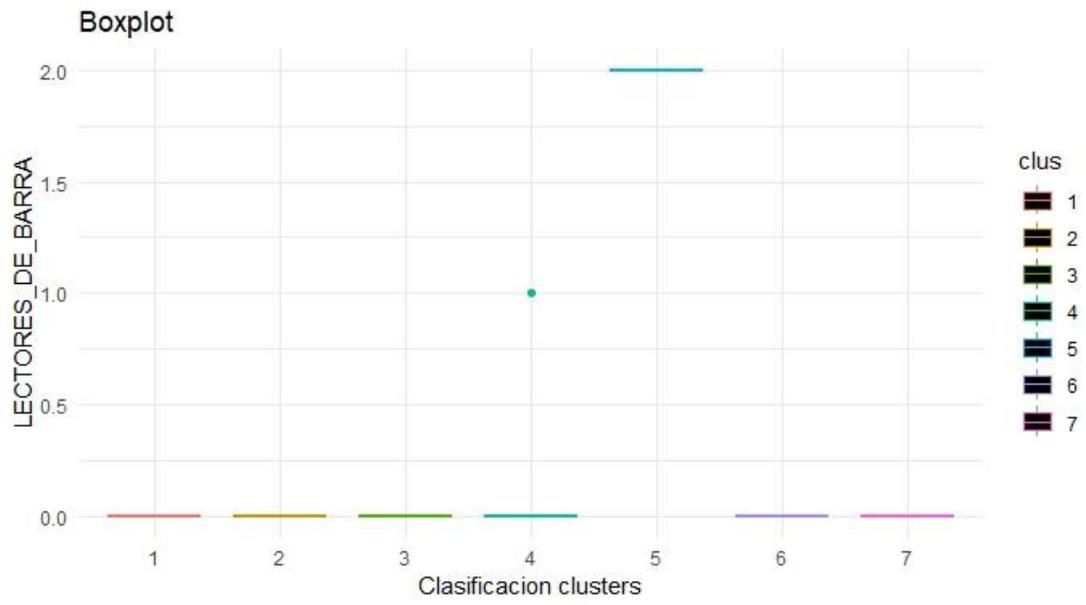
CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
439 SALVACERO_CIA_LTDA	2500	1	0	0	0	0	2	6
440 SAMANESA_SA	0	1	0	0	0	0	0	6
441 SANTANA_SOFTWARE_SASOF_SA	0	0	0	0	0	7000	0	6
442 SCHNIER_SVEN	7700	0	0	0	0	0	0	6
443 SEAKLAIR_SA	50000	0	0	0	0	0	0	6
444 SEGRATI_SA	140000	0	0	0	0	0	156	6
445 SELLAN_CASTRO_MARGARITA_PILAR	58700	0	0	0	0	0	0	6
446 SEMOLA_S_A	826000	1	0	0	5	37	22	6
447 SERVICIOS_DENTALES_SERVIDENT_SA	578000	0	0	0	0	0	0	6
448 SERVICIOS_MEDICOS_INTEGRALES_POPULARES_SA_SERVIMI...	167300	0	0	0	0	0	0	6
449 SERVIREPUESTOS_LOZADA_SA_SERVILOZA	86000	0	0	0	0	0	21	6
450 SIDELSUR_SA	30000	0	0	0	0	0	0	6
451 SIMED_S_A	3743926	0	0	0	0	9050	0	6
452 SISMODE_CIA_LTDA	0	0	0	0	0	40	0	6
453 SOCIEDAD_DE_HECHO_OROAGUA	0	1	0	0	0	4	0	6
454 SOCIEDAD_EN_PREDIOS_RUSTICOS_AGRPECUARIA_EL_TR...	0	0	0	0	1	0	0	6
455 SOITGAR_SA	64190	0	0	0	0	0	0	6
456 SOLMEDA_SA	0	1	0	0	0	2	0	6
457 SOLUCIONES_INTEGRALES_PARA_SUPERFICIES_NEOPINDEC...	132080	1	0	0	0	360	95	6
458 SOLUCIONES_TECNICAS_E_INGENIERIA_SA_SOTECINGENIER...	1546	0	0	0	0	0	0	6
459 SOPRISA_SA	1536970	0	0	0	0	10	0	6
460 SPARTAN_DEL_ECUADOR_PRODUCTOS_QUIMICOS_SA	0	0	0	0	0	0	100	6

