



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

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

TÍTULO:

**Reestructuración de procesos en una empresa importadora y
distribuidora de fertilizantes a nivel nacional.**

AUTOR:

Mena Freire, Nicolás Humberto

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

TUTOR:

Ing. Carrera Buri, Félix Mgs.

**Guayaquil, Ecuador
19 de septiembre del 2022**



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

CERTIFICACIÓN

Certificamos que el presente trabajo de integración curricular fue realizado en su totalidad por **Mena Freire, Nicolás Humberto**, como requerimiento para la obtención del título de **Licenciado en Negocios Internacionales**.

TUTOR

f. _____
Ing. Carrera Buri, Félix Mgs.

DIRECTORA DE LA CARRERA

f. _____
Ing. Hurtado Cevallos, Gabriela Mgs.

Guayaquil, a los 19 del mes de septiembre del año 2022



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE ECONOMÍA Y EMPRESA

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Mena Freire, Nicolás Humberto**

DECLARO QUE:

El Trabajo de Integración Curricular, **Reestructuración de procesos en una empresa importadora y distribuidora de fertilizantes a nivel nacional**, previo a la obtención del título de **Licenciado en Negocios Internacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

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

Guayaquil, a los 19 del mes de septiembre del año 2022

EL AUTOR

f. _____

Mena Freire, Nicolás Humberto



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

AUTORIZACIÓN

Yo, **Mena Freire, Nicolás Humberto**

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Integración Curricular, **Reestructuración de procesos en una empresa importadora y distribuidora de fertilizantes a nivel nacional**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 19 del mes de septiembre del año 2022

EL AUTOR:

f. _____

Mena Freire, Nicolás Humberto



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

REPORTE URKUND

URKUND		Lista de fuentes	Bloques	Abrir
Documento	MENA FREIRE NICOLAS HUMBERTO.doc (D143842030)			
Presentado	2022-09-09 12:34 (-05:00)			
Presentado por	nicolas.mena2108@gmail.com			
Recibido	felix.carrera01.ucsg@analysis.orkund.com			
Mensaje	TESIS MENA FREIRE NICOLAS HUMBERTO Mostrar el mensaje completo			
	0% de estas 27 páginas, se componen de texto presente en 0 fuentes.			
		Categoría	Enlace/nombre de archivo	
			Universidad Católica de Santiago de Guayaquil / D143500810	
			Universidad Católica de Santiago de Guayaquil / D127738726	
			ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO / D132229401	
			Universidad Central de Ecuador / D58673869	
			Universidad Nacional de Moquegua / D142648347	
			Universidad Técnica Particular de Loja / D103516441	
			https://link.springer.com/article/10.1007/s43545-021-00237-z	

TUTOR

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

AGRADECIMIENTO

Quiero agradecer a de todo corazón a mis padres que a pesar de distintas circunstancias que se dieron durante este periodo universitario siempre estuvieron ahí para apoyarme y darme animo a seguir adelante y dar lo mejor de mí.

También agradecer a varios miembros de mi familia que me brindaron apoyo, lo cual es muy importante para mí, estuvieron pendientes de mi trayecto universitario hasta el final.

Adicional a esto quiero agradecer a mis compañeros que formaron parte de este trayecto al ser un apoyo en distintas situaciones.

Además, agradecerle completamente a mi tutor, Ing. Félix Carrera, que me guio y apoyo en este proceso de tesis.

DEDICATORIA

Quiero dedicar este proyecto a mi familia que estuvo siempre apoyándome, más que nada a mis padres que fueron mi motor y quienes siempre estuvieron para mí sin dudarlo brindándome un apoyo incondicional.

A mis amigos que me dieron consejos para seguir con este proyecto. Y estuvieron para mí desde el día uno de clases hasta el último día.



