

**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS ECONÓMICAS, ADMINISTRATIVAS Y
EMPRESARIALES
CARRERA DE NEGOCIOS INTERNACIONALES**

TÍTULO:

**Análisis de la clasificación basado en el concepto Machine Learning por
tipo de cliente en la empresa Tecopesca C.A**

AUTORES

Camacho Rosero Micaella Patricia

Reyna Muentes Cristhian Aurelio

**Trabajo de integración curricular previo a la obtención del título de
Licenciado en Negocios Internacionales**

TUTOR:

Ing. Carrera Buri, Félix Miguel Mgs.

Guayaquil, Ecuador

21 de febrero del 2022



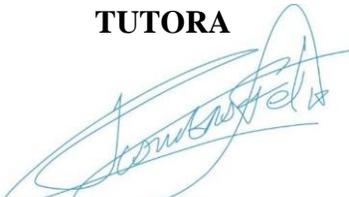
UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE CIENCIAS ECONÓMICAS, ADMINISTRATIVAS Y
EMPRESARIALES
CARRERA DE NEGOCIOS INTERNACIONALES**

CERTIFICACIÓN

Certificamos que el presente trabajo de integración curricular, fue realizado en su totalidad por **Camacho Rosero Micaella Patricia y Reyna Muentes Cristhian Aurelio**, como requerimiento para la obtención del título de en **Negocios Internacionales**.

TUTORA

f. 

Ing. Carrera Buri, Félix Miguel Mgs.

DIRECTOR DE LA CARRERA

f. 

Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.

Guayaquil, a los 21 días del mes de febrero del año 2022



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE CIENCIAS ECONÓMICAS, ADMINISTRATIVAS Y
EMPRESARIALES**

CARRERA DE NEGOCIOS INTERNACIONALES

DECLARACIÓN DE RESPONSABILIDAD

Nosotros, **Camacho Rosero Micaella Patricia**
Reyna Muentes Cristhian Aurelio

DECLARAMOS QUE:

El Trabajo de Integración Curricular, **Análisis de la clasificación basado en el concepto Machine Learning por tipo de cliente en la empresa Tecopesca C.A** previo a la obtención del título de **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 Titulación referido.

Guayaquil, a los 21 días del mes de febrero del año 2022

AUTORES

f. _____
Reyna Muentes Cristhian Aurelio

f. _____
Camacho Rosero Micaella Patricia



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE CIENCIAS ECONÓMICAS, ADMINISTRATIVAS Y
EMPRESARIALES
CARRERA DE NEGOCIOS INTERNACIONALES**

AUTORIZACIÓN

Nosotros, **Camacho Rosero Micaella Patricia**
Reyna Muentes Cristhian Aurelio

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 la clasificación basado en el concepto Machine Learning por tipo de cliente en la empresa Tecopesca C.A**, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y total autoría.

Guayaquil, a los 21 del mes de febrero del año 2022

f. _____
Reyna Muentes Cristhian Aurelio

f. _____
Camacho Rosero Micaella Patricia



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**
**FACULTAD DE CIENCIAS ECONÓMICAS, ADMINISTRATIVAS Y
EMPRESARIALES**
CARRERA DE NEGOCIOS INTERNACIONALES
REPORTE URKUND

← → ↻ 🔒 secure.orkund.com/old/view/121863012-960032-674408#q1bKLVayio7VUSrOTM/LTMt

URKUND

Documento	REYNAMUENTES CHRISTIANAURELIO Y CAMACHOROSERO MICAELLAPATRICIA FINAL.docx (D127738726)
Presentado	2022-02-12 16:46 (-05:00)
Presentado por	cr11sthianreyna@outlook.com
Recibido	felix.carrera01.ucsg@analysis.orkund.com
Mensaje	REYNA MUENTES CRISTHIAN AURELIO CAMACHO ROSERO MICAELLA PATRICIA TESIS FINAL Mostrar el mensaje completo 0% de estas 27 páginas, se componen de texto presente en 0 fuentes.

TUTORA

f. _____

Ing. Carrera Buri Félix Miguel

AGRADECIMIENTO

Este trabajo está dedicado especialmente a mis padres, hermanos y compañero de vida. Fueron un gran apoyo fundamental a lo largo de estos años.

Una especial dedicatoria a mi papá y mamá ya que fueron las personas que creyeron en mí y me ayudaron durante todo el camino. Son los pilares en mi vida y son los causantes de la persona que soy hoy en día.

-Micaella Camacho

Quiero agradecer principalmente a Dios por que gracias a el he cumplido todos los objetivos que me propongo y por llenarme de bendiciones y ponerme en este camino y en esta carrera.

Agradezco a mi madre por su apoyo en toda mi vida por formarme y por disciplinarme siempre por esforzarse para darme todo y que nunca me falte nada, a mi hermano y hermana que me han acompañado siempre, a mi familia que nunca me hizo falta en todos lo momentos especiales de mi vida, a mi Padre y a mi Abuelo que fueron gran parte de mi vida y que actualmente ya no se encuentran conmigo, pero fueron un gran ejemplo para mi y nunca dejaron de creer en mi y en mis capacidades, A todas mis tías que son mis segundas madres y a mi abuela que es una madre más, a mis primos y tíos que han sido gran parte de mi vida.

Adicional quiero agradecer a todos mis compañeros que me acompañaron durante toda esta etapa agradezco por todo lo que viví con ellos y su apoyo siempre.

Finalmente, a mi tutor, agradecerle por todo el conocimiento que me otorgo, por sus enseñanzas y por explotar mis capacidades dentro de mi vida universitaria.

-Cristhian Reyna

DEDICATORIA

Mamá, papá, hermanos y amigos personales en mi vida, este nuevo logro es en gran parte gracias a ustedes, he logrado concluir con éxito un proyecto importante para mi carrera estudiantil. Quisiera dedicar mi tesis a ustedes, personas de bien, seres que ofrecen amor, bienestar y paz.

-Micaella Camacho

Quisiera dedicar este trabajo de investigación principalmente a Dios, a todas las personas que creyeron en mi y en mis capacidades siempre, aquellas que me dieron las oportunidades de crecer en los distintos ámbitos de la vida, a mis padres y familia que me han acompañado a aquellas personas que a pesar de no llevar lazos de sangre en mi vida se han convertido en familiares y me han hecho la persona que soy hoy en día a todos aquellos que me han dado lecciones en la vida y me han formado como la persona que soy actualmente.

Finalmente dedico este trabajo a todos aquellos que han estado junto a mí en este proceso a mis padres y abuelos, primos y tías y a mi pareja y su familia, y a todos mis amigos que me acompañaron en este proceso.

-Cristhian Reyna



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE CIENCIAS ECONÓMICAS, ADMINISTRATIVAS Y
EMPRESARIALES
CARRERA DE NEGOCIOS INTERNACIONALES**

TRIBUNAL DE SUSTENTACIÓN

f. *Gabriela Hurtado*

ING. Hurtado Cevallos Gabriela Elizabeth MGS.

DIRECTOR DE CARRERA

f. *Augusto*

ING. Castro Peñarreta Ángel Aurelio MGS.

COORDINADOR DEL ÁREA

f. *Lucin Castillo*

EC. Lucin Castillo Virginia Carolina MGS.

OPONENTE

Tabla de Contenido

Introducción.....	2
Introducción.....	2
Justificación.....	4
Problemática.....	6
Alcance.....	9
Objetivos.....	10
Objetivo General.....	10
Objetivo Específico.....	10
Hipótesis.....	10
Operacionalización de las Variables.....	11
Capítulo II: Marco Teórico.....	11
Marco Teórico.....	11
Inteligencia Empresarial.....	11
Big Data.....	13
Cloud Computing.....	13
Vulnerabilidades y Riesgos de Trabajar en la Nube.....	14
Sistema de Soporte de Decisiones.....	15
Data Marts.....	15
Data Warehouse.....	16
Data Mining.....	17

Cuadrante Mágico de Gartner para Herramientas BI.....	18
Cubo OLAP	18
Machine Learning.....	19
Deep Learning	19
Redes Neuronales	20
Python.....	21
Keras.....	22
Redes Feedforward.....	23
Redes Convolucionales	24
Marco Conceptual.....	25
Algoritmos de Machine Learning.....	25
Grupos de Algoritmos de Machine Learning	25
Tipos de Algoritmos del Machine Learning.....	27
Marco Legal.....	31
Capítulo III: Metodología.....	31
Metodología.....	31
Pasos a Realizar	32
Conclusiones.....	45
Recomendaciones	46
Referencias	47
Anexos	51

Lista de Tablas

Tabla 1. <i>Operacionalización de las variables a utilizarse en el modelo</i>	11
Tabla 2. <i>Predicción del modelo</i>	36
Tabla 3. <i>Tabla con variable y predicción</i>	37
Tabla 4. <i>Predicción del algoritmo</i>	42
Tabla 5. <i>VARIABLES con predicción</i>	42
Tabla 6. <i>Matriz de confusión con probabilidad</i>	43
Tabla 7. <i>Matriz de confusión actual vs. predicho</i>	44

Lista de Figuras

Figura 1. <i>Algoritmo de Machine Learning</i>	9
Figura 2. <i>Proceso de un Data Mart</i>	16
Figura 3. <i>Proceso de recopilación de data</i>	17
Figura 4. <i>Funcionamiento del cubo OLAP</i>	19
Figura 5. <i>Red neuronal con capa oculta</i>	21
Figura 6. <i>Red neuronal Keras</i>	23
Figura 7. <i>Red neuronal</i>	24
Figura 8. <i>Gráfico de regresión</i>	27
Figura 9. <i>Red neuronal simple</i>	30
Figura 10. <i>Árbol de clasificación</i>	38

Resumen

La era digital se enfoca en las nuevas tecnologías, lo que implica profundos cambios y transformaciones en la sociedad en un mundo globalizado. Esto se manifiesta en una verdadera revolución tecnológica. La inteligencia artificial es parte de esta era, es la combinación de algoritmos propuestos para crear máquinas con capacidades similares a las de los humanos. Tiene aplicaciones en varios campos como la salud, las finanzas, el transporte y la educación.

Machine Learning es un método de análisis de datos que crea automáticamente modelos analíticos. Esta es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con una mínima intervención humana.

El propósito de este trabajo fue analizar la aplicabilidad de Machine Learning para clasificar a los clientes en función de las ventas totales y volumen de ventas en dólares representadas en cajas en la industria de conservas de atún. El análisis se desarrollará utilizando árboles clasificados y bosques aleatorios, mediante la recopilación de datos de la empresa TECOPESCA C.A.

Palabras claves: Machine Learning, árboles clasificados, bosques aleatorios, inteligencia artificial.

Abstract

The digital age focuses on new technologies, which implies profound changes and transformations in society in a globalized world. This manifests itself in a true technological revolution. Artificial intelligence is a similar part of this era, it is the combination of algorithms proposed to create machines with human-like capabilities. It has applications in various fields such as health, finance, transportation, and education.

Machine Learning is a data analysis method that automatically creates analytical models. This is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention.

The purpose of this work was to analyze the applicability of Machine Learning to classify customers based on total sales and sales volume in dollars represented in boxes in the tuna preservation industry. The analysis will be developed using classified trees and random forests, by collecting data from the company TECOPESCA C.A.

Keywords: Machine Learning, classify customers, total sales, random forests, classified trees, artificial intelligence

Résumé

L'ère numérique mise sur les nouvelles technologies, ce qui implique des changements et des transformations profondes de la société dans un monde globalisé. Cela se traduit par une véritable révolution technologique. L'intelligence artificielle est une partie similaire de cette époque, c'est la combinaison d'algorithmes proposés pour créer des machines avec des capacités de type humain. Il a des applications dans divers domaines tels que la santé, la finance, les transports et l'éducation.

Machine Learning est une méthode d'analyse de données qui crée automatiquement des modèles analytiques. Il s'agit d'une branche de l'intelligence artificielle basée sur l'idée que les systèmes peuvent apprendre des données, identifier des modèles et prendre des décisions avec une intervention humaine minimale.

Le but de ce travail était d'analyser l'applicabilité de l'apprentissage automatique pour classer les clients en fonction des ventes totales et du volume des ventes en dollars représentés dans des cases dans l'industrie de la conservation du thon. L'analyse sera développée à partir d'arbres classés et de forêts aléatoires, en collectant des données auprès de la société TECOPESCA C.A.

Mots clés : Machine Learning, forêts aléatoires, d'arbres classés, L'intelligence artificielle

Introducción

Introducción

La era digital forma parte de la sociedad hoy en día, donde la inteligencia artificial dejó de ser solo una idea y ha empezado ser una realidad, los trabajos que requerían mano de obra humana se están automatizando.

Cada día que pasa hay cambios constantes en la tecnología, todo para facilitar y dar una mejor vida humana. Machine Learning consta en estos avances digitales, siendo una disciplina científica creando sistemas que aprenden de cálculos previos para así dar decisiones y resultados (González, 2017).

El término de machine learning fue usado por primera vez en 1959, pero no llegó a su reconocimiento hasta hace poco, siendo más usada por el incremento de la tecnología y uso de ella (Perez, 2020).

Machine learning está sujeta a la tecnología que presenciamos más día a día, las diferentes empresas u organizaciones pueden lucrarse y tomar mejores decisiones sin necesidad de que haya alguna participación humana ya que permite a la computadora aprender por sí mismo y al mismo tiempo realizar tareas de forma autónoma sin ser programados (Asociación para el Progreso de la Dirección , 2019).

Las empresas que trabajan con grandes cantidades de datos han optado y han dado el reconocimiento respectivo sobre el valor que tiene machine learning, ya que gracias a este puede tomar decisiones más asertivas, sacando las estadísticas de los datos

que tenga la empresa para disminuir el riesgo. Esto logra una ventaja competitiva sobre otras empresas que todavía no lo usan (Perez, 2020).

Un ejemplo que se usa globalmente y está más cerca de nosotros de lo que pensamos es en las plataformas digitales, así como redes sociales, machine learning es usado para hacer recomendaciones personalizadas para el usuario según el comportamiento. Analiza el historial de compras, reproducciones del usuario, o intereses en general y lo compara con lo que hacen otros usuarios para crear recomendaciones personalizadas (León, 2019).

Por otro lado, podemos ver también machine learning en los servicios o comercio financiero que son usados de manera nacional. Las empresas financieras o bancos como tal la utilizan para identificar los insights importantes, una sensación o creencia profunda de las personas o consumidores (Terán, 2021).

Así logran prevenir algún intento de fraude. Gracias a estos insights podemos reconocer las oportunidades que estas entidades tienen al momento de ayudar inversionistas saber cuándo comprar y vender evitando pérdidas y generar ganancias (Arévalo, 2018).

Tomando en cuenta que este tipo de gestión puede depender netamente de las probabilidades, siendo un mercado todavía con una probabilidad relativamente baja, donde machine learning puede generar un amplio beneficio y a su vez también ganancias, siendo algo automático donde los humanos no pueden competir. El uso de machine learning se basa en estas entidades financieras en consumir grandes cantidades de datos a gran velocidad (Azabache, 2019).

De modo empresarial, machine learning es usado en estas grandes entidades donde manejan una base de datos bastante amplia, donde existen diferentes categorías ya se diferentes productos, proveedores, clientes, ubicaciones, entre otros. Esta información que tienen simula una tendencia así creando los insights, ayudando a futuras proyecciones. Con estas proyecciones pueden alcanzar respuestas certeras y así eliminando o previniendo el riesgo (León, 2019).

El uso de machine learning disminuye el riesgo que pueden haber usado los patrones de comportamiento, siendo una reducción de predicción de la información. El clásico machine learning usualmente se clasifica según la forma en que un algoritmo aprende a ser más preciso en sus predicciones. Hay cuatro enfoques básicos: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi-supervisado y aprendizaje reforzado. El tipo de algoritmo de datos que los científicos eligen utilizar depende del tipo de datos que quieran predecir (Arévalo, 2018)

Justificación

El uso de Machine Learning se ve justificada debido a que dentro de la empresa existe una amplia base de datos, donde la es de gran importancia y poder clasificar a los clientes de la mejor manera donde haya una probabilidad de error mínima. Así mismo lograr tomar una decisión, donde esto ayude a proyectar a la medida una planificación para lograr una mejor segmentación.

Existe el uso de datos estadísticos y matemáticos en conjunto de la tecnología y programas de computación, donde en otras palabras se podría conocer como inteligencia artificial (Fasce, 2020).

La justificación también va de la mano con esto debido a que mediante a esta inteligencia artificial logra minimizar o en tales casos eliminar los riesgos que pueda haber en los distintos escenarios que presenta la empresa, donde el producto será clasificando acorde al sector que mejor convenga, eliminando de toda manera el riesgo. En otras palabras, es que controla y ordena a todo lo que posee información dentro de la inteligencia artificial mediante empleo de machine learning.