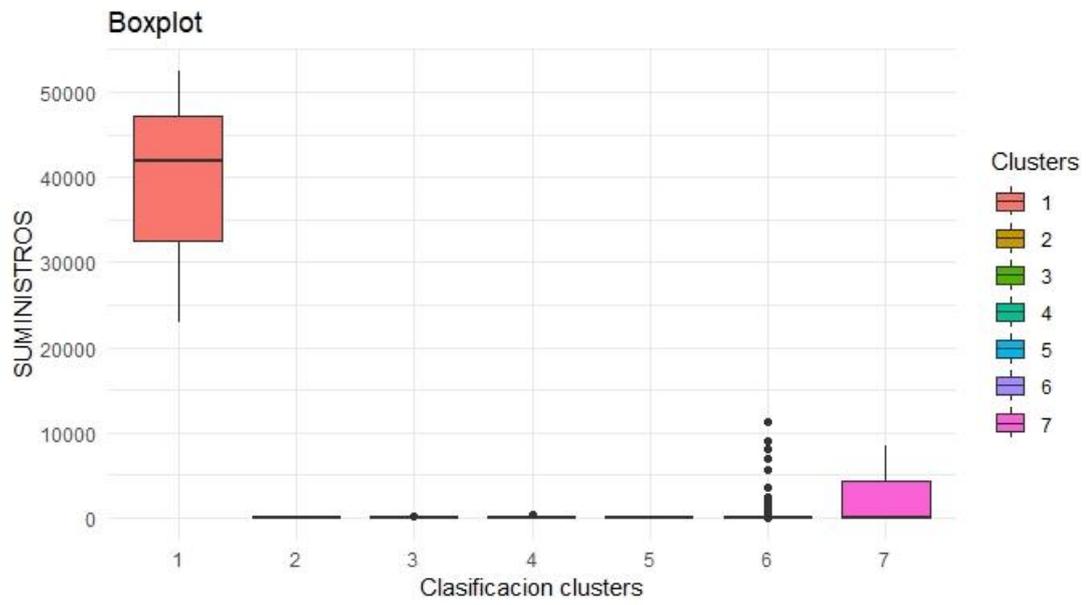
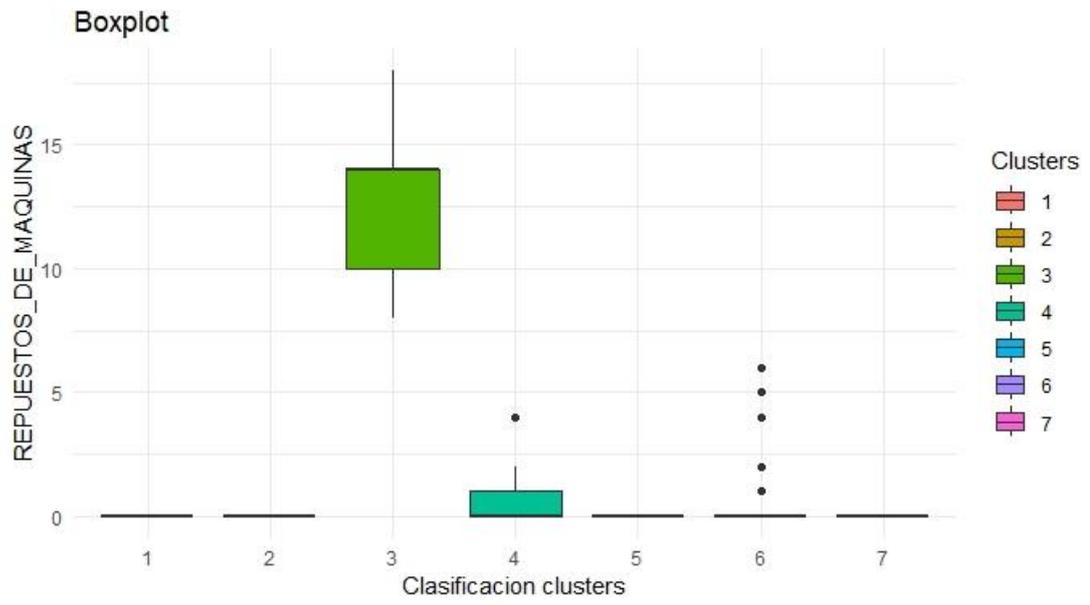
CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
461 STOWHAS_SIMBALL_CARLOS_ALBERTO	0	0	0	0	0	1400	0	6
462 SUPRALIVE_SA	463720	0	0	0	0	0	267	6
463 SUPRINSA_SU_PROVEEDOR_INDUSTRIAL_S_A	30000	1	0	0	0	0	52	6
464 SWISSOIL_DEL_ECUADOR_SA_SWISSOIL	0	0	0	0	0	2000	12	6
465 TAPIA_MARTINEZ_GABRIEL	0	1	0	0	0	0	1	6
466 TECNOLOGISTICA_TLG_SA	0	0	0	0	0	1	0	6
467 TECNOQUIMICAS_DEL_ECUADOR_SA	0	0	0	0	0	204	0	6
468 TELECOMPU_SA	63000	0	0	0	0	0	0	6
469 TERMINAL_DE_CARGAS_DEL_ECUADOR_SA_TERMICARGA	115730	1	0	0	0	0	63	6
470 TERMOEK_TERMOCOGIBLES_Y_EMPAQUES_SA	172000	0	0	0	0	0	0	6
471 TERRAENCANTO_SAS	0	0	0	0	0	5	0	6
472 TEXFASHION_SA	9140	0	0	0	0	0	0	6
473 TICKETSHOW_SA	21230	0	0	0	0	4	1050	6
474 TIENDAS_INDUSTRIALES_ASOCIADAS_TIA_S_A	10305000	0	0	0	0	0	0	4
475 TIENDAS_TUTI_TTDE_SA	6676160	40	0	0	0	4	0	4
476 TINI_TREND_SAS	20000	1	0	0	0	0	0	6
477 TORO_YEPEZ_GONZALO	114000	0	0	0	0	0	0	6
478 TRANSCITY_SA	149000	0	0	0	0	0	0	6
479 TROPICALAGRO_SA	336290	0	0	0	0	0	0	6
480 TUCHOK_SA	111960	4	0	0	0	0	15	6
481 UNIDAD_EDUCATIVA_SAN_FRANCISCO_DE_ASIS_GUAYAS	4500	0	0	0	0	0	0	6
482 UNIFRUTTI_ECUADOR_SA	325000	0	0	0	0	0	0	6