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

TRIBUNAL DE SUSTENTACIÓN

Gabriela Hurtado

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

María Soledad Rea F

f. _____
Ing. Rea Fajardo, María Soledad, Mgs.
COORDINADOR DEL ÁREA

Rosa Margarita Zumba

f. _____
Lcda. Zumba Córdova, Rosa Margarita, PHD
OPONENTE



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

CALIFICACIÓN

Mena Freire, Nicolás Humberto

ÍNDICE

RESUMEN	XII
ABSTRACT	XIII
RÉSUMÉ	XIV
CAPÍTULO 1	2
INTRODUCCIÓN.....	2
PROBLEMÁTICA	5
<i>Algoritmo de Procesos</i>	7
JUSTIFICACIÓN	7
ALCANCE	9
OBJETIVO GENERAL.....	10
OBJETIVOS ESPECÍFICOS	10
HIPÓTESIS	10
CAPÍTULO 2: MARCO TEÓRICO	10
MACHINE LEARNING.....	10
APRENDIZAJE SUPERVISADO	12
CLASIFICACIÓN.....	15
ÁRBOL DE DECISIÓN PARA CLASIFICAR	16
ENTROPÍA	17
GANANCIA DE INFORMACIÓN	18
ÍNDICE GINI.....	19
CHI- CUADRADO.....	19
BOSQUES ALEATORIOS	20
MARCO CONCEPTUAL	24
INTELIGENCIA ARTIFICIAL.....	24
BUSINESS INTELLIGENCE	26
MACHINE LEARNING.....	27
APRENDIZAJE SUPERVISADO	28
MARCO LEGAL	30
CAPÍTULO 3: METODOLOGÍA	32
OPERACIONALIZACION DE VARIABLES	32
PASOS POR REALIZAR	33
RESULTADOS	41
CONCLUSIÓN	44
RECOMENDACIONES	45
REFERENCIAS	46
ANEXOS	50

INDICE DE FIGURAS

FIGURA 1 ALGORITMO DE PROCESOS	7
FIGURA 2 ESQUEMA GENERAL DEL APRENDIZAJE SUPERVISADO (DEY, 2017)	13
FIGURA 3 FUNCIÓN PARA PREDECIR	14
FIGURA 4 APRENDIZAJE SUPERVISADO (SIMEONE, 2018)	14
FIGURA 5 ÁRBOL DE DECISIÓN	16
FIGURA 6 GANANCIA DE INFORMACIÓN	18
FIGURA 7 ECUACIÓN MATEMÁTICA DE CHI-CUADRADO.....	20
FIGURA 8 ÁRBOL DE DECISIÓN VS BOSQUE ALEATORIO (MERINO & CHACÓN, 2017)	23
FIGURA 9 CAPTURA DE R-STUDIO DEL INICIO DE CREACIÓN DEL MODELO.....	33
FIGURA 10 CAPTURA DE R-STUDIO/CAMBIO A VALORES NUMÉRICOS DE LAS VARIABLES A USAR	34
FIGURA 11 CAPTURA DE R-STUDIO MOSTRANDO LOS PAQUETES PARA ÁRBOL DE DECISIÓN	35
FIGURA 12 CAPTURA DE R-STUDIO, FUNCIÓN DE PLANTAR SEMILLA.....	35
FIGURA 13 CAPTURA DE R-STUDIO, CREACIÓN DE ENTRENAMIENTO DE DATOS ...	36
FIGURA 14 CAPTURA DE R-STUDIO, CREACIÓN DEL ÁRBOL DE DECISIÓN	36
FIGURA 15 CAPTURA DE RESULTADOS DEL ÁRBOL.....	36
FIGURA 16 CAPTURA DE R-STUDIO, FUNCIÓN DEL GRAFICO DEL ÁRBOL	37
FIGURA 17 CAPTURA DE GRAFICO DEL ÁRBOL	37
FIGURA 18 CAPTURA DE LAS LIBRERÍAS A USAR EN EL BOSQUE ALEATORIO	38
FIGURA 19 CAPTURA DE FUNCIÓN PARA REALIZAR EL BOSQUE ALEATORIO	39
FIGURA 20 CAPTURA DE LOS DATOS DEL BOSQUE ALEATORIO.....	39
FIGURA 21 CAPTURA DE FUNCIÓN PARA LA PREDICCIÓN.....	39
FIGURA 22 CAPTURA DE LA FUNCIÓN DE MATRIZ DE CONFUSIÓN	40
FIGURA 23 CAPTURA DE LA FUNCIÓN DE PROBABILIDAD DE ÉXITO.....	40
FIGURA 24 CAPTURA DE FUNCIÓN PARA OBSERVAR LOS VALORES CLASIFICADOS EN EL DATAFRAME.....	40
FIGURA 25 RESULTADOS DE LA MATRIZ	41
FIGURA 26 RESULTADOS DE LA PREDICCIÓN	42