En el Ecuador existen varias empresas que han entendido que el uso de una automatización inteligente es capaz de transformar el modelo de su negocio, trayendo múltiples beneficios y ahorros a la compañía, Este tipo de conceptos ha llevado a que empresas como PRONACA S.A. empiecen a implementar un modelo basado en la industria 4.0, en mayor o menor grado, en todas sus líneas de producción (Mcmillan, 2020).

Los modelos de Machine Learning son muy versátiles y útiles para el mundo empresarial. Su capacidad para adaptarse a los cambios y el aprendizaje automático, permite pronosticar variables futuras que han alcanzado aciertos superiores al 90%. Por ello, esta tecnología tiene un enorme potencial de cara al futuro y las grandes empresas ya están utilizando Machine Learning o preparan planes a corto plazo para su implementación dentro de la compañía (Coelho, 2020).

Con el análisis de datos es fácil comprobar cuándo es el mejor momento para incrementar la producción, cesar la producción, reducir los presupuestos, focalizarse en una tarea específica, etc. (Coelho, 2020).

El aspecto iterativo del machine learning es importante porque a medida que los modelos son expuestos a nuevos datos, éstos pueden adaptarse de forma independiente. Aprenden de cálculos previos para producir decisiones y resultados confiables y repetibles. Es una ciencia que no es nueva pero que ha cobrado un nuevo impulso (Fasce, 2020).

A partir de un error cometido, el aprendizaje automático registra las variables y en un futuro permite que no se repita. Además, puede implementar acciones preventivas. La solidez del sistema depende en gran medida del tiempo que lleve integrado al proceso (Vásquez & Malada, 2021).

Es un ámbito muy importante dentro de grandes empresas donde tienen una enorme cantidad de datos y ayude a una toma de decisión más acertada y a mayor velocidad, con machine learning se va creando patrones de comportamiento y así lograr esa respuesta asertiva.

Problemática

La problemática radica en que el mercado en el cual está compitiendo el producto es un mercado totalmente monopolizado por lo cual para implementar estrategias para competir en el mercado es necesario saber cuáles son sus clientes con más acogida y la zona de donde provienen, para esto la clasificación de clientes es necesaria para una implementación de un plan de acción el cual sea exitoso.

La clasificación de productos organiza los productos en cuatro categorías según el comportamiento de compra de los. La clasificación de clientes en estas categorías distintas ayuda a los especialistas a implementar estrategias más efectivas y a orientar

sus esfuerzos en función de las expectativas de los consumidores de esa clase de Clientes (Prieto, 2020).

Al realizar una clasificación de clientes tomamos en cuenta varios factores que estarán incluidos en nuestra base de datos estos serán de suma importancia para que así el algoritmo logre clasificar, cabe recalcar que en esta base de datos la información que esté incluida ahí debe ser de calidad.

En otro estudio uno de los principales problemas es la necesidad de personal experimentado en estos puestos. Asimismo, este trabajador puede desempeñar sus funciones de manera eficiente durante un período de tiempo, después del cual su productividad decae.

Los problemas identificados en la etapa de clasificación son el acondicionamiento de los arándanos en la línea de producción y la selección y clasificación de los arándanos para su envasado y posterior venta.

Los problemas específicos encontrados fueron los siguientes:

- Empleo de muchos trabajadores en el proceso de selección y clasificación de los arándanos, lo que origina un sobre costo de mano de obra
- Se requiere que los trabajadores tengan experiencia para reconocer y diferenciar los frutos de calidad y los frutos defectuosos para poder realizar la clasificación.

En otro ejemplo donde el Banco Alimentario de La Plata es una sociedad civil ubicado en Argentina presenta el mismo problema donde ellos van y consiguen

donaciones, deben ingresarlas al depósito para clasificarlas, y realizar la logística para enlazar la mercadería a los comedores, según con qué productos se cuenta y las necesidades que poseen las instituciones (Giraldez, 2019).

Teniendo como problemas específicos:

- Como trabajan los operarios está probada y mecanizada, pero hay puntos en los que se podría mejorar y agilizar el proceso.
- Los productos de las donaciones son registrados en una planilla a mano y luego son cargados al sistema, lo que hace que el registro se haga dos veces.
- La carga de las donaciones en el sistema es manual, por lo que está sujeta a fallos del operador.

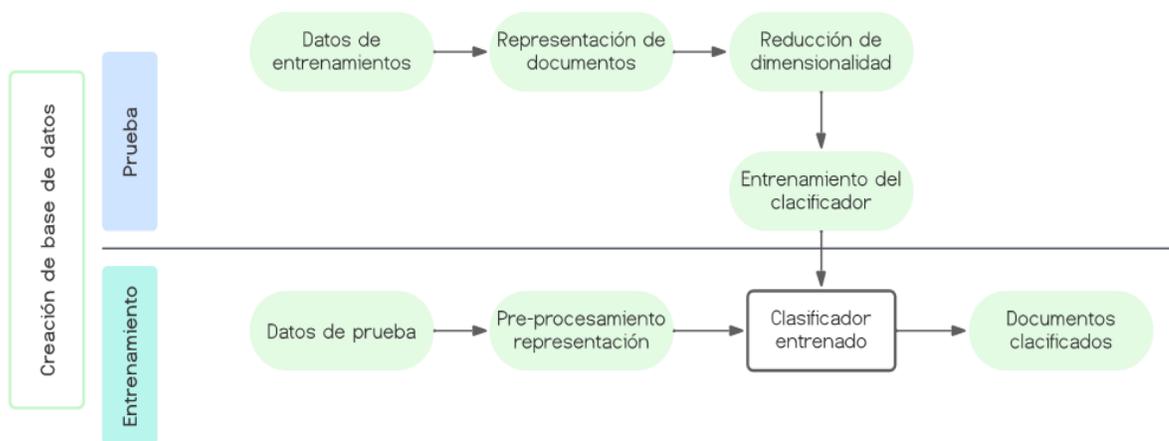
La implementación de modelos de Machine Learning a escala es uno de los retos más importantes para las empresas que desean crear valor a través de la Inteligencia Artificial, y a medida que los modelos se vuelven más complejos, cada vez es más difícil.

Solo un porcentaje muy pequeño de los proyectos de Machine Learning llegan a la producción. Un proyecto de Machine Learning puede fallar por muchas razones, entre ellas la implementación. Es fundamental que todos los responsables de la toma de decisiones comprendan plenamente cómo funciona el despliegue y cómo reducir el riesgo de fracaso al llegar a este paso crucial. Como ya sabemos un modelo desplegado puede definirse como cualquier unidad de código que se integra perfectamente en un entorno de producción y que puede recibir una entrada y devolver una salida, los problemas al implementar un modelo de Machine Learning surgen al momento de

finalizar las pruebas y cuando se inicia la implementación, es justamente en este paso en que aparecen algunos de los problemas más comunes (González, 2017).

Figura 1

Algoritmo de Machine Learning.



Alcance

La aplicación del machine learning con respecto a la clasificación permite cumplir con la adaptación y evaluación de clasificar la base de datos por Tipo de Cliente. La herramienta tendrá como objetivo clasificar una nueva instancia en relación con las variables establecidas que a largo plazo pueden causar grandes problemas a las compañías en un uso que deberían implementar.

El tipo de investigación es descriptiva e inferencial y va dirigido a aquellas empresas de Ecuador que no realizan la suficiente indagación con respecto al análisis de datos. Para aquellos estudiantes y profesores que quisieran realizar algún método mucho más preciso del que se usa actualmente. Los aportes conceptuales van desde un mejor

manejo de los datos adquiridos y también aportes a las exportaciones que las empresas realizan continuamente.

Objetivos

Objetivo General

- Determinar la viabilidad en la aplicación de machine learning en la clasificación por tipo de clientes.

Objetivo Específico

- Aplicar el concepto de machine learning sobre el objetivo de estudio.
- Clasificar los datos obtenidos a través de una recolección de Datos.
- Plantear y describir del análisis de los resultados.
- Determinar la toma de decisión conforme al tipo de cliente y verificar, la efectividad.

Hipótesis

- La aplicación de ML clasifica de manera correcta a los tipos de clientes tomando como variables el volumen de venta en dólares y las ventas expresadas en las cajas.

Operacionalización de las Variables

Tabla 1

Operacionalización de las variables a utilizarse en el modelo.

Variables	
Zona	En esta variable se encuentran todos los clientes clasificados por zonas tomando como Zona 0 a Manta, a la zona 1 Imbabura -Carchi-Ibarra-Tulcán, Zona 2 Guayaquil-Santa Elena, Zona 3 Pichincha, Zona 4 Manabí-Portoviejo y Zona 5 austro
Volumen de ventas en dólares	En esta variable se encuentra un total de ventas en dólares entre los años 2018, 2019 y 2020 de cada cliente.
Ventas en Cajas	En esta variable se expresa el total de ventas de cada cliente en cajas entre los años 2018, 2019 y 2020.

Capítulo II: Marco Teórico

Marco Teórico

Inteligencia Empresarial

Inteligencia Empresarial, o BI, por sus siglas en inglés refiere al análisis de los datos recopilados y almacenados en una organización. La capacidad de combinar estas dos fuentes de forma compleja, combinada con las diversas herramientas que permiten el análisis de estos datos. Mediante el uso de herramientas como el Explorador de datos, los datos analíticos se pueden realizar con mayor precisión, lo que permite a las

organizaciones tomar mejores decisiones. Esto, a su vez, es beneficioso para las organizaciones, ya que les permite tomar decisiones.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc. (Ilino, 2007).

Además, mejora la toma de decisiones en la organización. El software centrado en BI le permite obtener una ventaja competitiva sobre sus competidores en el mercado al encontrar información oculta en datos que no se pueden vincular, visualizar y analizar por completo.

Las Soluciones de BI con sus principales características generales es que se encuentran preparado para que Unicode y a su vez que sean compatibles con los datos que se encuentran acumulados en varios idiomas. En el documento de Especificaciones Técnicas de cada herramienta tiene una lista que accede a la verificación de los tipos de fuentes de datos con los que se puede asociar siendo compatibles. Las bases de datos en ciertas ocasiones con los datos de proveedores pequeños no tienen esas fuentes para que sean compatibles debido a esto, estos pueden moverse los datos a SQL Server o a Excel, siendo esto una herramienta que se encuentran en todas las listas de las especificaciones técnicas para poder cargar los datos a la aplicación BI. La tecnología de Inteligencia

Empresarial incluye conjuntos de consultas, informes, paneles y métodos OLAP predefinidos y personalizados. También se puede utilizar para analizar e informar.

Big Data

Esta disciplina implica de la gestión de grandes cantidades de datos que recopila de diversas fuentes. La mayoría de las veces, los datos provienen de sensores y otros dispositivos. La diferencia que existe con los datos recopilados para las bases tradicionales, en los datos de BD puede tener un porcentaje de error debido a que los sensores con los que se proceden llegan a tener fallas, siendo las mismas que pueden ser pasadas por alto.

No obstante, el cambio de mentalidad ocasiona el uso de datos masivos es que, aunque el nivel de errores puede ser alto, el resto de la información puede ayudar a ratificar la veracidad del dato. En el libro de Viktor Mayer-Schonberger encontramos esta frase: “Usar todos los datos nos permite ver detalles que nosotros nunca hubiésemos visto con una cantidad limitada o pequeña de datos. Big Data nos da una especial vista clara de la granular: subcategorías y submercados de muestras que no se pueden evaluar” (Schonberger, 2014).

Cloud Computing

Se puede entender Cloud Computing como servicios que mediante servicios a través de la conectividad y gran escala de Internet. La informática en la nube democratiza el paso a medios de software del grupo mundial, ya que es una aplicación de software que sirve a varios clientes. La zona múltiple es lo que contrasta la computación en la nube de la subcontratación única y los modelos de proveedores de servicios de aplicaciones crecidamente antiguos.

Actualmente, las pequeñas empresas tienen el espacio de rendir el poder de la tecnología progresada de modo escalable. Los servicios en la Nube se han clasificado en tres grupos que incluyen: Infraestructura (IaaS), Software (SaaS) y Plataforma (PaaS), los mismos que, como se mencionó anteriormente, se ofrecen a través de una red. Como dijo (Grobauer & Stöcker, 2011).

Vulnerabilidades y Riesgos de Trabajar en la Nube

Ataques a las vulnerabilidades ocasionan incorrectamente trabajo del método, la abusó de usuarios no autorizados, la carencia de la integridad de los datos, entre otros. Los Riesgos y Vulnerabilidades son unas de las principales desventajas del trabajo en la Nube, a pesar de que los proveedores de servicios en la Nube cada vez ofrecen productos más seguros, el cliente siempre está expuesto al hecho de que el producto tenga una vulnerabilidad que no haya sido detectada por el fabricante y pueda ser víctima de un ataque (Grobauer, Walloschek, & Stocker, 2011).

Business Intelligence en la Nube

Los servicios en la Nube han formado que las soluciones para BI se pongan la magnitud de empresas pequeñas y medianas que no gozaban de la oportunidad de los servicios de la tecnología por sus altos costos proporción en deposición de Infraestructura (hardware, software, seguridades físicas y lógicas) como de la puesta a punto.

Permite ofrecer servicios de computación a través de Internet. Estos servicios están orientados a que se paguen por uso, SaaS (software as a service), siendo un nuevo modelo de negocio para las empresas de tecnología. La ventaja para las empresas de unir

los dos conceptos, la inteligencia de negocio y la computación en la nube, es que ya no necesitan una infraestructura de servidores de aplicaciones, comunicaciones, etc., no tiene que realizar fuertes desembolsos pensando que lo van a recuperar a futuro, sino que solo van a pagar por lo que necesitan usar en cada momento (López, Ante, & Portie, 2021).

Sistema de Soporte de Decisiones

Esta es una herramienta de inteligencia empresarial que se enfoca en el análisis de datos empresariales, estas herramientas suelen tener una interfaz amigable que permite mostrar los resultados a través de un gráfico para facilitar el análisis de los resultados.

Una decisión es el resultado de un proceso por el cual se elige entre distintas opciones, es decir, de un sistema de apoyo a las decisiones, conocido también como DSS por sus siglas en inglés, Decisión Support System. Este es un sistema informático o software que sustenta el proceso de toma de decisiones. De esta manera, este proceso induce a que los comprometidos en el proyecto, se reúnan para diagramar y crear opciones para tomar una decisión de importancia (Simon, 2021).

Data Marts

Los Data Marts son almacenes de datos especializados, permiten clasificar los datos y así facilitar el procesamiento de datos para su posterior estudio específico en el almacén de datos. La información de Data Marts puede provenir de un almacén de datos o de una base de datos relacional. El uso de Data Marts puede ayudar a que el almacén de datos funcione más fácilmente, especialmente para un almacén de datos muy grande,

donde el peso de los datos conduce a una disminución en el rendimiento de las consultas. Su implementación es fundamental en un negocio, pues de esta manera la empresa puede integrar y combinar la información en un Data Mart para descubrir oportunidades y amenazas indetectables en las bases de datos operacionales.

Figura 2

Proceso de un Data Mart.



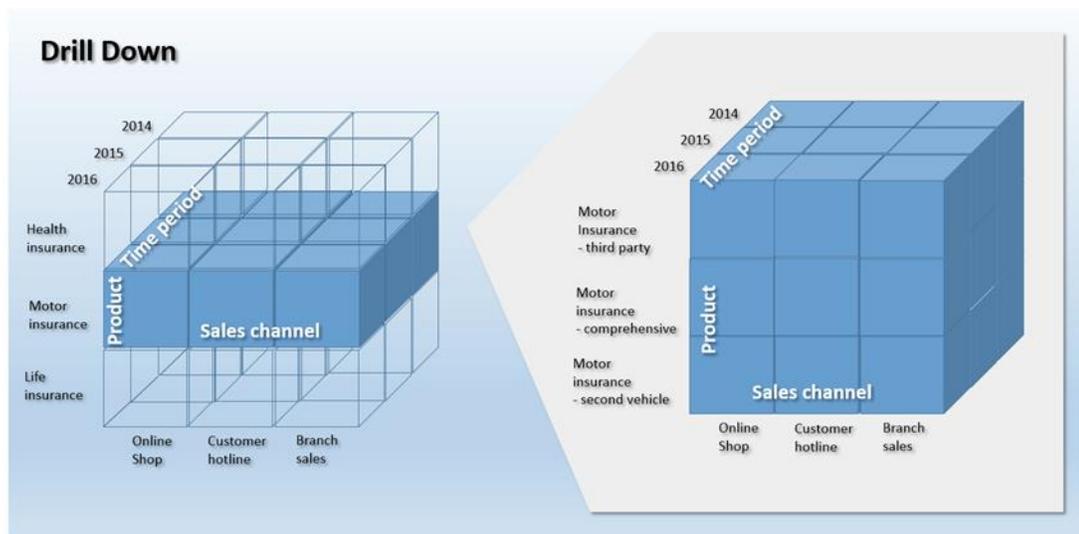
Data Warehouse

Es un repositorio donde podemos recopilar datos de diferentes fuentes. Estos datos se transforman para crear metadatos, los cuales se ubican en el archivo y permiten ser combinados según las necesidades de las organizaciones para obtener información que permita la mejora del sistema. El almacén de datos es responsable de la limpieza e integridad de los datos, lo que facilita la extracción de datos posterior. Los datos históricos permiten detectar tendencias, el análisis de estos datos ayuda a conocer al negocio a mayor profundidad y tomar decisiones más acertadas. Data Warehouse es una

arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas. Un data warehouse es una arquitectura conocida ya en muchas empresas modernas (Banda, 2017).