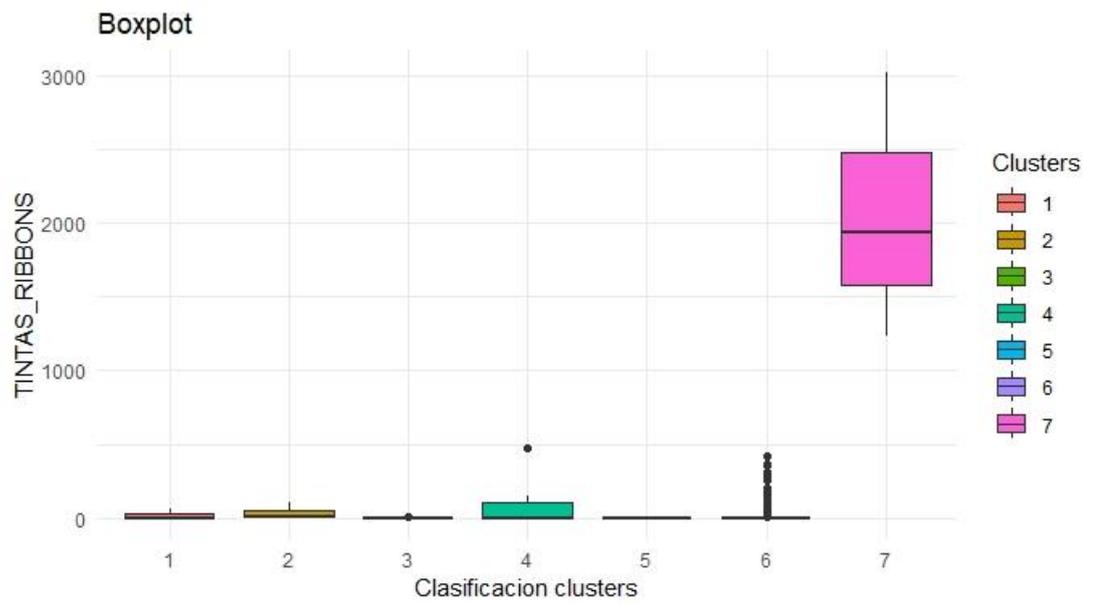
CLIENTES	ETIQUETAS	IMPRESORAS	LECTORES_DE_BARRA	MATERIA_PRIMA	REPUESTOS_DE_MAQUINAS	SUMINISTROS	TINTAS_RIBBONS	ctus
483 UNITEDEXPORTS_SA	89440	0	0	0	0	0	5	6
484 UTOPIA_MAYORISTA_SAS	2790	0	0	0	0	0	0	6
485 VALDERREY_LOROÑO_MICHAEL_DEIVI	0	0	0	0	0	0	0	6
486 VALENCIA_COLOBON_EDMUNDO_NELFOR	0	0	0	0	0	0	16	6
487 VALENCIA_HERNANDEZ_EDDIE_BISMARCK	17500	1	0	0	0	0	0	6
488 VALVERDE_GUEVARA_JAVIER_GIOVANNI	0	3	0	0	0	0	0	6
489 VELASCO_CHONATA_MARIA_MARGARITA	10000	0	0	0	0	0	0	6
490 VELASQUEZ_PROCEL_LUIS_OSWALDO	300000	0	0	0	0	0	0	6
491 VERA_BALSECA_ROMEO_SEGUNDO	45000	0	0	0	0	0	0	6
492 VERA_DEMERA_DIANA_KAROLINA	30000	0	0	0	0	0	0	6
493 VILLA_ZHAGUI_MAURO_LEONCIO	40000	1	0	0	0	0	2	6
494 VILLAO_VELEZ_VERONICA_ELIZABETH	0	0	0	0	0	0	4	6
495 VILLAVICENCIO_NARVAEZ_SANTIAGO_FRANCISCO	0	0	0	0	0	0	1	6
496 VIVAS_SALTOS_LUIS_JAVIER	15000	0	0	0	0	0	0	6
497 VIVIFY_DEL_ECUADOR_SA	800	1	0	0	0	0	36	6
498 WINDOW_WORLD_SA	145000	0	0	0	0	0	0	6
499 ZAMBRANO_MACIAS_LUIS_FRANCISCO_GREGORIO	400000	0	0	0	0	0	0	6
500 ZAMORA_ALVIA_MARIA_ELENA	37500	0	0	0	0	0	0	6
501 ZAPVELCORP_SA	577500	0	0	0	0	0	0	6