RESUMEN

Con un mundo globalizado existen nuevas maneras y oportunidades a las cuales nos tenemos que adaptar, más que nada a las nuevas herramientas tecnológicas. Para las empresas es importante que realicen un análisis de estas para que así la toma de decisiones sea acertada a largo plazo, dando como resultados se puedan reestructurar procesos para que estos sean óptimos. En este caso, la inteligencia artificial es de los conceptos más llamativos e innovadores para las empresas en los últimos tiempos por el gran aporte que este genera en el ámbito empresarial.

Dentro de los conceptos de inteligencia artificial, tenemos al Machine Learning, este brinda los recursos necesarios para realizar un análisis de lo que una empresa necesite por medio del manejo de datos que esta posee. El objetivo que se tiene en esta investigación es el analizar la eficiencia de los vendedores de la empresa Quimasa S.A, por medio de distintos procesos de Machine Learning, ya que esta empresa no ha implementado este tipo de análisis. Se obtuvo los datos de ventas de la empresa del último año y se hizo uso de procesos de Aprendizaje Supervisado, los cuales fueron Árbol de Decisión y Bosques Aleatorios, para así dar a conocer a la empresa el rendimiento que hubo en las ventas que se realizaron y se pueda mejorar a largo plazo para que se cometan los mismos errores.

Palabras Claves: Machine Learning, Aprendizaje Supervisado, Árbol de decisión, Bosques Aleatorios, Inteligencia Artificial.

ABSTRACT

In a globalized world there are new ways and opportunities to which we must adapt, especially to new technological tools. It is important for companies to make an analysis of these tools to make the right decisions in the long term, resulting in the ability to restructure processes so that they are optimal. In this case, artificial intelligence is one of the most striking and innovative concepts for companies in recent times due to the great contribution it generates in the business environment.

Within the concepts of artificial intelligence, we have Machine Learning, which provides the necessary resources to perform an analysis of what a company needs through data management that it has. The objective of this research is to analyze the efficiency of the salesmen of Quimasa S.A., by means of different Machine Learning processes, since this company has not implemented this type of analysis. The sales data of the company for the last year was obtained and the use of Supervised Learning processes was made, which were Decision Tree and Random Forests, to let the company, know the performance of the sales that were made and to improve in the long term so that the same mistakes are not made.

Keywords : *Machine Learning, artificial intelligence, decision tree, random forest, supervised learning*

RÉSUMÉ

Dans un monde globalisé, il existe de nouveaux moyens et de nouvelles opportunités auxquels nous devons nous adapter, notamment aux nouveaux outils technologiques. Il est important pour les entreprises de les analyser afin de prendre les bonnes décisions à long terme, ce qui permet de restructurer les processus pour les rendre optimaux. Dans ce cas, l'intelligence artificielle est l'un des concepts les plus marquants et les plus novateurs pour les entreprises ces derniers temps, en raison de la grande contribution qu'elle génère dans l'environnement des affaires.

Dans les concepts d'intelligence artificielle, nous avons le Machine Learning, cela fournit les ressources nécessaires pour effectuer une analyse de ce dont une entreprise a besoin à travers la gestion des données qu'elle possède. L'objectif de cette présente recherche est d'analyser l'efficacité des vendeurs de l'entreprise Quimasa S.A., au moyen de différents processus de Machine Learning, car cette entreprise n'a pas mis en place ce type d'analyse. Les données de vente de l'entreprise pour l'année dernière ont été obtenues et les processus d'apprentissage supervisé ont été utilisés, à savoir l'arbre de décision et les forêts aléatoires, afin d'informer l'entreprise sur la performance des ventes réalisées et de s'améliorer sur le long terme pour ne pas reproduire les mêmes erreurs.