Figura 3

Proceso de recopilación de data.



Data Mining

Se usa el término Minería de Datos o Data Mining, para describir los procesos que se encargan de encontrar información que, a primera vista, en un conjunto de datos no es fácil de observar. Estos datos se cargan y analizan por medio de herramientas que usan técnicas que permiten encontrar modelos. Su objetivo principal es descubrir automáticamente enormes bases de datos mediante el uso de diferentes técnicas y tecnologías. El objetivo es encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos que se han ido recopilando con el tiempo (Bello, 2021).

Cuadrante Mágico de Gartner para Herramientas BI

El Cuadrante Mágico de Gartner es una herramienta desarrollada por la empresa de investigación y consultoría Gartner Inc. Se ocupan de analizar el mercado de las nuevas tecnologías. Los cuadrantes clasifican a las aplicaciones en: Líderes (Leaders), Aspirantes (Challengers), Visionarios (Visionaries) y Nichos Específicos (Niche Players). Esta herramienta clasifica las soluciones informáticas de las principales compañías en el mercado, presentando los resultados de una manera gráfica, según el resultado de una investigación que permite analizar a cada una de ellas basado en los mismos criterios (Ali, Cemil, Merve, & Abdullah, 2021).

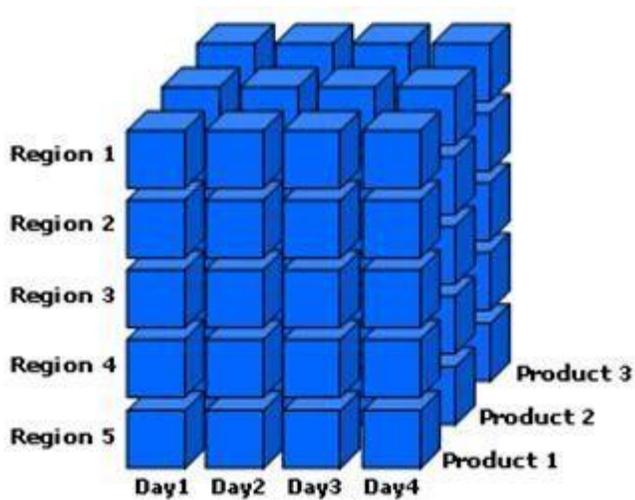
Cubo OLAP

La tecnología OLAP permite un uso más eficiente de los almacenes de datos para el análisis en línea al proporcionar respuestas rápidas a consultas analíticas complejas y repetitivas. Técnicas de agregación de datos y modelado de datos OLAP multidimensionales que organizan y resumen grandes cantidades de datos para que puedan evaluarse rápidamente utilizando herramientas gráficas y analíticas en línea.

OLAP proporciona varias ventajas a analistas del negocio como, por ejemplo, un modelo de datos intuitivo y multidimensional que facilita la selección, recorrido y exploración de los datos. Un lenguaje analítico de consulta que proporciona la capacidad de explorar las complejas relaciones existentes entre los datos empresariales (Kohl, 2018).

Figura 4

Funcionamiento del cubo OLAP.



Machine Learning

Machine learning es una forma de la IA que permite a un sistema aprender de los datos en lugar de aprender mediante la programación explícita. Sin embargo, machine learning no es un proceso sencillo. Conforme el algoritmo ingiere datos de entrenamiento, es posible producir modelos más precisos basados en datos. Un modelo de machine learning es la salida de información que se genera cuando se entrena su algoritmo de machine learning con datos. Después del entrenamiento, al proporcionar un modelo con una entrada, se le dará una salida. Por ejemplo, un algoritmo predictivo creará un modelo predictivo. A continuación, cuando proporcione el modelo predictivo con datos, recibirá un pronóstico basado en los datos que entrenaron al modelo (International Business Machine Corporation, 2017).

Deep Learning

El deep learning es un tipo de machine learning que entrena a una computadora para que realice tareas como las hacemos los seres humanos, como el reconocimiento

del habla, la identificación de imágenes o hacer predicciones. En lugar de organizar datos para que se ejecuten a través de ecuaciones predefinidas, el deep learning configura parámetros básicos acerca de los datos y entrena a la computadora para que aprenda por cuenta propia reconociendo patrones mediante el uso de muchas capas de procesamiento (Pushpika, Oosterhout, Pursey, Duran, & Essen, 2019).

En el enfoque *Deep Learning* se usan estructuras lógicas que se asemejan en mayor medida a la organización del sistema nervioso de los mamíferos, teniendo capas de unidades de proceso (neuronas artificiales) que se especializan en detectar determinadas características existentes en los objetos percibidos. La visión artificial es una de las áreas donde el Deep Learning proporciona una mejora considerable en comparación con algoritmos más tradicionales (Mao, He, Morrison, & Coca, 2020).

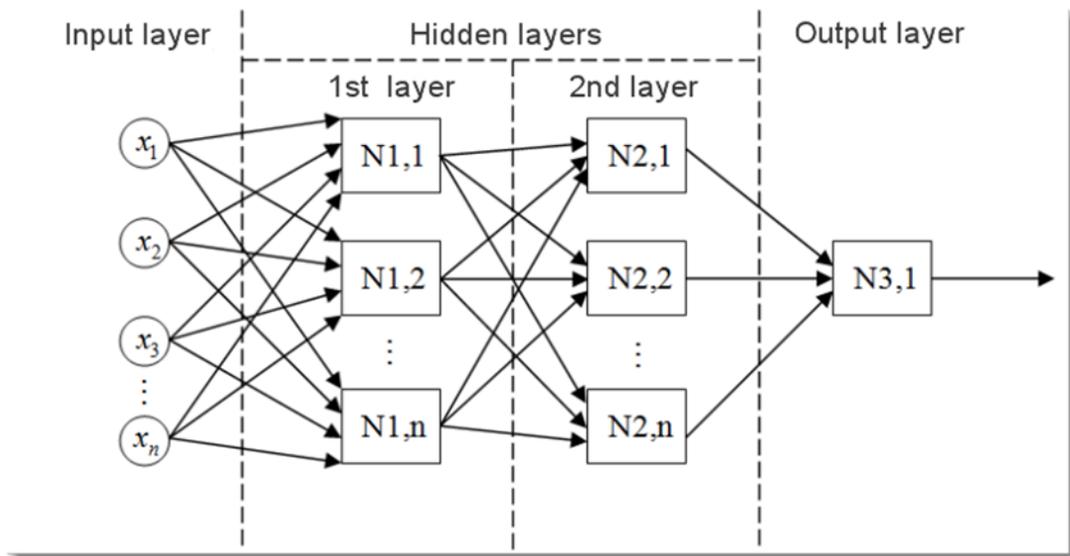
Redes Neuronales

Las redes neuronales artificiales son un modelo inspirado en el funcionamiento del cerebro humano. Está formado por un conjunto de nodos conocidos como neuronas artificiales que están conectadas y transmiten señales entre sí. Estas señales se transmiten desde la entrada hasta generar una salida. El objetivo principal de este modelo es aprender modificándose automáticamente a si mismo de forma que puede llegar a realizar tareas complejas que no podrían ser realizadas mediante la clásica programación basada en reglas. De esta forma se pueden automatizar funciones que en un principio solo podrían ser realizadas por personas (Lean, 2019).

Como se ha mencionado el funcionamiento de las redes se asemeja al del cerebro humano. Las redes reciben una serie de valores de entrada y cada una de estas entradas llega a un nodo llamado neurona. Las neuronas de la red están a su vez agrupadas en capas que forman la red neuronal. Cada una de las neuronas de la red posee a su vez un peso, un valor numérico, con el que modifica la entrada recibida (Lean, 2019).

Figura 5

Red neuronal con capa oculta.



Python

Es un lenguaje de programación versátil multiplataforma y multiparadigma que se destaca por su código legible y limpio. Una de las razones de su éxito es que cuenta con una licencia de código abierto que permite su utilización en cualquier escenario. Esto hace que sea uno de los lenguajes de iniciación de muchos programadores siendo impartido en escuelas y universidades de todo el mundo. Sumado a esto cuenta con grandes compañías que hacen de este un uso intensivo (Robledano, 2019).

Python es ideal para trabajar con grandes volúmenes de datos ya que, el ser multiplataforma, favorece su extracción y procesamiento, por eso lo eligen las empresas de Big Data. A nivel científico, tiene una gran biblioteca de recursos con especial énfasis en las matemáticas para aspirantes a programadores en áreas especializadas. Y si esto fuera poco, es posible crear videojuegos, aunque no es tan eficiente como Java o C# (Robledano, 2019).

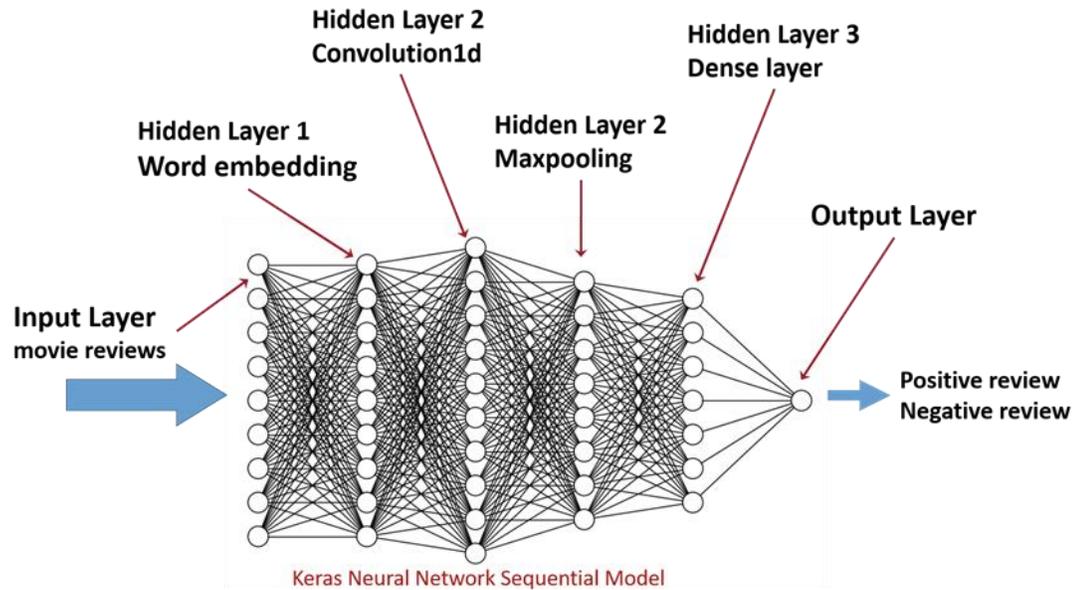
Keras

Keras es un framework de alto nivel para el aprendizaje, escrito en Python y capaz de correr sobre los frameworks TensorFlow, CNTK, o Theano. Fue desarrollado con el objeto de facilitar un proceso de experimentación rápida. Lo que haremos en este experimento es entrenar modelos de clasificación de imágenes (Burgal, 2018).

Proporciona los componentes básicos para desarrollar modelos complejos de aprendizaje profundo. A diferencia de los marcos independientes, este software de código abierto no se ocupa de operaciones simples de bajo nivel. En su lugar, utiliza las bibliotecas de los marcos de aprendizaje automático asociados para este propósito. Actúan como una especie de motor de backend para Keras. Dado que la idea es ser modular, las capas deseadas para la red neuronal que se está desarrollando se conectan entre sí sin que el usuario de Keras deba comprender o controlar el backend real del marco seleccionado (Barnett, Irene, & Husted, 2020).

Figura 6

Red Neuronal Keras.

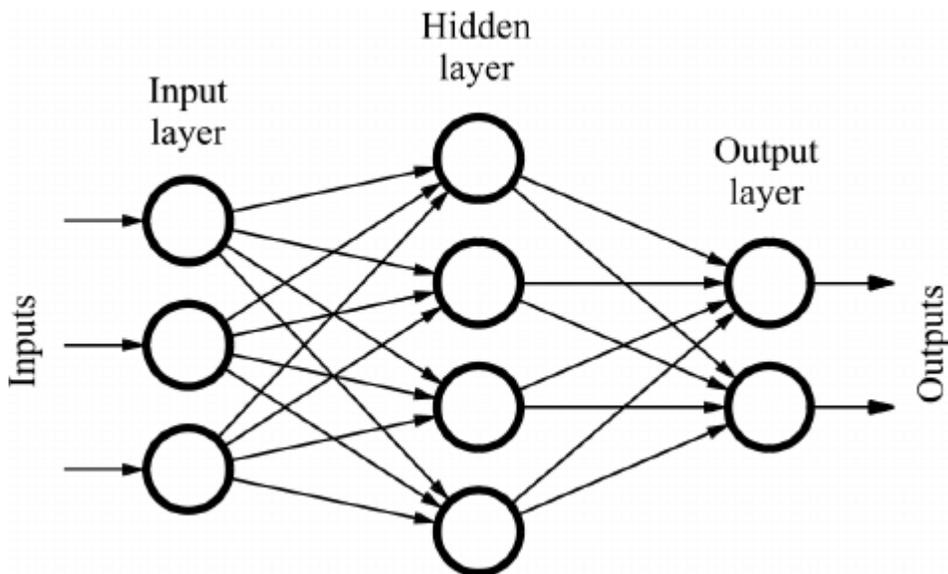


Redes Feedforward

Una red neuronal prealimentada es una red neuronal artificial donde las conexiones entre las unidades no forman un ciclo. Estas son diferentes de las redes neuronales recurrentes. La red neuronal prealimentada fue la primera y más sencilla forma de red neuronal artificial ideada. En esta red, la información se mueve en una única dirección: adelante. De los nodos de entrada, a través de los nodos escondidos hacia los nodos de salida. No hay ningún ciclo o bucle en estas redes (Asociación para el Progreso de la Dirección , 2019).

Figura 7

Red neuronal.



Redes Convolucionales

Las convolutional neural networks o redes neuronales convolucionales (CNN), solucionan este problema ya que asumen ciertas características espaciales de los inputs que permiten simplificar las arquitecturas de la red reduciendo, en gran medida, el número de variables de entrada (Torralba, 2021).

Por tanto, son especialmente útiles en problemas de visión artificial, especialmente en el reconocimiento de objetos. Las redes CNN utilizan distintos tipos de capas o capas. La capa más importante y la que da nombre a la web es la capa de convolución. Esta capa funciona con pequeños filtros 3D que se mueven alrededor de la imagen para obtener el resultado de esta capa.

Marco Conceptual

Algoritmos de Machine Learning

Conceptos como inteligencia artificial (IA) o machine learning (aprendizaje automático) son ineludibles en el contexto actual. El aprendizaje automático es esa rama de la informática que otorga a la IA la capacidad de aprender tareas. Para lograrlo, los programadores se basan en los algoritmos del machine learning.

El término aprendizaje automático se confunde a menudo con el de Inteligencia Artificial, cuando en realidad es un subcampo. Se define como la capacidad del ordenador para aprender sin ser programado explícitamente (Asociación para el Progreso de la Dirección , 2019).

Grupos de Algoritmos de Machine Learning

Aprendizaje Supervisado. Según la Asociación para el Progreso de la Dirección (2019), en el aprendizaje supervisado, la máquina se enseña con el ejemplo. De este modo, el operador proporciona al algoritmo de aprendizaje automático un conjunto de datos conocidos que incluye las entradas y salidas deseadas, y el algoritmo debe encontrar un método para determinar cómo llegar a esas entradas y salidas.