Anexo 3 Diagrama de cajas por clúster con respecto a cada categoría de producto











DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **Castillo Moncada Rogger Francisco** con C.C: #0953518545 y **Vera Álava, Andy Jahir**, con C.C: #1316564788 autores del trabajo de integración curricular: **Análisis de clasificación por tipo de cliente de la empresa Labelprintsa S.A. en base a un modelo de clusterización por Machine Learning** previo a la obtención del título de **Licenciado en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaramos 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.- Autorizamos 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, **07 de febrero de 2023**

f. _____
Castillo Moncada, Rogger Francisco

C.C: **0953518545**

f. _____
Vera Álava, Andy Jahir

C.C: **1316564788**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TRABAJO DE INTEGRACIÓN CURRICULAR

TEMA Y SUBTEMA:	Análisis de clasificación por tipo de cliente de la empresa Labelprintsa S.A. en base a un modelo de clusterización por Machine Learning		
AUTOR(ES)	Castillo Moncada, Rogger Francisco y Vera Álava, Andy Jahir		
REVISOR(ES)/TUTOR(ES)	Carrera Buri, Félix Miguel		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Economía y Empresa		
CARRERA:	Negocios Internacionales		
TÍTULO OBTENIDO:	Licenciado en Negocios Internacionales		
FECHA DE PUBLICACIÓN:	07 de febrero de 2023	No. DE PÁGINAS:	101
ÁREAS TEMÁTICAS:	Tecnología, Innovación, Automatización		
PALABRAS CLAVES/ KEYWORDS:	Tecnología, Innovación, Automatización, Análisis de datos, Inteligencia artificial, aprendizaje automático		
RESUMEN/ABSTRACT:	<p>La tecnología y la innovación son fundamentales para el crecimiento y competitividad de las empresas en la actualidad, en donde la automatización de procesos mediante el análisis de datos es una herramienta clave para mejorar la eficiencia y la calidad del servicio ofrecido a los clientes. Sin embargo, el manejo y entendimiento de conceptos de inteligencia artificial, como el aprendizaje automático, sigue siendo un desafío debido a la complejidad de estas tecnologías y a factores prácticos como costos, personal, formación y postura de los empresarios. A pesar de estos desafíos, es esencial que las empresas se adapten a la digitalización de los procesos para poder competir en los mercados globales. La clasificación y segmentación de clientes son técnicas fundamentales para las empresas, ya que permiten adaptar las estrategias y campañas de marketing de manera más personalizada a los diferentes tipos de clientes objetivos, y establecer estrategias eficientes para lograr una ventaja competitiva en el mercado. La presente investigación se enfoca en el desarrollo de un modelo de clasificación de clientes utilizando Machine Learning (ML) y clusterización para la empresa Labelprintsa. El objetivo es establecer estrategias personalizadas para diferentes grupos de clientes debido al aumento de la demanda de métodos de impresión y al crecimiento de la empresa. La aplicación de ML se justifica debido al crecimiento de la empresa y la necesidad de manejar una gran cantidad de datos para aplicar estrategias de retención o adquisición de clientes.</p>		
ADJUNTO PDF:	SI <input checked="" type="checkbox"/>	NO	
CONTACTO CON AUTOR/ES:	Teléfono: +593-98 165 6902 +593 99 372 5248	E-mail: roggercastillo0@gmail.com Andyvera2992@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):			