Mots-clés : Intelligence artificielle, apprentissage automatique, apprentissage supervise, arbre de décision, forêts aléatoires

Capítulo 1

Introducción

Mediante el avance de la era tecnológica, las empresas cada vez optan por tener nuevas metodologías en sus procesos, ya que, es importante conocer lo que saben y de qué manera se puede manejar ese conocimiento. Es por esto por lo que hoy en día las empresas están en búsqueda de una reestructuración de sus procesos para que de esta manera se pueda optimizar los que ya están implementados o cambiarlos y así tener un mejor desempeño en los mercados. (Landriscini, 2017)

Uno de los grandes avances tecnológicos es la inteligencia artificial, ya que esta se centra en el dominio del conocimiento relacionado con el lenguaje, la inteligencia, el razonamiento, el aprendizaje y la resolución de problemas. Convirtiéndose en parte muy importante para las empresas que desean adaptarse a nuevos modelos metodológicos. Para Russo (2020) y Ludermir (2021), dichas técnicas pueden resolver problemas que son cada vez más complejos, trayendo así eficiencia, sentido y agilidad. (Pinho, et al., 2022)

Dentro de las distintas formas en que las empresas podrían realizar una reestructuración en donde tengan una base clara de lo que está pasando y cómo manejarlo es por medio del concepto de machine learning, el cual a través de sus procesos podrá arrojar resultados confiables para que así se puedan tomar decisiones acertadas. (Athey & Imbens, 2019)

El efecto que tiene el Machine Learning es de vital importancia, este concepto por medio de sus procesos hace uso de los datos para realizar análisis de comportamientos y de alguna manera logra identificar tendencias futuras, es por esto por lo que mientras existan una gran cantidad de información, hay una gran posibilidad de conocer con anticipación cual será el comportamiento de un grupo determinado de personas. (Sandoval, 2019)

Los procesos en que se puede hacer uso de Machine Learning deben ser sectores que donde se logre generar algoritmos, puesto que estos

poseen la capacidad de aprender por si solos, sin tener la necesidad de estar constantemente programándolos, ya que, el Machine Learning es una rama de la Inteligencia Artificial. (Hansen, 2018)

Es importante saber cuándo una empresa debe considerar la posible reestructuración en alguno de los procesos que manejan, ya que, existen diversas empresas en el Ecuador las cuales son una base importante para el desarrollo y crecimiento de la economía, como es el caso de la empresa Quimasa S.A. Quimasa es una empresa importadora que produce, comercializa, formula y distribuye materias primas, en donde se encarga de atender al sector productivo del Ecuador, ya que su línea de negocios va direccionado a los productos agrícolas, alimenticia e industrial, productos plásticos y nutrición animal. Dado el gran alcance que tienen a nivel nacional e internacional, poseen su propia red de almacenes, facilitando su logística al poder transportar grandes cantidades de producto en óptimas condiciones hasta el lugar solicitado.

Todo esto se logra llevar a cabo por medio de distintos procesos los cuales han seguido usando durante sus años de funcionamiento y hasta la actualidad, en donde estos brindaron grandes aportes al crecimiento de la empresa, aunque hoy en día existan alternativas más fiables para llevar a cabo su funcionamiento. (Rebollo & Sánchez, 2012)

Fred R. David (2013) explica que todo proceso interno de la organización debe ser analizado y evaluado para determinar si es un proceso que requiere cambio y mejora o un cambio que se puede generar como tal en la empresa. Esto ha sido demostrado en los últimos años ya que las compañías han logrado entender que es fundamental realizar un ajuste de personal en caso de que no se esté logrando los objetivos establecidos y de esta manera se alcance una mayor productividad.