Mientras el operador conoce las respuestas correctas al problema, el algoritmo identifica patrones en los datos, aprende de las observaciones y hace predicciones. El algoritmo realiza predicciones y es corregido por el operador, y este proceso sigue hasta que el algoritmo alcanza un alto nivel de precisión y rendimiento.

Aprendizaje sin Supervisión. Aquí, el algoritmo de aprendizaje automático estudia los datos para identificar patrones. No hay una clave de respuesta o un operador humano para proporcionar instrucción. En cambio, la máquina determina las correlaciones y las relaciones mediante el análisis de los datos disponibles.

En un proceso de aprendizaje no supervisado, se deja que el algoritmo de aprendizaje automático interprete grandes conjuntos de datos y dirija esos datos en consecuencia. Así, el algoritmo intenta organizar esos datos de alguna manera para describir su estructura. Esto podría significar la necesidad de agrupar los datos en grupos u organizarlos de manera que se vean más organizados.

A medida que evalúa más datos, su capacidad para tomar decisiones sobre los mismos mejora gradualmente y se vuelve más refinada.

Aprendizaje por Refuerzo. Según la Asociación para el Progreso de la Dirección (2019) el aprendizaje por refuerzo se centra en los procesos de aprendizajes reglamentados, en los que se proporcionan algoritmos de aprendizaje automáticos con un conjunto de acciones, parámetros y valores finales.

Al definir las reglas, el algoritmo de aprendizaje automático intenta explorar diferentes opciones y posibilidades, monitorizando y evaluando cada resultado para determinar cuál es el óptimo.

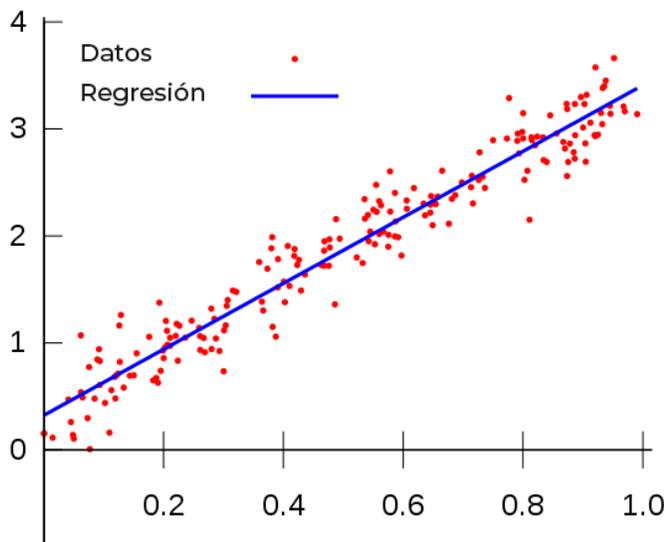
En consecuencia, este sistema enseña la máquina a través del proceso de ensayo y error. Aprende de experiencias pasadas y comienza a adaptar su enfoque en respuesta a la situación para lograr el mejor resultado posible.

Tipos de Algoritmos del Machine Learning

Algoritmos de Regresión. Según la Asociación para el Progreso de la Dirección (2019), en las tareas de regresión, el programa de aprendizaje automático debe estimar y comprender las relaciones entre las variables. El análisis de regresión se enfoca en una variable dependiente y una serie de otras variables cambiantes, lo que lo hace particularmente útil para la predicción y el pronóstico.

Figura 8

Gráfico de regresión.



Algoritmos Bayesianos. Según la Asociación para el Progreso de la Dirección (2019), este tipo de algoritmos por clasificación están basados en el teorema de Bayes y clasifican cada valor como independiente de cualquier otro. Lo que permite predecir una clase o categoría en función de un conjunto dado de características, utilizando la probabilidad.

A pesar de su simplicidad, el clasificador funciona sorprendentemente bien y se usa a menudo porque supera a los métodos de clasificación más sofisticados.

Algoritmos de Agrupación. Según la Asociación para el Progreso de la Dirección (2019), se utilizan en el aprendizaje no supervisado, y sirven para categorizar datos no etiquetados, es decir, datos sin categorías o grupos definidos.

El algoritmo funciona mediante la búsqueda de grupos dentro de los datos, con el número de grupos representados por la variable K . A continuación, funciona de manera iterativa para asignar cada punto de datos a uno de los K grupos según las características proporcionadas.

Algoritmos de Árbol de Decisión. Según la Asociación para el Progreso de la Dirección (2019), un árbol de decisión es una estructura de árbol similar a un diagrama de flujo que utiliza un método de bifurcación para ilustrar cada resultado posible de una decisión. Cada nodo dentro del árbol representa una prueba en una variable específica, y cada rama es el resultado de esa prueba.

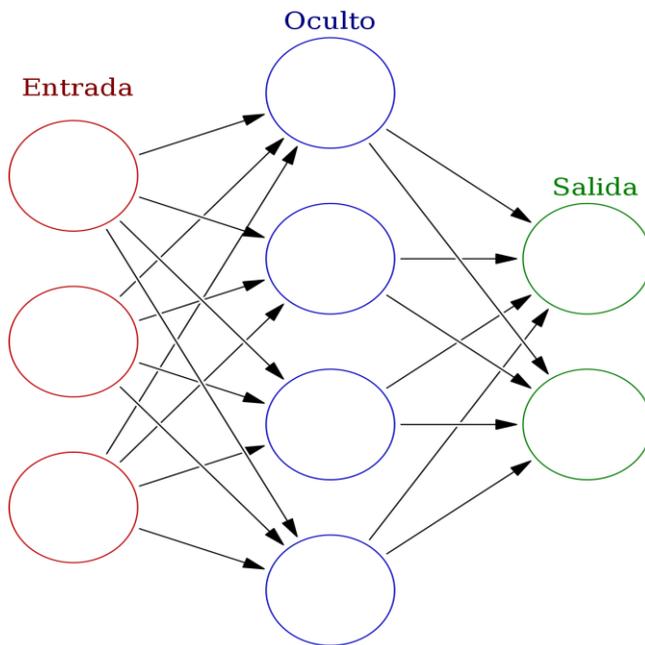
Algoritmos de Redes Neuronales. Según la Asociación para el Progreso de la Dirección (2019), una red neuronal artificial (RNA) comprende unidades dispuestas en una serie de capas, cada una de las cuales se conecta a las capas anexas. Las RNA se inspiran en los sistemas biológicos, como el cerebro, y en cómo procesan la información.

Por lo tanto, son esencialmente un gran número de elementos de procesamiento interconectados, que trabajan al unísono para resolver problemas específicos.

También aprenden con el ejemplo y la experiencia, y son extremadamente útiles para modelar relaciones no lineales en datos de alta dimensión, o donde la relación entre las variables de entrada es difícil de entender.

Figura 9

Red neuronal simple.



Algoritmos de Reducción de Dimensión. Según la Asociación para el Progreso de la Dirección (2020), la reducción de dimensión reduce el número de variables que se consideran para encontrar la información exacta requerida.

Algoritmos de Aprendizaje Profundo. Según la Asociación para el Progreso de la Dirección (2019), los algoritmos de aprendizaje profundo ejecutan datos a través de varias capas de algoritmos de redes neuronales, las cuales pasan a una representación simplificada de los datos a la siguiente capa.

La mayoría funciona bien en conjuntos de datos que tienen hasta unos cientos de características o columnas. Sin embargo, un conjunto de datos no estructurado, como el de una imagen, tiene una cantidad tan grande de características que este proceso se vuelve engorroso o completamente inviable.

Marco Legal

En el ámbito legal se han generado varios planes alineados a una normativa legal entre los cuales se destacan: 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, Art. 313) que permiten alinearse y ser partícipes de la adopción de estas tecnologías emergentes para el cambio y adopción de una cultura de transformación digital para un mejor desarrollo a nivel empresarial en el ámbito local con proyección internacional, lo que se convierte en un impacto positivo.

Capítulo III: Metodología

Metodología

En la ciudad de Manta, Ecuador Tecopesca C.A, Es una empresa que recoge la ancestral tradición pesquera de las costas ecuatorianas y la combina con tecnología de punta elaborando así productos en conserva de la más alta calidad, Tecopesca llega con sus productos a través de marcas propias o distribuidores a los mercados nacional e

internacional. Tecopesca es parte del Grupo Corporativo Visión. El número total de ventas acumulado tomando en cuenta los últimos 3 años registrados fue de \$8,229,698.70 y el número total de ventas representado en cajas es de 154806.27.

El registro se obtuvo de una base de datos proporcionado por la empresa Tecopesca C.A del año 2018 al año 2020, Los datos obtenidos son tipo

El trabajo es de corte transaccional debido a que se recolectaron datos de ventas a través del tiempo donde se realizó un análisis con respecto a las variables que se interrelacionan, la técnica para recolección de datos que se empleo es proporcionada por la empresa Tecopesca C.A que nos ha proporcionado las ventas en dólares de los años 2018-2019-2020, y el total de ventas en cajas de los mismos años

El objetivo principal del estudio radica en la importancia del uso de Machine Learning mediante un árbol de clasificación y luego un modelo de árbol aleatorio para obtener una clasificación más exacta. La investigación es de corte correlacional de tipo exploratorio ya que busca clasificar a los clientes a través de las variables.

Pasos a Realizar

La aplicación de El árbol de clasificación y Random forest en Machine Learning es utilizada para la clasificación de las distintas variables que se desee tomar, el machine learning nos será útil para la clasificación de las variables mediante los datos de entrenamiento y los datos de prueba ambas provienen de la misma base de datos, y validar la clasificación de el árbol de decisión con los árboles aleatorios los cuales me

van a corregir la clasificación de los clientes creada por el árbol de decisión y con los resultados confirmar si el modelo es adecuado.

Una vez que ya se realizó una explicación breve de los árboles de clasificación y de los bosques aleatorios se procederá a mencionar los pasos a seguir en el programa R studio.

Primero se procederá a cargar la base de datos que se estableció a partir de la base que proporciono la empresa denominado en la base de datos como Clientes y se procede a cargar la base de datos con la secuencia ***Clientes = read.csv("../DATA/clientestesis.csv", header = TRUE, sep = ",")***, a continuación en la base de datos Clientes se procede a establecer nuestra variable que se va a clasificar como factor mediante el algoritmo ***Clientes\$TipoDeCliente = factor(Clientes\$TipoDeCliente)***, de esta forma nuestra variable será tomada como factor para luego poder realizar la clasificación, después se procede a cargar la librería caret que de acuerdo con Berrendero (2017), permite utilizar un código unificado para aplicar reglas de clasificación muy distintas, implementadas en diferentes paquetes.

```
1 Clientes = read.csv("../DATA/clientestesis.csv",header = TRUE, sep = ",")
2 Clientes$TipoDeCliente = factor(Clientes$TipoDeCliente)
3 library(caret)
```

Una vez cargados los datos y dado a que este modelo de árbol de clasificación esta derivado del Random Forest procedemos a plantar una semilla con la secuencia ***set.seed(123)***, es uso principal de la semilla es reproducir secuencias de números aleatorios a estos números se los denomina pseudoaleatorios esto quiere decir que si se

conoce la semilla se puede predecir la secuencia, la semilla que se utilizó para este estudio fue la 123.

```
5 set.seed(123)
```

Luego de que la semilla fue plantada se procede a realizar la partición de datos para dividirlos en una parte que será tomada para el entrenamiento y otra será tomada de prueba, el investigador puede decidir el porcentaje que pueda usar para la partición de datos en la cual el porcentaje mínimo que puede tomar para el entrenamiento es del 50% dependiendo de la cantidad de observaciones que tenga sin embargo un rango confiable para la clasificación de datos es de 70% para el entrenamiento y 30% para la prueba dependiendo de la cantidad de datos que se tenga, en nuestro caso se procedió a tomar un porcentaje de entrenamiento el cual será del 70%, se procedió a darle el nombre de *identificador.En* a nuestro entrenamiento, además se señaló nuestra variable clasificadora la cual es *TipoDeCliente*(El tipo de cliente), esto será expresado en el siguiente algoritmo: *identificador.En = createDataPartition(Clientes\$TipoDeCliente, p=0.7, list = F)*.

```
7 identificador.En = createDataPartition(Clientes$TipoDeCliente, p=0.7, list = F)
8 identificador.En
```

Una vez que se haya realizado la partición de datos se procede a llamar a la librería *rpart* la cual realiza arboles de decisión, pero su función principal es a partir de una base de datos o un conjunto crea un árbol de decisión que se puede usar para pronosticar.

```
10 library(rpart)
```

Después de cargar la librería se procede a crear el modelo de árbol de clasificación con los datos de entrenamiento con el siguiente algoritmo **Modelo = rpart(TipoDeCliente ~.,data = Clientes[identificador.En,],method = "class",)**, donde se toma la variable clasificadora la cual es **TipoDeClientes** para su clasificación, es decir que se clasifica a los tipos de clientes en función de las demás variables, y se añade a la fórmula el método el cual es **method = "class"** esto significa que la metodología que se usará es de clasificación.

```
12 Modelo = rpart(TipoDeCliente ~.,data = Clientes[identificador.En,],method = "class",)
13 Modelo
```

Procedemos a correr el modelo para verificar que el árbol exista en la base de datos, para luego ser graficado con el comando **Modelo**.

```
> Modelo
n= 508

node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 508 132 1 (0.74015748 0.16535433 0.06692913 0.02755906)
 2) TotalCajas< 92 377 21 1 (0.94429708 0.03713528 0.01856764 0.00000000)
   4) TotalCajas>=14.5 292 8 1 (0.97260274 0.01712329 0.01027397 0.00000000) *
   5) TotalCajas< 14.5 85 13 1 (0.84705882 0.10588235 0.04705882 0.00000000)
     10) TotalVentas< 711.5 78 7 1 (0.91025641 0.06410256 0.02564103 0.00000000) *
     11) TotalVentas>=711.5 7 3 2 (0.14285714 0.57142857 0.28571429 0.00000000) *
 3) TotalCajas>=92 131 61 2 (0.15267176 0.53435115 0.20610687 0.10687023)
   6) TotalVentas< 17286.61 88 21 2 (0.17045455 0.76136364 0.06818182 0.00000000)
     12) TotalVentas< 10333.36 47 16 2 (0.27659574 0.65957447 0.06382979 0.00000000)
       24) TotalVentas>=5793.21 23 11 2 (0.47826087 0.52173913 0.00000000 0.00000000)
         48) TotalCajas< 209.5 15 5 1 (0.66666667 0.33333333 0.00000000 0.00000000) *
         49) TotalCajas>=209.5 8 1 2 (0.12500000 0.87500000 0.00000000 0.00000000) *
       25) TotalVentas< 5793.21 24 5 2 (0.08333333 0.79166667 0.12500000 0.00000000) *
     13) TotalVentas>=10333.36 41 5 2 (0.04878049 0.87804878 0.07317073 0.00000000) *
 7) TotalVentas>=17286.61 43 22 3 (0.11627907 0.06976744 0.48837209 0.32558140)
   14) TotalVentas< 98096.86 30 9 3 (0.16666667 0.10000000 0.70000000 0.03333333)
     28) TotalCajas>=402 22 9 3 (0.22727273 0.13636364 0.59090909 0.04545455)
       56) TotalCajas< 614.5 9 4 1 (0.55555556 0.22222222 0.22222222 0.00000000) *
       57) TotalCajas>=614.5 13 2 3 (0.00000000 0.07692308 0.84615385 0.07692308) *
     29) TotalCajas< 402 8 0 3 (0.00000000 0.00000000 1.00000000 0.00000000) *
   15) TotalVentas>=98096.86 13 0 4 (0.00000000 0.00000000 0.00000000 1.00000000) *
```

Una vez verificado el modelo utilizamos la función **predict** que con nuestros datos de entrenamientos **identificador.En** generar un vector con los valores predichos por el modelo que hemos entrenado y usamos la siguiente función **Predicción =**

predict(Modelo, Clientes[-identificador.En,]), una vez que tengamos la predicción procedemos a usar *Predicción* para observar los datos que nos dio la predicción.

```
15 Predicción = predict(Modelo,Clientes[-identificador.En,])
16 Predicción|
```

Se procede a revisar el pronóstico para verificar que este en la base de datos:

Tabla 2

Predicción del modelo.

```
> Predicción
   1          2          3          4
1  0.91025641 0.06410256 0.02564103 0.00000000
2  0.91025641 0.06410256 0.02564103 0.00000000
6  0.91025641 0.06410256 0.02564103 0.00000000
9  0.91025641 0.06410256 0.02564103 0.00000000
15 0.91025641 0.06410256 0.02564103 0.00000000
17 0.91025641 0.06410256 0.02564103 0.00000000
18 0.91025641 0.06410256 0.02564103 0.00000000
19 0.91025641 0.06410256 0.02564103 0.00000000
28 0.91025641 0.06410256 0.02564103 0.00000000
29 0.91025641 0.06410256 0.02564103 0.00000000
32 0.91025641 0.06410256 0.02564103 0.00000000
37 0.91025641 0.06410256 0.02564103 0.00000000
38 0.91025641 0.06410256 0.02564103 0.00000000
43 0.91025641 0.06410256 0.02564103 0.00000000
44 0.91025641 0.06410256 0.02564103 0.00000000
46 0.91025641 0.06410256 0.02564103 0.00000000
47 0.91025641 0.06410256 0.02564103 0.00000000
49 0.91025641 0.06410256 0.02564103 0.00000000
56 0.91025641 0.06410256 0.02564103 0.00000000
58 0.91025641 0.06410256 0.02564103 0.00000000
60 0.91025641 0.06410256 0.02564103 0.00000000
62 0.91025641 0.06410256 0.02564103 0.00000000
65 0.91025641 0.06410256 0.02564103 0.00000000
68 0.91025641 0.06410256 0.02564103 0.00000000
70 0.91025641 0.06410256 0.02564103 0.00000000
71 0.91025641 0.06410256 0.02564103 0.00000000
78 0.91025641 0.06410256 0.02564103 0.00000000
79 0.14285714 0.57142857 0.28571429 0.00000000
87 0.91025641 0.06410256 0.02564103 0.00000000
88 0.91025641 0.06410256 0.02564103 0.00000000
95 0.91025641 0.06410256 0.02564103 0.00000000
```

Luego de realizar la predicción se procede a crear una nueva columna en la base de datos que vamos a denominar “Predicción” en la que se incluirán los datos de prueba y los de entrenamiento siendo escritos de la siguiente forma:

Cientes[identificador.En,"Prediccion"]=predict(Modelo,Cientes[identificador.En,])

Cientes[-identificador.En,"Prediccion"]=predict(Modelo,Cientes[identificador.En,])

Se pueden visualizar de la siguiente manera el algoritmo y la predicción:

```
18 Cientes[identificador.En,"Prediccion"]=predict(Modelo,Cientes[identificador.En,])
19
20
21 Cientes[-identificador.En,"Prediccion"]=predict(Modelo,Cientes[-identificador.En,])
```

Tabla 3

Tabla con variable y predicción.

	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
1	1	52.80	1.00	1	NA
2	1	64.80	1.00	1	NA
3	1	24.60	1.00	1	0.9102564
4	1	473.59	1.00	1	0.9102564
5	1	52.80	1.00	1	0.9102564
6	1	57.82	1.00	1	NA
7	1	57.82	1.00	1	0.9102564
8	1	79.30	1.00	1	0.9102564
9	1	64.80	1.00	1	NA
10	1	72.00	1.00	1	0.9102564
11	1	72.00	1.00	1	0.9102564
12	1	44.40	1.00	1	0.9102564
13	1	51.66	1.00	1	0.9102564
14	1	46.80	1.00	1	0.9102564
15	1	24.06	1.00	1	NA
16	1	73.10	1.00	1	0.9102564

Una vez verificado que el modelo se procede a llamar a la librería *rpart.plot*

Esta librería nos ayudará a crear el árbol de clasificación usando el conjunto de datos y una fórmula de predicción para graficarlos mediante el siguiente algoritmo:

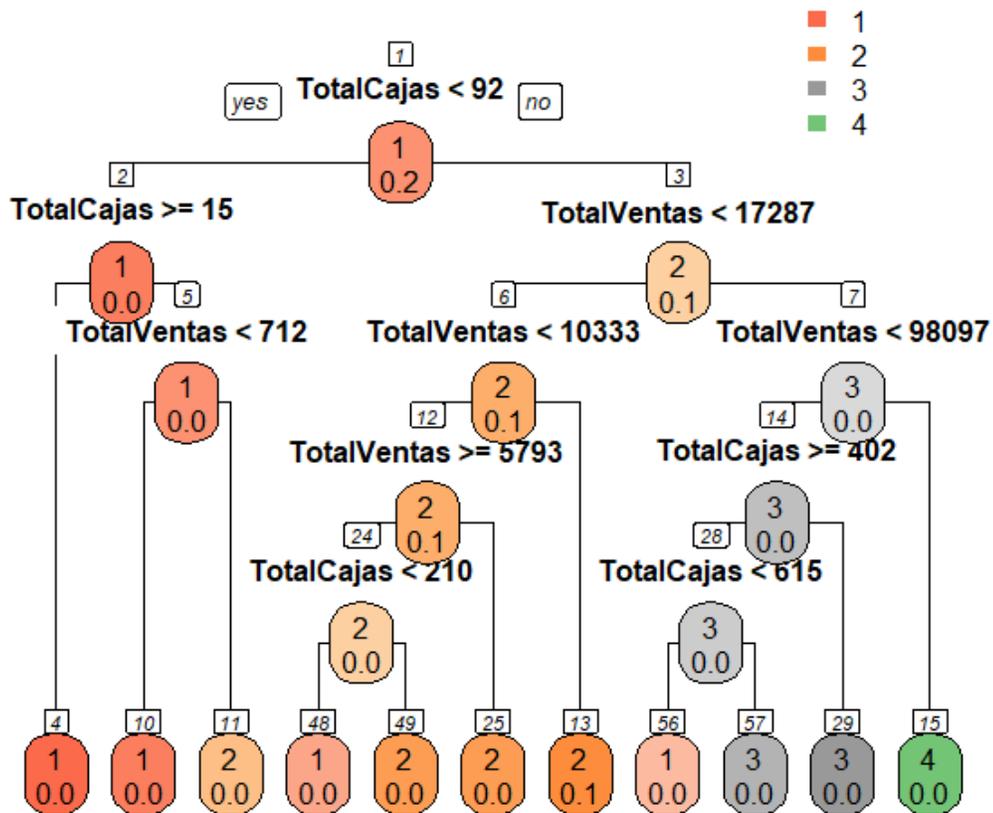
```
rpart.plot(Modelo,digits = -1, type = 1, extra = 10,cex = 1,nn=TRUE,fallen.leaves = TRUE)
```

A continuación, Podemos observar el algoritmo usado para la creación de el árbol de clasificación y el árbol graficado:

```
23 rpart.plot(Modelo,digits = -1, type = 1, extra = 10,cex = 1,nn=TRUE,fallen.leaves = TRUE)
```

Figura 10

Árbol de clasificación.



En el árbol de clasificación podemos observar cómo se clasifico a los tipos de clientes para esto es necesario explicar qué tipo de cliente corresponde a cada número tenemos a los clientes los clientes poco frecuentes están representados con el número 1, los clientes numero 2 son clientes clasificados como normales, los clientes con el numero 3 son los frecuentes y los clasificados con el numero 4 son premium a través del árbol graficado podemos obtener la clasificación de los clientes a partir de las variables que se tomaron los cuales fueron volumen de ventas en dólares y numero de cajas vendidas, con estas dos variables y el tipo de cliente, el árbol los clasifica en el primer layer por total de venta en cajas tomando como ejemplo para la explicación de el mismo tomaremos a un cliente que compro 193.00 cajas y en total de compra en dólares es de 17456.00.

Como él nos indica el primer layer ya que su compra en cajas fue mayor de 92 nos vamos al lado derecho, en el segundo layer tomamos la clasificación izquierda ya que nuestro total de ventas no es mayor a 98097 y en la siguiente layer tenemos un numero de cajas menor a 615 por lo que caemos en la categoría 3, de esta manera se clasifican los clientes en el árbol de clasificación.

Una vez obtenido el modelo de mi árbol de clasificación procedemos a realizar un bosque aleatorio para obtener un mejor rendimiento de generalización para un rendimiento durante un entrenamiento similar, para la creación del bosque aleatorio se utilizará la misma base de datos proporcionada por la empresa TECOPESCA C.A.

Para los bosques aleatorios se debe tener las variables independientes de forma secuencial las cuales en nuestro caso son los clientes, el volumen de ventas en dólares y

el total de venta en caja, y nuestra variable dependiente que es el tipo de cliente, primero se procederá a cargar la base de datos la cual como ya se menciono es la proporcionada por la empresa TECOPESCA C.A a la cual le pondremos de nombre *Cientesrf* y se procede con la siguiente secuencia *Cientesrf = read.csv("../DATA/clientestesis.csv")*, a continuación se procede a convertir nuestra variable clasificatoria *TipoDeCliente* en factor mediante el siguiente algoritmo *Cientesrf\$TipoDeCliente = factor(Cientesrf\$TipoDeCliente)* esto se realiza ya que necesitamos que la variable dependiente sea un factor ya que el arbol de decisión toma la analogía de regresión lineal pero inversa por lo mismo se necesita que la variable dependiente tenga secuencialidad mas no las independientes como en la regresión.

```
1 Cientesrf = read.csv("../DATA/clientestesis.csv")
2 Cientesrf$TipoDeCliente = factor(Cientesrf$TipoDeCliente)
3 library(randomForest)
4 set.seed(123)
```

Una vez cargados los datos y transformados nuestra variable en factor se procede a llamar a la librería *randomForest* la cual es usada para la regresión y clasificación y es muy útil en los casos nen donde existen grupos de modelos muy ligeros, lo que hace esta función es crear un modelo más robusto a través del modelo antes creado y obtener mejores resultados, una vez llamada la librería se procede a plantar una semilla la cual como ya se mencionó se utiliza para generar número aleatorios.

Luego de plantar la semilla se procede a la partición de datos para dividirlos en entrenamiento y prueba como ya se mencionó anteriormente en el anterior ejemplo, tomaremos el 70% para el entrenamiento y el 30% para la prueba, a estos datos le otorgaremos el nombre de *datosEntrenados*

Para luego colocar nuestra variable *TipoDeCliente* y todo esto se expresa con el siguiente algoritmo *datosEntrenados = createDataPartition(Clientesrf\$TipoDeCliente, p=0.7, list = F)*.

```
datosEntrenados = createDataPartition(Clientesrf$TipoDeCliente, p=0.7, list = F)
```

Una vez creados nuestros datos de entrenamientos se procede a crear el modelo de bosque aleatorio en el cual a partir de la base de datos de *Clientesrf* se toma las variables que se van a predecir las cuales son *X(Total de ventas en cajas, Volumen de ventas en dólares y la zona de donde compra el cliente)*, y luego la variable *Y(Tipo de cliente)* para este modelo se toman los datos entrenados los cuales están nombrados como *datosEntrenados* y le damos el número de veces que quiero que se repita la prueba *ntree = 2000* en este caso 2000 veces todo se resume en el siguiente algoritmo.

```
8 RandomTreeModel = randomForest(x=Clientesrf[datosEntrenados,1:3],
9                               y=Clientesrf[datosEntrenados,4],
10                              ntree = 2000, keep.forest = TRUE)
11
```

Una vez creado el modelo de árbol aleatorio se procede a realizar la predicción para después realizar la matriz de confusión y saber si se garantiza el árbol, con la función *predict* en donde se toma el modelo de *RandomTreeModel* y usamos los datos no entrenados es decir el 30% restante.

```
Predicción = predict (RandomTreeModel, Clientesrf[-datosEntrenados,])
Predicción
```

Procedemos a observar *Predicción* para verificar que este en la base de datos.

Tabla 4

Predicción del algoritmo.

```
> Predicción
 1  2  6  9 15 17 18 19 28 29 32 37 38 43 44 46 47 49 56 58 60 62 65 68
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
70 71 78 79 87 88 95 99 101 103 109 112 120 123 124 126 131 132 133 138 142 144 149 150
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
157 162 171 173 180 183 187 188 189 192 193 203 204 206 207 208 216 219 220 228 231 242 246 247
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
248 254 259 260 261 264 266 268 271 274 275 287 297 301 304 311 312 318 321 322 324 325 327 329
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
332 334 338 343 351 354 356 358 369 375 376 377 384 386 387 388 389 393 394 397 408 411 422 428
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
429 432 435 436 438 439 442 443 445 446 458 461 465 467 468 470 473 474 475 481 485 488 492 493
 1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
494 495 500 502 505 507 509 512 513 514 516 517 525 528 531 534 537 538 540 549 550 552 570 578
 1  1  1  1  1  1  1  2  2  2  2  2  2  2  3  1  3  1  1  1  2  2  2  2
581 582 586 590 595 596 602 603 606 608 612 614 621 622 623 624 626 627 634 640 642 644 647 651
 1  1  1  2  2  2  2  2  2  2  2  2  2  2  1  1  2  2  2  2  1  2  2  3
652 654 655 656 657 663 666 669 670 671 672 674 688 690 696 697 699 701 705 710 714 715 717 722
 3  3  3  3  4  2  3  3  3  3  2  3  3  3  1  2  1  3  4  4  4  4  4  4
724
 4
Levels: 1 2 3 4
```

Una vez realizada la predicción procedemos a verificar que este en la base de datos.

Tabla 5

Variables con predicción.

	zona	TotalVentas	TotalCajas	TipoDeCliente	Predicción
1	1	52.80	1.00	1	NA
2	1	64.80	1.00	1	NA
3	1	24.60	1.00	1	1
4	1	473.59	1.00	1	1
5	1	52.80	1.00	1	1
6	1	57.82	1.00	1	NA
7	1	57.82	1.00	1	1
8	1	79.30	1.00	1	1
9	1	64.80	1.00	1	NA
10	1	72.00	1.00	1	1
11	1	72.00	1.00	1	1
12	1	44.40	1.00	1	1
13	1	51.66	1.00	1	1
14	1	46.80	1.00	1	1
15	1	24.06	1.00	1	NA

Una vez creada la predicción procedemos a crear la matriz en probabilidad con el algoritmo **Probabilidad = predict(RandomTreeModel, Clientesrf[-datosEntrenados,], type = "prob")** en el cual tomamos los datos de prueba y predecimos con la función **predict** con esta matriz podemos observar la probabilidad que tienen los clientes de pertenecer a un tipo de cliente.

```
Probabilidad = predict(RandomTreeModel, Clientesrf[-datosEntrenados,], type = "prob")
Probabilidad
```

Tabla 6

Matriz de confusión con probabilidad.

```
> Probabilidad
      1      2      3      4
1  1.0000 0.0000 0.0000 0.0000
2  1.0000 0.0000 0.0000 0.0000
6  1.0000 0.0000 0.0000 0.0000
9  1.0000 0.0000 0.0000 0.0000
15 1.0000 0.0000 0.0000 0.0000
17 1.0000 0.0000 0.0000 0.0000
18 0.9860 0.0000 0.0140 0.0000
19 0.9610 0.0000 0.0390 0.0000
28 1.0000 0.0000 0.0000 0.0000
29 1.0000 0.0000 0.0000 0.0000
32 1.0000 0.0000 0.0000 0.0000
37 1.0000 0.0000 0.0000 0.0000
38 1.0000 0.0000 0.0000 0.0000
43 1.0000 0.0000 0.0000 0.0000
```

Una vez realizada la predicción procedemos a realizar la matriz de confusión con los datos de prueba con la variable que se va a predecir las cuales son **[-datosEntrenados, "TipoDeCliente"]**.

Y se procede a cruzarlas con el modelo de predicción entre el actual y el predicho con **table (Clientesrf[-datosEntrenados, "TipoDeCliente"], Predicción, dnn = c("Actual", "Predicho"))**.

```

20 table(Clientesrf[-datosEntrenados, "TipoDeCliente"], Predicción, dnn = c("Actual", "Predicho"))
21
22

```

Tabla 7

Matriz de confusión actual vs. predicho.

	Predicho			
Actual	1	2	3	4
1	152	7	2	0
2	4	26	6	0
3	1	3	10	0
4	0	0	0	6

Una vez realizada la matriz de confusión podemos observar los resultados, del lado Izquierdo podemos observar los valores actuales y del lado superior el predicho, lo que nos quiere decir esto es la efectividad con la cual se clasificaron clientes, tomando como ejemplo a los clientes tipo 1 que serían los poco frecuentes se clasifico 151 correctamente y 9 se clasificaron incorrectamente, eso nos indica que el modelo es confiable al momento de corregir con efectividad la clasificación de los clientes.

Conclusiones

En la actualidad el uso del Machine Learning está entre las herramientas más eficientes para el análisis de datos dentro de las empresas, este mismo nos permite resumir, ordenar, clasificar, predecir y visualizar los resultados los cuales mediante algoritmos facilitan a las empresas el análisis de los distintos datos que se usen para obtener un resultado en el campo de aplicación.

Sin embargo, existen algunas limitaciones ya que no se pueden considerar variables subjetivas o reconocer datos cualitativos en el Machine Learning ya que para que se desarrolle un algoritmo que arroje un probabilístico se necesitan datos cuantitativos.