Las empresas en el Ecuador tienen cierto temor a reestructurar sus procesos al no querer lidiar con nuevos métodos o conceptos, sin importar si estos proveerán un mejor funcionamiento en ellas. Esto es parte

fundamental ya que, estas estructuras deben ser revisadas constantemente para que se pueda garantizar un mejor funcionamiento, pero cada vez son más las organizaciones que toman el riesgo de dar ese paso adquirir nuevos procesos dando como resultado importantes cambios y avances significativos dentro de la misma. Así habrá más empresas las cuales busquen invertir en estos conceptos como lo es el Machine Learning y manejen con mayor eficacia su información y puedan tomar las decisiones correctas para la reestructuración que vaya a ser aplicada en el área analizada. (Hansen, 2018)

Esto es necesario para manejar una empresa de la manera más adecuada y certera posible, como se puede observar a nivel mundial, compañías multinacionales optan por llevar a cabo reestructuraciones constantes gracias a la implementación de este concepto, puesto que, hacen análisis de cada área dentro de su organización obteniendo así información relevante el cual les permite actuar con tiempo y buscar soluciones de manera anticipada. De esta manera se diferenciarán de las demás al adaptarse cualquier tipo de ventaja innovadora que les puedan ofrecer como es uso del aprendizaje automático.

Estas grandes empresas a nivel mundial tienen como visión tener una mejora continua, ser más eficientes y eficaces en sus procesos, tener una mejor calidad logística hasta el punto de entrega al comprador. Con estos procesos mejor optimizados gracias al uso de estas herramientas tecnológicas, dan una ventaja competitiva considerable, la gestión de negocios mejora y se pueden observar pequeñas mejoras en un periodo de tiempo corto, lo que conlleva a que a largo plazo el rendimiento de estos procesos tendrá una gran mejora, permitiendo de esta manera análisis claros con la información que posee la empresa.

La implementación de estas nuevas tecnologías da la oportunidad de encontrar modelos que aporten a los procesos administrativos y así exista un incremento a la posibilidad de tener éxito a largo plazo.

Problemática

El problema se da puesto que la empresa tiene ciertas dificultades a la hora de llevar sus procesos en el área de ventas al no tener el control adecuado de su base de datos además de no tener claro que tan eficientes son los vendedores de la empresa. Esto se da por no tener las herramientas necesarias y hacer uso de procesos tradicionales los cuales logran aportar poca organización llevando como consecuencia el no tener una forma de analizar datos y tomar decisiones.

Es importante analizar el rendimiento de los trabajadores del área de ventas ya que esto es parte importante de formas adecuadas para un buen manejo administrativo, el cual permitirá al área de importaciones tener una mejor proyección de los productos que serán necesarios en cierta cantidad de días, así como también, en el área operativa manejarán de mejor manera el producto que sea requerido para distribuir a lo largo del país.

Existe una cantidad considerable de empresas a nivel nacional que siguen haciendo uso de la herramienta Excel para el manejo de sus datos. Esta pueda aportar si llegase a ser una compañía pequeña o también como las microempresas en donde su base de datos no es extensa, pero para empresas que grandes en donde hay un gran manejo de datos, este proceso tradicional ya no es el más adecuado.

Considerando el proceso que se lleva a cabo a través de métodos como es el uso de Excel para la organización de sus datos y con estos pasar a analizarlos. Esto puede llegar a provocar una discordancia en las ventas que se hagan llevando así un efecto en cadena en donde no habrá conformidad de parte de las distintas áreas de la empresa que puedan ser afectadas, principalmente el área operativa. Es por esto por lo que hoy en día existen nuevas formas, programas, que permitan manejar grandes cantidades de datos para que estos se puedan analizar de forma adecuada reduciendo de manera considerable las probabilidades de error en el estudio de la información.

Se puede apreciar que es de vital importancia realizar una reestructuración en sus procesos en donde estos permitan conocer el desempeño que tiene el trabajador con respecto a los objetivos planteados de la empresa esto se da con el manejo adecuado de todos los datos que obtiene la organización. Mediante la implementación de Machine Learning, facilitara la automatización de los procesos como también la toma de decisiones basado en la información que se posee. Haciendo que con el tiempo esta aprenda de manera automática y logre integrarse al desarrollo de la empresa adaptándose a los cambios que existan mediante la alimentación constante de los datos que existan.