La utilización del Machine Learning en las empresas actualmente en los diferentes sectores de la industria tales como de logística, producción, telecomunicaciones, automotriz y transporte optan por el uso del machine learning sin embargo al no contar con una base de datos y la escasa información que se tiene muchas veces no permite el uso adecuado de las herramientas para ser más efectiva a la hora de su utilización.

Las técnicas que se usaron para el desarrollo nos permitieron crear un modelo clasificadorio para los tipos de clientes utilizando valores de los diferentes años registrados. Los modelos utilizados fueron un árbol de clasificación el cual nos clasifico a nuestros clientes mediante sus variables de compras en total de cajas y de ventas en dólares este modelo los clasifico por tipo de clientes y el Bosque aleatorio nos permitió garantizar y corregir los mismos que los clientes estén clasificados correctamente, Sin embargo esto no significa que estos sean los únicos algoritmos de Machine Learning para la clasificación, Machine Learning posee una amplia gama de modelos clasificadorios los cuales pueden usarse para determinar los diferentes tipos de análisis que se necesiten determinar.

La viabilidad del uso de la inteligencia es muy grande ya que garantiza resultados con un alto porcentaje de confianza para después poder realizar una clasificación más ajustada con datos que se obtengan en el futuro y que así se disminuya la cantidad de errores al clasificar.

Recomendaciones

Como recomendaciones para trabajos futuros que use el Machine Learning como herramienta en recomendaría obtener una base de datos amplia y de calidad para un adecuado estudio, esto quiere decir que no exista manipulación de datos porque esto podría causar que la investigación no nos de datos con la eficacia con lo que lo haría si fuesen datos de calidad. Además, se le recomienda dar una delimitación debida a la información dado a que se necesita una base de datos extensa para el funcionamiento de los modelos de clasificación, es decir mientras mayor sea la base de datos mayor eficaz.

Se debe considerar la mayor cantidad de variables para que influyan en el modelo de clasificación ya que mientras mayor sea la cantidad de variables mejor será el algoritmo clasificadorio y por ende la empresa podrá tomar una mejor decisión al momento de realizar un análisis de los datos.

Además se debe obtener una mayor cantidad de datos para obtener un algoritmo clasificadorio más eficiente para el estudio ya que al tener una base de datos con pocas observaciones es decir con pocos clientes se puede obtener ciertos errores que el programa no logra aplicar de la forma correcta y por último se recomendaría contrastar la información obtenida con otro algoritmo de clasificación para determinar cómo se comportan las mismas variables desde otro modelos de clasificación para un análisis más detallado.

Referencias

- Ali, U., Cemil, K., Merve, K., & Abdullah, K. (25 de March de 2021). Board structure, financial performance, corporate social responsibility performance, CSR committee, and CEO duality: Disentangling the connection in healthcare. *Corporate Social Responsibility and Environmental Management*, 28(6), 1730-1748. doi:10.1002/csr.2141
- Arévalo, J. (2018). *Sistema de Recomendación para Empresas de Retail Ecuador*. Obtenido de Repositorio Universidad de las Fuerzas Armadas: <http://repositorio.espe.edu.ec/bitstream/21000/18792/1/T-ESPE-038816.pdf>
- Asociación para el Progreso de la Dirección . (4 de Abril de 2019). *¿Cuáles son los tipos de algoritmos del machine learning?* Obtenido de <https://www.apd.es/algoritmos-del-machine-learning/>
- Azabache, P. (2019). *Optimización del Margen Comercial de Préstamos de Consumo: Algoritmo de Optimización con Machine Learning*. Obtenido de https://repositorio.esan.edu.pe/bitstream/handle/20.500.12640/1743/2019_MAM_17-2_05_R.pdf?sequence=1&isAllowed=y
- Banda, M. (3 de Mayo de 2017). *Todo lo que necesitas saber sobre almacenamiento de datos*. Obtenido de <https://www.powerdata.es/data-warehouse>
- Barnett, M., Irene, H., & Husted, B. (27 de January de 2020). Beyond Good Intentions: Designing CSR Initiatives for Greater Social Impact. *Journal of Management*, 46(6), 937-964. doi:10.1177/0149206319900539
- Bello, E. (20 de Diciembre de 2021). *IBES*. Obtenido de ¿Qué es el minado de Datos o Data Mininig?: <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/>
- Burgal, J. (20 de Junio de 2018). *Deep Learning Básico con Keras*. Obtenido de <https://enmilocalfunciona.io/deep-learning-basico-con-keras-parte-1/>

- Coelho, F. (1 de Junio de 2020). *Digital 55*. Obtenido de Qué es Machine Learning: casos de éxito en empresas: <https://www.digital55.com/innovacion/que-es-machine-learning-casos-exito-empresas/>
- Fasce, G. (2020). *SAS Institute Inc*. Obtenido de Aprendizaje automático: https://www.sas.com/es_es/insights/analytics/machine-learning.html
- Giraldez, M. (18 de Diciembre de 2019). *Necesidades logísticas de los Bancos de Alimentos*. Obtenido de Web Picking: <https://webpicking.com/necesidades-logisticas-de-los-bancos-de-alimentos/>
- González, A. (19 de Agosto de 2017). *Clever data*. Obtenido de ¿Qué es Machine Learning?: <https://cleverdata.io/que-es-machine-learning-big-data/>
- Grobauer, T., Walloschek, & Stocker, E. (2011). *Understanding Cloud Computing Vulnerabilities*. IEEE Security Privacy.
- Grobauer, W., & Stöcker, V. (2011). *Understanding Cloud Computing Vulnerabilities*. IEEE Security & Privacy.
- Ilino, A. (3 de Enero de 2007). *Sinnex*. Obtenido de https://www.sinnexus.com/business_intelligence/
- International Business Machine Corporation. (10 de Agosto de 2017). *¿Qué es Machine Learning?* Obtenido de <https://www.ibm.com/mx-es/analytics/machine-learning#:~:text=Machine%20learning%20es%20una%20forma,m%C3%A1s%20precisos%20basados%20en%20datos.>
- Kohl, J. (1 de Febrero de 2018). *¿Qué son los cubos OLAP?* Obtenido de Auribox Training: <https://blog.auriboxtraining.com/business-intelligence/los-cubos-olap/>
- Lean, G. (22 de Octubre de 2019). *Qué son las redes neuronales y sus funciones*. Obtenido de <https://www.atriainnovation.com/que-son-las-redes-neuronales-y-sus-funciones/#:~:text=Las%20redes%20neuronales%20artificiales%20son,entrada%20hasta%20generar%20una%20salida.>

- León. (2019). *Diseño de un modelo predictivo a través de la técnica de minería de datos 'Random Forest' para la detección de fraude bypass en redes telefónicas en el Ecuador*. Obtenido de <http://201.159.223.180/bitstream/3317/14343/1/T-UCSG-PRE-ING-CIS-251.pdf>
- López, M., Ante, L., & Portie, J. (9 de Marzo de 2021). *BI o Cloud, Inteligencia de Negocio en la Nube*. Obtenido de CIC: <https://www.cic.es/inteligencia-de-negocio-en-la-nube/>
- Mao, Y., He, J., Morrison, A., & Coca, A. (1 de June de 2020). Effects of tourism CSR on employee psychological capital in the COVID-19 crisis: from the perspective of conservation of resources theory. *Current Issues in Tourism* , 24, 2716-2734. doi:10.1080/13683500.2020.1770706
- McMillan, J. (3 de Diciembre de 2020). *La Automatización Avanza en Ecuador*. Obtenido de Ecomex: <https://www.e-comex.com/la-automatizacion-avanza-en-el-ecuador/>
- Perez, G. (4 de Mayo de 2020). *Machine Learning*. Obtenido de <https://www.aicor.com/machine-learning/>
- Prieto, E. (21 de Agosto de 2020). *Es de Latino*. Obtenido de Qué es la clasificación de productos y por qué es importante para sus esfuerzos de marketing.: <https://www.esdelatino.com/que-es-la-clasificacion-de-productos-y-por-que-es-importante-para-sus-esfuerzos-de-marketing/>
- Pushpika, V., Oosterhout, H., Pursey, H., Duran, P., & Essen, M. (21 de May de 2019). Strategic CSR: A Concept Building Meta-Analysis. *Journal of Management Studies*, 57(2), 314-350. doi:10.1111/joms.12514
- Robledano, A. (23 de Septiembre de 2019). *Qué es Python: Características, evolución y futuro*. Obtenido de OpenWebinars: <https://openwebinars.net/blog/que-es-python/>
- Schonberger, M. (2014). *Big Data La Revolución*. Los Datos.

Simon, S. (17 de Diciembre de 2021). *Sistema de toma de decisiones*. Obtenido de <https://www.euroinnova.edu.es/blog/que-es-un-sistema-de-toma-de-decisiones>

Terán, J. (12 de Septiembre de 2021). *El uso de Deep Learning y tipos de aplicaciones*. Obtenido de <https://economia3.com/uso-del-deep-learning/>

Torralba, P. (21 de Mayo de 2021). *IEBS*. Obtenido de Qué son las Redes Neuronales Convolucionales: <https://www.iebschool.com/blog/redes-neuronales-convolucionales/>

Vásquez, T., & Malada, C. (06 de Julio de 2021). *AUTYCOM*. Obtenido de 6 beneficios del aprendizaje automático para la industria: <https://www.autycom.com/beneficios-del-aprendizaje-automatico/>

Anexos

Anexo 1 Scripts Completo Rstudio

```
tesis.R* x Clientes x R tesis.R* x r tesis 2.R* x Untitled2* x
Source on Save Run Source
1 Clientes = read.csv("../DATA/clientestesis.csv",header = TRUE, sep = ",")
2 Clientes$TipoDeCliente = factor(Clientes$TipoDeCliente)
3 library(caret)
4 #plantar una semilla
5 set.seed(123)
6 #Entrenando datos para las clasificacion
7 identificador.En = createDataPartition(Clientes$TipoDeCliente, p=0.7, list = F)
8 identificador.En
9
10 library(rpart)
11 #Construyendo el modelo de arbol de decicion
12 Modelo = rpart(TipoDeCliente ~.,data = Clientes[identificador.En,],method = "class",)
13 Modelo
14 #Prediccion
15 Prediccion = predict(Modelo,Clientes[-identificador.En,])
16 Prediccion
17
18 Clientes[identificador.En,"Prediccion"]=predict(Modelo,Clientes[identificador.En,])
19
20 Clientes[-identificador.En,"Prediccion"]=predict(Modelo,Clientes[-identificador.En,])
21
22 #Funcion rpartPLOT]
23 rpart.plot(Modelo,digits = -1, type = 1, extra = 10,cex = 1,nn=TRUE,fallen.leaves = TRUE)
```

```
r tesis 2.R* x tesis.R* x Clientes x R tesis.R* x Untitled2* x
Source on Save Run Source
1 library(randomForest)
2 library(lattice)
3 library(ggplot2)
4 library(caret)
5 library(ROCR)
6
7 Clientesrf = read.csv("../DATA/clientestesis.csv")
8
9 #Transformo la variable clasificatoria a factor##
10 Clientesrf$TipoDeCliente = factor(Clientesrf$TipoDeCliente)
11 head(Clientesrf$TipoDeCliente)
12
13 ###Plantamos la semilla
14 set.seed(123)
15
16
17 #Entrenamos los datos
18 datosEntrenados = createDataPartition(Clientesrf$TipoDeCliente, p=0.7, list = F)
19
20
21 #Armos el Bosque Aleatorio
22 RandomTreeModel = randomForest(x=Clientesrf[datosEntrenados,1:3],
23                               y=Clientesrf[datosEntrenados,4],
24                               ntree = 2000, keep.forest = TRUE)
25 # Realizo la prediccion
26 Prediccion = predict (RandomTreeModel, Clientesrf[-datosEntrenados,])
27 Prediccion
28
29 #Probabilidad
30 Probabilidad = predict(RandomTreeModel, Clientesrf[-datosEntrenados,], type ="prob")
31 Probabilidad
32
33 #Valores clasificados
34 Clientesrf[datosEntrenados,"Prediccion"]= predict(RandomTreeModel,Clientesrf[datosEntrenados,])
35
36
37
38 ##Realizamos la matriz de confusion##
39 table(Clientesrf[-datosEntrenados, "TipoDeCliente"], Prediccion, dnn = c("Actual","Predicho"))
40
41
42
```

Anexo 2 Predicción de la base de datos

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
1	52.80	1.00	1	NA
2	64.80	1.00	1	NA
3	24.60	1.00	1	0.9102564
4	473.59	1.00	1	0.9102564
5	52.80	1.00	1	0.9102564
6	57.82	1.00	1	NA
7	57.82	1.00	1	0.9102564
8	79.30	1.00	1	0.9102564
9	64.80	1.00	1	NA
10	72.00	1.00	1	0.9102564
11	72.00	1.00	1	0.9102564
12	44.40	1.00	1	0.9102564
13	51.66	1.00	1	0.9102564
14	46.80	1.00	1	0.9102564
15	24.06	1.00	1	NA
16	73.10	1.00	1	0.9102564
17	36.08	1.00	1	NA
18	24.06	1.00	1	NA
19	43.96	1.00	1	NA
20	73.10	1.00	1	0.9102564
21	86.93	1.00	1	0.9102564
22	31.00	1.00	1	0.9102564
23	51.66	1.00	1	0.9102564
24	54.32	1.00	1	0.9102564
25	101.40	1.75	1	0.9102564
26	105.60	2.00	1	0.9102564
27	96.00	2.00	1	0.9102564
28	102.21	2.00	1	NA
29	111.62	2.00	1	NA
30	115.20	2.00	1	0.9102564
31	115.64	2.00	1	0.9102564
32	146.20	2.00	1	NA
33	89.43	2.00	1	0.9102564
34	54.83	2.00	1	0.9102564
35	116.61	2.00	1	0.9102564
36	137.04	2.00	1	0.9102564
37	50.40	3.00	1	NA
38	125.40	3.00	1	NA
39	155.52	3.00	1	0.9102564
40	148.04	3.00	1	0.9102564
41	123.85	3.00	1	0.9102564
42	113.39	3.00	1	0.9102564
43	103.04	3.00	1	NA
44	134.40	4.00	1	NA
45	66.00	4.00	1	0.9102564
46	63.72	4.00	1	NA
47	120.28	5.00	1	NA
48	249.92	5.00	1	0.9102564
49	320.72	5.00	1	NA

	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
47	1	240.72	5.00	1	NA
50	5	110.95	5.00	1	0.9102564
51	5	113.10	5.00	1	0.9102564
52	1	310.56	5.00	1	0.9102564
53	1	202.34	5.00	1	0.9102564
54	1	353.84	5.00	1	0.9102564
55	1	134.54	5.00	1	0.9102564
56	1	220.94	5.00	1	NA
57	1	123.56	5.00	1	0.9102564
58	1	422.00	6.00	1	NA
59	1	421.30	6.00	1	0.9102564
60	1	388.80	6.00	1	NA
61	5	3475.90	6.00	1	0.1428571
62	1	356.87	6.00	1	NA
63	1	413.54	6.00	1	0.9102564
64	1	594.93	6.00	1	0.9102564
65	1	365.76	6.00	1	NA
66	1	412.54	7.00	1	0.9102564
67	1	210.94	7.00	1	0.9102564
68	1	462.56	8.00	1	NA
69	1	249.44	8.00	1	0.9102564
70	1	619.94	8.00	1	NA
71	5	275.22	8.00	1	NA
72	1	456.64	8.00	1	0.9102564
73	1	234.54	8.00	1	0.9102564
74	1	243.54	8.00	1	0.9102564
75	3	362.16	9.00	1	0.9102564
76	1	622.08	10.00	1	0.9102564
77	3	627.00	10.00	1	0.9102564
78	4	685.00	10.00	1	NA
79	6	729.00	10.00	1	NA
80	1	677.00	10.00	1	0.9102564
81	3	685.00	10.00	1	0.9102564
82	1	700.00	11.00	1	0.9102564
83	5	668.00	11.00	1	0.9102564
84	1	655.00	11.00	1	0.9102564
85	1	633.60	12.00	1	0.9102564
86	1	699.84	12.00	1	0.9102564
87	4	627.42	12.00	1	NA
88	4	654.65	12.00	1	NA
89	3	676.00	12.00	1	0.9102564
90	2	643.00	12.00	1	0.9102564
91	1	695.00	12.00	1	0.9102564
92	1	658.00	12.00	1	0.9102564
93	6	447.88	13.00	1	0.9102564
94	5	465.65	13.00	1	0.9102564
95	1	475.00	13.00	1	NA
96	1	477.00	13.00	1	0.9102564
97	5	453.00	14.00	1	0.9102564
98	4	543.00	14.00	1	0.9102564

	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
98	4	543.00	14.00	1	0.9102564
99	4	532.00	14.00	1	NA
100	6	523.00	14.00	1	0.9102564
101	1	568.00	14.00	1	NA
102	6	598.00	14.00	1	0.9102564
103	1	547.00	14.00	1	NA
104	2	586.00	14.00	1	0.9102564
105	1	598.00	14.00	1	0.9102564
106	1	531.00	14.00	1	0.9102564
107	4	612.00	15.00	1	0.9726027
108	4	634.00	15.00	1	0.9726027
109	5	648.00	15.00	1	NA
110	2	698.00	15.00	1	0.9726027
111	3	678.00	15.00	1	0.9726027
112	3	890.00	15.00	1	NA
113	2	862.00	15.00	1	0.9726027
114	1	958.00	15.00	1	0.9726027
115	3	838.00	15.00	1	0.9726027
116	1	891.00	15.00	1	0.9726027
117	1	932.00	15.00	1	0.9726027
118	4	900.00	16.00	1	0.9726027
119	3	926.00	16.00	1	0.9726027
120	6	975.00	16.00	1	NA
121	5	910.00	16.00	1	0.9726027
122	2	927.00	16.00	1	0.9726027
123	1	890.00	16.00	1	NA
124	2	836.00	16.00	1	NA
125	1	989.00	16.00	1	0.9726027
126	1	925.00	16.00	1	NA
127	1	991.44	17.00	1	0.9726027
128	4	882.00	17.00	1	0.9726027
129	1	897.00	17.00	1	0.9726027
130	2	783.00	17.00	1	0.9726027
131	1	943.00	17.00	1	NA
132	1	907.00	17.00	1	NA
133	6	836.00	17.00	1	NA
134	6	801.00	18.00	1	0.9726027
135	4	977.00	18.00	1	0.9726027
136	4	797.00	18.00	1	0.9726027
137	5	858.00	18.00	1	0.9726027
138	3	991.00	18.00	1	NA
139	1	889.00	18.00	1	0.9726027
140	2	775.00	19.00	1	0.9726027
141	3	943.00	19.00	1	0.9726027
142	4	881.00	19.00	1	NA
143	2	810.00	19.00	1	0.9726027
144	1	790.00	19.00	1	NA
145	3	1328.00	20.00	1	0.9726027
146	5	1033.24	20.00	1	0.9726027
147	6	1017.60	20.00	1	0.9726027

▲	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
148	6	767.95	20.00	1	0.9726027
149	2	1065.00	20.00	1	NA
150	4	1157.00	20.00	1	NA
151	2	1212.00	20.00	1	0.9726027
152	1	1168.00	20.00	1	0.9726027
153	6	1045.00	20.00	1	0.9726027
154	4	1180.00	21.00	1	0.9726027
155	5	1048.00	21.00	1	0.9726027
156	6	1215.00	21.00	1	0.9726027
157	1	1074.00	21.00	1	NA
158	1	1200.00	21.00	1	0.9726027
159	4	1011.79	22.00	1	0.9726027
160	4	976.35	22.00	1	0.9726027
161	2	1107.00	22.00	1	0.9726027
162	2	1155.00	22.00	1	NA
163	1	1155.00	22.00	1	0.9726027
164	1	1126.00	22.00	1	0.9726027
165	4	1210.00	23.00	1	0.9726027
166	4	1209.00	23.00	1	0.9726027
167	2	1006.00	23.00	1	0.9726027
168	3	1120.00	23.00	1	0.9726027
169	2	1211.00	23.00	1	0.9726027
170	6	1174.00	23.00	1	0.9726027
171	1	1082.00	23.00	1	NA
172	1	1101.00	23.00	1	0.9726027
173	1	1097.00	23.00	1	NA
174	4	1226.50	24.00	1	0.9726027
175	3	1143.00	24.00	1	0.9726027
176	4	1118.00	24.00	1	0.9726027
177	4	1177.00	24.00	1	0.9726027
178	5	1197.00	24.00	1	0.9726027
179	6	1023.00	24.00	1	0.9726027
180	5	1061.00	24.00	1	NA
181	3	1006.00	24.00	1	0.9726027
182	4	1197.00	25.00	1	0.9726027
183	4	978.00	25.00	1	NA
184	4	1054.00	25.00	1	0.9726027
185	5	1061.00	25.00	1	0.9726027
186	5	1120.00	25.00	1	0.9726027
187	1	1034.00	25.00	1	NA
188	6	1132.00	25.00	1	NA
189	2	1171.00	25.00	1	NA
190	6	1133.00	25.00	1	0.9726027
191	1	1216.00	25.00	1	0.9726027
192	1	1234.00	25.00	1	NA
193	4	982.00	26.00	1	NA
194	4	1142.00	26.00	1	0.9726027
195	5	1078.00	26.00	1	0.9726027
196	5	1042.00	26.00	1	0.9726027
197	2	1101.00	26.00	1	0.9726027

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion	
198	4	1149.00	27.00	1	0.9726027
199	5	1192.00	27.00	1	0.9726027
200	5	1026.00	27.00	1	0.9726027
201	5	1052.00	27.00	1	0.9726027
202	5	1048.00	27.00	1	0.9726027
203	5	1199.00	27.00	1	NA
204	1	1145.00	27.00	1	NA
205	3	1136.00	27.00	1	0.9726027
206	1	1014.00	27.00	1	NA
207	5	1103.00	28.00	1	NA
208	2	1109.00	28.00	1	NA
209	1	1173.00	28.00	1	0.9726027
210	1	1201.00	28.00	1	0.9726027
211	3	1405.45	29.00	1	0.9726027
212	4	1475.88	29.00	1	0.9726027
213	4	3874.00	29.00	1	0.9726027
214	4	36578.00	29.00	1	0.9726027
215	6	6876.00	29.00	1	0.9726027
216	6	4567.00	29.00	1	NA
217	1	3876.00	29.00	1	0.9726027
218	6	3567.00	29.00	1	0.9726027
219	2	5678.00	29.00	1	NA
220	2	3214.00	29.00	1	NA
221	1	3567.00	29.00	1	0.9726027
222	1	3006.90	30.00	1	0.9726027
223	2	1426.25	30.00	1	0.9726027
224	4	1289.76	30.00	1	0.9726027
225	5	4132.94	30.00	1	0.9726027
226	5	34565.00	30.00	1	0.9726027
227	1	23456.00	30.00	1	0.9726027
228	1	3007.50	30.00	1	NA
229	1	1424.00	31.00	1	0.9726027
230	1	1349.00	31.00	1	0.9726027
231	5	1325.00	31.00	1	NA
232	1	1317.00	31.00	1	0.9726027
233	1	1361.00	31.00	1	0.9726027
234	4	1308.00	32.00	1	0.9726027
235	5	1292.00	32.00	1	0.9726027
236	2	1370.00	32.00	1	0.9726027
237	1	1313.00	32.00	1	0.9726027
238	1	1290.00	32.00	1	0.9726027
239	3	1383.00	32.00	1	0.9726027
240	1	1303.00	32.00	1	0.9726027
241	3	1396.00	33.00	1	0.9726027
242	4	1384.00	33.00	1	NA
243	2	1318.00	33.00	1	0.9726027
244	1	1404.00	33.00	1	0.9726027
245	5	1299.00	34.00	1	0.9726027
246	5	1312.00	34.00	1	NA
247	2	1354.00	34.00	1	NA

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion	
248	6	1401.00	34.00	1	NA
249	2	1335.00	34.00	1	0.9726027
250	6	1401.00	34.00	1	0.9726027
251	3	1367.00	34.00	1	0.9726027
252	3	1319.00	34.00	1	0.9726027
253	2	1377.00	35.00	1	0.9726027
254	1	1404.00	35.00	1	NA
255	1	1422.00	35.00	1	0.9726027
256	3	1366.00	35.00	1	0.9726027
257	1	1371.00	35.00	1	0.9726027
258	4	1303.00	36.00	1	0.9726027
259	1	1324.00	36.00	1	NA
260	1	1382.00	36.00	1	NA
261	1	1391.00	36.00	1	NA
262	1	1361.00	36.00	1	0.9726027
263	1	1419.00	36.00	1	0.9726027
264	5	1299.00	37.00	1	NA
265	5	1424.00	37.00	1	0.9726027
266	6	1415.00	37.00	1	NA
267	6	1353.00	37.00	1	0.9726027
268	5	1296.00	37.00	1	NA
269	5	1400.00	37.00	1	0.9726027
270	5	1401.00	37.00	1	0.9726027
271	2	1425.00	37.00	1	NA
272	3	1732.34	38.00	1	0.9726027
273	5	1408.00	38.00	1	0.9726027
274	4	1396.00	38.00	1	NA
275	1	1347.00	38.00	1	NA
276	3	1291.00	38.00	1	0.9726027
277	2	1306.00	38.00	1	0.9726027
278	1	1338.00	38.00	1	0.9726027
279	1	1390.00	39.00	1	0.9726027
280	6	1297.00	39.00	1	0.9726027
281	3	1416.00	39.00	1	0.9726027
282	3	1413.00	39.00	1	0.9726027
283	4	3080.00	40.00	1	0.9726027
284	5	1328.00	40.00	1	0.9726027
285	5	1409.00	40.00	1	0.9726027
286	5	1379.00	40.00	1	0.9726027
287	2	1411.00	40.00	1	NA
288	4	1319.00	41.00	1	0.9726027
289	2	1380.00	41.00	1	0.9726027
290	2	1386.00	41.00	1	0.9726027
291	1	1373.00	41.00	1	0.9726027
292	3	1290.00	41.00	1	0.9726027
293	5	1396.00	41.00	1	0.9726027
294	3	1383.00	41.00	1	0.9726027
295	2	1319.00	41.00	1	0.9726027
296	5	2163.14	42.00	1	0.9726027
297	2	2333.00	42.00	1	NA

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion	
299	1	2372.00	42.00	1	0.9726027
300	3	2414.00	42.00	1	0.9726027
301	4	3400.00	43.00	1	NA
302	1	3407.00	43.00	1	0.9726027
303	5	2453.00	43.00	1	0.9726027
304	6	2673.00	43.00	1	NA
305	5	2893.00	43.00	1	0.9726027
306	4	1546.00	43.00	1	0.9726027
307	5	3836.00	43.00	1	0.9726027
308	5	2984.00	43.00	1	0.9726027
309	2	2093.00	43.00	1	0.9726027
310	1	1894.00	43.00	1	0.9726027
311	4	1876.00	44.00	1	NA
312	6	1845.00	44.00	1	NA
313	1	2456.00	44.00	1	0.9726027
314	3	2136.00	44.00	1	0.9726027
315	1	1984.00	44.00	1	0.9726027
316	6	1770.44	45.00	1	0.9726027
317	5	1894.00	45.00	1	0.9726027
318	2	2435.00	45.00	1	NA
319	3	1784.00	45.00	1	0.9726027
320	1	1643.00	45.00	1	0.9726027
321	1	1547.00	45.00	1	NA
322	1	2654.00	45.00	1	NA
323	4	3450.00	46.00	1	0.9726027
324	4	2745.00	46.00	1	NA
325	5	3888.00	46.00	1	NA
326	2	1996.00	46.00	1	0.9726027
327	3	3496.00	46.00	1	NA
328	1	2385.00	46.00	1	0.9726027
329	1	1456.00	47.00	1	NA
330	2	3851.00	47.00	1	0.9726027
331	4	2145.00	47.00	1	0.9726027
332	2	1675.00	47.00	1	NA
333	2	2346.00	47.00	1	0.9726027
334	1	1749.00	47.00	1	NA
335	6	1843.00	48.00	1	0.9726027
336	4	1894.00	48.00	1	0.9726027
337	3	1789.00	48.00	1	0.9726027
338	6	1889.00	48.00	1	NA
339	3	1689.00	48.00	1	0.9726027
340	3	2224.18	49.00	1	0.9726027
341	4	2254.00	49.00	1	0.9726027
342	2	2213.83	50.00	1	0.9726027
343	3	2435.00	50.00	1	NA
344	4	2174.00	50.00	1	0.9726027
345	3	1998.00	50.00	1	0.9726027
346	5	2845.00	50.00	1	0.9726027
347	3	2843.00	50.00	1	0.9726027
348	1	2784.00	50.00	1	0.9726027

	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
349	2	2458.00	50.00	1	0.97260274
350	1	3175.00	50.00	1	0.97260274
351	5	2215.40	51.00	1	NA
352	5	2399.90	52.00	1	0.97260274
353	4	4275.00	52.00	1	0.97260274
354	4	4395.00	52.00	1	NA
355	4	3491.00	52.00	1	0.97260274
356	5	3487.00	52.00	1	NA
357	5	3723.00	52.00	1	0.97260274
358	1	4090.00	52.00	1	NA
359	3	3462.00	52.00	1	0.97260274
360	4	3294.00	53.00	1	0.97260274
361	4	3507.00	53.00	1	0.97260274
362	5	4886.00	53.00	1	0.97260274
363	1	3872.00	53.00	1	0.97260274
364	2	4504.00	53.00	1	0.97260274
365	1	3151.00	53.00	1	0.97260274
366	2	4150.00	54.00	1	0.97260274
367	5	4764.00	54.00	1	0.97260274
368	6	4738.00	54.00	1	0.97260274
369	5	4589.00	54.00	1	NA
370	3	4583.00	54.00	1	0.97260274
371	1	3875.00	54.00	1	0.97260274
372	1	3641.00	54.00	1	0.97260274
373	2	2426.46	55.00	1	0.97260274
374	4	2386.89	55.00	1	0.97260274
375	5	3876.00	55.00	1	NA
376	1	3317.00	55.00	1	NA
377	3	4925.00	55.00	1	NA
378	1	4580.00	55.00	1	0.97260274
379	2	3962.00	55.00	1	0.97260274
380	5	3694.00	56.00	1	0.97260274
381	4	4194.00	56.00	1	0.97260274
382	5	3990.00	56.00	1	0.97260274
383	6	4115.00	56.00	1	0.97260274
384	3	4145.00	56.00	1	NA
385	5	3981.00	56.00	1	0.97260274
386	5	3455.00	56.00	1	NA
387	1	3295.00	56.00	1	NA
388	1	3451.00	56.00	1	NA
389	2	4793.00	56.00	1	NA
390	6	2349.48	57.00	1	0.97260274
391	6	4564.00	57.00	1	0.97260274
392	3	3295.00	57.00	1	0.97260274
393	1	3713.00	57.00	1	NA
394	1	2105.00	58.00	1	NA
395	2	2817.28	58.00	1	0.97260274
396	4	5860.00	58.00	1	0.97260274
397	6	5858.00	58.00	1	NA
398	1	5859.00	58.00	1	0.97260274

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion	
399	2	5855.00	58.00	1	0.97260274
400	5	5851.00	59.00	1	0.97260274
401	3	5853.00	59.00	1	0.97260274
402	1	5852.00	59.00	1	0.97260274
403	1	5851.00	59.00	1	0.97260274
404	2	2683.01	60.00	1	0.97260274
405	1	5857.00	60.00	1	0.97260274
406	6	5856.00	60.00	1	0.97260274
407	4	5855.00	60.00	1	0.97260274
408	6	5850.00	60.00	1	NA
409	1	5852.00	60.00	1	0.97260274
410	1	5850.00	60.00	1	0.97260274
411	5	5856.00	61.00	1	NA
412	5	5855.00	61.00	1	0.97260274
413	3	5861.00	61.00	1	0.97260274
414	6	5852.00	61.00	1	0.97260274
415	6	5860.00	61.00	1	0.97260274
416	6	5856.00	61.00	1	0.97260274
417	2	5861.00	61.00	1	0.97260274
418	1	5860.00	62.00	1	0.97260274
419	2	5851.00	62.00	1	0.97260274
420	1	5857.00	62.00	1	0.97260274
421	5	5850.00	62.00	1	0.97260274
422	5	5851.00	62.00	1	NA
423	1	5857.00	62.00	1	0.97260274
424	5	5854.00	63.00	1	0.97260274
425	4	5857.00	63.00	1	0.97260274
426	4	5854.00	63.00	1	0.97260274
427	1	5857.00	63.00	1	0.97260274
428	5	5850.00	63.00	1	NA
429	5	5853.00	64.00	1	NA
430	6	5851.00	64.00	1	0.97260274
431	5	5861.00	64.00	1	0.97260274
432	2	5861.00	64.00	1	NA
433	5	5853.00	64.00	1	0.97260274
434	6	5853.00	64.00	1	0.97260274
435	3	5861.00	64.00	1	NA
436	1	5858.00	64.00	1	NA
437	1	5855.00	64.00	1	0.97260274
438	1	5851.00	65.00	1	NA
439	1	5856.00	65.00	1	NA
440	4	5851.00	65.00	1	0.97260274
441	6	5851.00	65.00	1	0.97260274
442	2	5859.00	65.00	1	NA
443	5	5853.00	65.00	1	NA
444	5	5860.00	65.00	1	0.97260274
445	2	5852.00	65.00	1	NA
446	1	5854.00	65.00	1	NA
447	1	5853.00	65.00	1	0.97260274
448	6	2703.13	66.00	1	0.97260274

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion	
449	4	2759.00	66.00	1	0.97260274
450	5	4733.00	66.00	1	0.97260274
451	3	4315.00	66.00	1	0.97260274
452	3	2942.00	66.00	1	0.97260274
453	4	3319.45	67.00	1	0.97260274
454	4	1808.00	67.00	1	0.97260274
455	4	1498.00	67.00	1	0.97260274
456	1	1969.00	67.00	1	0.97260274
457	3	4680.00	67.00	1	0.97260274
458	2	3022.00	67.00	1	NA
459	1	4832.00	67.00	1	0.97260274
460	6	3976.00	67.00	1	0.97260274
461	6	3830.00	67.00	1	NA
462	1	5050.00	67.00	1	0.97260274
463	1	4484.00	67.00	1	0.97260274
464	4	3126.08	68.00	1	0.97260274
465	1	21802.00	68.00	1	NA
466	2	14914.00	68.00	1	0.97260274
467	3	15491.00	68.00	1	NA
468	6	16141.00	68.00	1	NA
469	5	17154.00	68.00	1	0.97260274
470	1	16354.00	68.00	1	NA
471	3	14610.00	68.00	1	0.97260274
472	1	21335.00	68.00	1	0.97260274
473	2	14493.00	68.00	1	NA
474	1	16964.00	68.00	1	NA
475	1	14323.00	68.00	1	NA
476	4	16244.00	69.00	1	0.97260274
477	4	21800.00	69.00	1	0.97260274
478	4	20586.00	69.00	1	0.97260274
479	4	14761.00	69.00	1	0.97260274
480	2	18879.00	69.00	1	0.97260274
481	1	19098.00	69.00	1	NA
482	6	13245.00	69.00	1	0.97260274
483	1	12010.00	69.00	1	0.97260274
484	1	13788.00	69.00	1	0.97260274
485	1	21616.00	69.00	1	NA
486	1	16932.00	69.00	1	0.97260274
487	2	13250.00	69.00	1	0.97260274
488	1	17441.00	69.00	1	NA
489	1	15570.00	69.00	1	0.97260274
490	4	21037.00	70.00	1	0.97260274
491	1	19349.00	70.00	1	0.97260274
492	6	19994.00	70.00	1	NA
493	1	16064.00	70.00	1	NA
494	1	17201.00	70.00	1	NA
495	2	3304.80	75.00	1	NA
496	4	3722.71	75.00	1	0.97260274
497	3	3527.51	77.00	1	0.97260274
498	2	3461.82	78.00	1	0.97260274

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
499	3	3412.79	80.00 1	0.97260274
500	6	3960.55	82.00 1	NA
501	6	3906.53	82.00 1	0.97260274
502	2	3909.55	88.00 1	NA
503	4	3900.99	90.00 1	0.97260274
504	4	4342.43	90.00 1	0.97260274
505	5	5909.52	90.00 1	NA
506	2	3833.12	97.00 1	0.08333333
507	5	5110.42	100.00 1	NA
508	5	4769.17	100.00 1	0.08333333
509	2	4693.99	111.00 1	NA
510	5	5849.64	118.00 1	0.66666667
511	2	5861.30	120.00 1	0.66666667
512	3	4834.50	133.00 1	NA
513	6	6947.89	133.00 1	NA
514	6	4834.50	133.00 1	NA
515	6	6562.99	137.00 1	0.66666667
516	2	7439.32	144.00 1	NA
517	3	7372.85	144.00 1	NA
518	5	7737.24	150.00 1	0.66666667
519	4	7109.76	156.00 1	0.66666667
520	4	8633.57	162.00 1	0.66666667
521	6	8960.55	168.00 1	0.66666667
522	2	9286.09	192.00 1	0.66666667
523	6	10229.06	198.00 1	0.66666667
524	3	9753.35	207.00 1	0.66666667
525	3	10337.96	208.00 1	NA
526	4	9743.01	237.00 1	0.12500000
527	2	12175.20	294.00 1	0.04878049
528	1	11400.00	300.00 1	NA
529	3	14871.32	300.00 1	0.04878049
530	4	18067.27	414.00 1	0.55555556
531	3	21465.43	426.00 1	NA
532	2	24387.06	483.00 1	0.55555556
533	3	27456.39	550.00 1	0.55555556
534	2	26839.23	562.00 1	NA
535	4	27523.43	573.00 1	0.55555556
536	2	26086.32	606.00 1	0.55555556
537	6	35792.30	763.00 1	NA
538	1	321.21	5.61 2	NA
539	1	472.10	6.50 2	0.91025641
540	1	657.90	9.00 2	NA
541	1	492.48	9.00 2	0.91025641
542	1	723.00	10.00 2	0.14285714
543	5	666.00	10.00 2	0.91025641
544	6	680.00	11.00 2	0.91025641
545	1	639.00	11.00 2	0.91025641
546	1	965.00	13.00 2	0.14285714
547	4	774.00	14.00 2	0.14285714
548	4	931.00	14.00 2	0.14285714

	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
549	6	955.00	14.00	2	NA
550	4	5572.00	285.00	2	NA
551	4	11168.00	105.00	2	0.04878049
552	4	10347.00	226.00	2	NA
553	5	5489.00	105.00	2	0.08333333
554	2	10045.00	242.00	2	0.12500000
555	4	7529.00	212.00	2	0.12500000
556	3	11377.00	128.00	2	0.04878049
557	6	16774.00	115.00	2	0.04878049
558	1	5434.00	329.00	2	0.08333333
559	4	10011.00	185.00	2	0.66666667
560	6	11809.00	237.00	2	0.04878049
561	4	11477.00	282.00	2	0.04878049
562	2	5439.00	214.00	2	0.08333333
563	1	5407.00	109.00	2	0.08333333
564	3	5474.00	244.00	2	0.08333333
565	5	6341.00	94.00	2	0.66666667
566	4	5492.00	110.00	2	0.08333333
567	4	11218.00	126.00	2	0.04878049
568	5	11197.00	212.00	2	0.04878049
569	2	11725.00	111.00	2	0.04878049
570	2	11559.00	282.00	2	NA
571	6	9579.00	232.00	2	0.12500000
572	6	7453.00	157.00	2	0.66666667
573	3	11501.00	186.00	2	0.04878049
574	2	11296.00	350.00	2	0.04878049
575	4	11583.00	239.00	2	0.04878049
576	5	11224.00	209.00	2	0.04878049
577	4	11452.00	309.00	2	0.04878049
578	1	11782.00	315.00	2	NA
579	3	11811.00	313.00	2	0.04878049
580	4	11239.00	134.00	2	0.04878049
581	4	1412.00	33.00	2	NA
582	1	1365.00	33.00	2	NA
583	1	1411.00	35.00	2	0.97260274
584	4	1357.00	36.00	2	0.97260274
585	5	1357.00	37.00	2	0.97260274
586	5	1578.40	37.41	2	NA
587	5	5168.00	130.00	2	0.08333333
588	6	5449.00	209.00	2	0.08333333
589	4	5727.00	224.00	2	0.08333333
590	4	4352.00	150.00	2	NA
591	6	5469.00	175.00	2	0.08333333
592	4	10721.00	239.00	2	0.04878049
593	4	5509.00	338.00	2	0.08333333
594	6	2599.00	48.00	2	0.97260274
595	2	2303.00	48.00	2	NA
596	6	2122.00	49.00	2	NA
597	1	3654.90	50.00	2	0.97260274
598	4	11050.00	283.00	2	0.04878049

	zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
599	3	11040.00	159.00	2	0.04878049
600	6	10973.00	103.00	2	0.04878049
601	4	10941.00	218.00	2	0.04878049
602	4	11825.00	212.00	2	NA
603	6	11257.00	180.00	2	NA
604	1	11571.00	341.00	2	0.04878049
605	1	11473.00	305.00	2	0.04878049
606	4	11389.00	343.00	2	NA
607	6	11389.00	338.00	2	0.04878049
608	1	5385.00	299.00	2	NA
609	4	8751.00	277.00	2	0.12500000
610	4	11171.00	329.00	2	0.04878049
611	1	16868.00	311.00	2	0.04878049
612	1	11743.00	151.00	2	NA
613	5	11251.00	228.00	2	0.04878049
614	6	5385.00	295.00	2	NA
615	2	9511.00	219.00	2	0.12500000
616	4	5363.00	246.00	2	0.08333333
617	4	6798.00	297.00	2	0.12500000
618	4	4583.00	359.00	2	0.08333333
619	1	8811.00	123.00	2	0.66666667
620	4	11585.00	308.00	2	0.04878049
621	1	16263.00	164.00	2	NA
622	2	12719.00	298.00	2	NA
623	2	12132.00	232.00	2	NA
624	4	11221.00	355.00	2	NA
625	1	10914.00	127.00	2	0.04878049
626	5	5908.00	338.00	2	NA
627	1	11075.00	234.00	2	NA
628	4	8758.00	275.00	2	0.12500000
629	1	4521.00	192.00	2	0.08333333
630	4	10945.00	128.00	2	0.04878049
631	4	5419.00	325.00	2	0.08333333
632	3	11420.00	135.00	2	0.04878049
633	6	5043.00	140.00	2	0.08333333
634	4	3973.46	90.00	2	NA
635	4	5736.78	118.00	2	0.08333333
636	4	5535.39	122.00	2	0.08333333
637	6	5345.09	140.00	2	0.08333333
638	6	7252.90	164.00	2	0.66666667
639	1	12878.51	189.00	2	0.04878049
640	2	9123.63	195.00	2	NA
641	4	10437.67	222.00	2	0.04878049
642	4	11051.83	228.00	2	NA
643	5	11537.60	229.00	2	0.04878049
644	4	11968.48	233.00	2	NA
645	6	10909.06	240.00	2	0.04878049
646	6	11859.59	241.00	2	0.04878049
647	4	15298.17	363.00	2	NA

zona	TotalVentas	TotalCajas	TipoDeCliente	Prediccion
649	21905.65	449.00	2	0.55555556
650	21842.82	585.00	2	0.55555556
651	35504.80	839.00	2	NA
652	37781.64	911.00	2	NA
653	42772.68	996.00	2	0.00000000
654	58386.82	1158.00	2	NA
655	86916.17	1619.00	2	NA
656	75234.85	1707.00	2	NA
657	209423.57	3238.00	2	NA
658	731.00	10.00	3	0.14285714
659	647.00	10.00	3	0.91025641
660	657.00	12.00	3	0.91025641
661	915.00	13.00	3	0.14285714
662	13914.00	873.00	3	0.04878049
663	4970.00	552.00	3	NA
664	38253.00	144.00	3	0.00000000
665	27090.00	515.00	3	0.55555556
666	16804.00	820.00	3	NA
667	36559.00	339.00	3	0.00000000
668	37452.00	390.00	3	0.00000000
669	33923.00	215.00	3	NA
670	33775.00	343.00	3	NA
671	23875.00	169.00	3	NA
672	5160.00	637.00	3	NA
673	3474.00	294.00	3	0.08333333
674	28345.00	738.00	3	NA
675	3561.00	44.00	3	0.97260274
676	3430.00	44.00	3	0.97260274
677	3269.08	46.00	3	0.97260274
678	22705.00	751.00	3	0.00000000
679	17456.00	193.00	3	0.00000000
680	24556.00	98.00	3	0.00000000
681	18264.00	629.00	3	0.00000000
682	18724.00	626.00	3	0.00000000
683	38987.00	498.00	3	0.55555556
684	20280.00	195.00	3	0.00000000
685	22821.00	876.00	3	0.00000000
686	16282.00	284.00	3	0.04878049
687	34666.00	184.00	3	0.00000000
688	28838.00	202.00	3	NA
689	3291.00	140.00	3	0.08333333
690	34017.00	810.00	3	NA
691	26516.00	838.00	3	0.00000000
692	5292.00	445.00	3	0.08333333
693	36596.00	800.00	3	0.00000000
694	19140.00	788.00	3	0.00000000
695	26448.00	623.00	3	0.00000000
696	3255.00	90.00	3	NA
697	5304.34	144.00	3	NA

698	1	19188.89	265.00	3	0.00000000
699	3	14601.55	295.00	3	NA
700	1	12155.41	325.00	3	0.04878049
701	5	23290.95	555.00	3	NA
702	3	40082.34	920.00	3	0.00000000
703	2	74995.61	1601.00	3	0.00000000
704	3	78664.73	1777.00	3	0.00000000
705	2	159091.70	3310.00	3	NA
706	1	198350.00	3171.00	4	0.00000000
707	2	222135.00	4517.00	4	0.00000000
708	2	200016.00	2717.00	4	0.00000000
709	5	137681.00	1078.00	4	0.00000000
710	5	119651.00	3733.00	4	NA
711	3	151655.00	3404.00	4	0.00000000
712	1	149840.00	3100.00	4	0.00000000
713	2	153805.00	877.00	4	0.00000000
714	3	222352.00	3629.00	4	NA
715	1	152460.00	3521.00	4	NA
716	5	122256.00	1049.00	4	0.00000000
717	3	166732.00	2799.00	4	NA
718	1	168556.00	1221.00	4	0.00000000
719	1	117529.00	2570.00	4	0.00000000
720	1	150199.00	3552.00	4	0.00000000
721	1	33184.53	628.00	4	0.00000000
722	3	140737.59	2883.00	4	NA
723	2	155244.94	3068.00	4	0.00000000
724	5	224922.64	4586.00	4	NA
725	1	623169.54	19760.00	4	0.00000000



DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **Camacho Rosero Micaella Patricia** con C.C: **0924168644**, **Reyna Muentes Cristhian Aurelio** con C.C: **0923220891**; autores del trabajo de integración curricular: **Análisis de la clasificación basado en el concepto Machine Learning por tipo de cliente en la empresa Tecopesca C.A**, previo a la obtención del título 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 titulación 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 titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, **21 de febrero de 2022**

Nombre: **Reyna Muentes Cristhian Aurelio**
C.C: **0924168644**

Nombre: **Camacho Rosero Micaella Patricia**
C.C: **0923220891**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Análisis de la clasificación basado en el concepto Machine Learning por tipo de cliente en la empresa Tecopesca C.A		
AUTOR(ES)	Reyna Muentes Cristhian Aurelio Camacho Rosero Micaella Patricia		
REVISOR(ES)/TUTOR(ES)	Félix Carrera Buri		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Facultad de Ciencias Económicas, Administrativas y Empresariales		
CARRERA:	Negocios Internacionales		
TÍTULO OBTENIDO:	Licenciado en Negocios Internacionales		
FECHA DE PUBLICACIÓN:	21 de febrero de 2022	No. DE PÁGINAS:	(80 de páginas)
ÁREAS TEMÁTICAS:	Business Intelligence, Análisis de Datos, Computer Science		
PALABRAS CLAVES/ KEYWORDS:	Machine Learning, árboles clasificados, bosques aleatorios, inteligencia artificial.		
RESUMEN/ABSTRACT:	<p>La era digital se enfoca en las nuevas tecnologías, lo que implica profundos cambios y transformaciones en la sociedad en un mundo globalizado. Esto se manifiesta en una verdadera revolución tecnológica. La inteligencia artificial es parte de esta era, es la combinación de algoritmos propuestos para crear máquinas con capacidades similares a las de los humanos. Tiene aplicaciones en varios campos como la salud, las finanzas, el transporte y la educación. Machine Learning es un método de análisis de datos que crea automáticamente modelos analíticos. Esta es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con una mínima intervención humana. El propósito de este trabajo fue analizar la aplicabilidad de Machine Learning para clasificar a los clientes en función de las ventas totales y volumen de ventas en dólares representadas en cajas en la industria de conservas de atún. El análisis se desarrollará utilizando árboles clasificados y bosques aleatorios, mediante la recopilación de datos de la empresa TECOPESCA C.A.</p>		
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Teléfono: 0984435841	E-mail: cr11shianreyna@outlook.com Micaella.camacho99@gmail.com	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UIC):	Nombre: Román Bermeo, Cynthia Lizbeth		
	Teléfono: +593-984228698 Extensión:		
	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):			