En la actualidad este concepto tiene gran importancia en el sector administrativo al tener procesos y herramientas en donde su propósito principal es mejorar los objetivos de la empresa a largo plazo, ofreciendo una forma de análisis predictivo, esto quiere decir que no solo toma en cuenta lo que haya pasado, sino que hace predicciones de que es lo que podría pasar.

Una reestructuración de procesos tendrá éxito si es realizada con la ayuda de estas nuevas herramientas tecnológicas las cuales aportan grandes cambios y mejoras. Para las empresas es importante invertir en este tipo de reestructuras, si desean tener un funcionamiento acorde a sus objetivos planteados con oportunidades de crecimiento constante, con el buen manejo y análisis de su información.

Algoritmo de Procesos

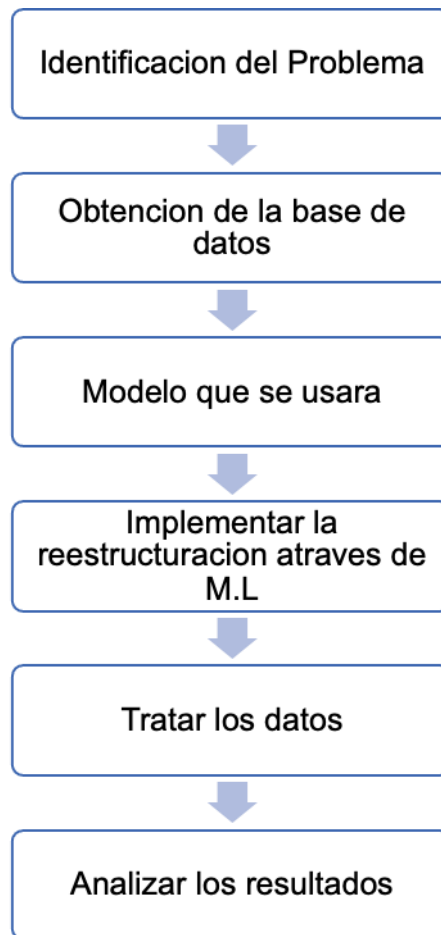


Figura 1 Algoritmo de Procesos

Justificación

El realizar esta investigación muestra lo importante que es llevar a cabo una reestructura en los procesos de las empresas para que haya un control y buen manejo de datos, puesto que, se suele dar poca relevancia al tratado de información que posee la empresa, dando como resultado un equipo de ventas que no esté rindiendo de la manera adecuada.

Es de vital relevancia que se conozca acerca de estos temas, ya que, para las personas que están por iniciar un negocio o las que tiene ya un negocio estructurado, sepan que hay que aprovechar del avance tecnológico que existe para lograr obtener una ventaja ante los demás. En el Ecuador las empresas siguen haciendo uso para el manejo de datos herramientas de

Excel, abriendo paso a una gran probabilidad de que existan errores comunes que pueden cometer las personas.

El hacer uso de las herramientas que posee el Machine Learning proporcionara un gran beneficio a las empresas que busquen reestructurarse al tener resultados significativos en las distintas áreas donde lo implementen y sean necesarios. Como seria el área de ventas, al implementar estas herramientas que nos brinda el M.L, dará paso tener un manejo adecuado del control de los trabajadores en dicha área de esta manera se podrá analizar lo que se haga sobre las ventas y así tomar decisiones acertadas referente a las necesidades del momento. (Rabbi, 2019)

Reestructurar procesos dentro de una empresa suele ser muy complicado al no tener clara la manera en que se implementarían los cambios, el uso de machine Learning permite encontrar patrones a través de una extensa cantidad de datos, aportan un alto valor operacional permitiendo así una mejora en la postura de las empresas. El diseño experimental aplicado a la toma de decisiones aumenta las posibilidades de éxito. (Gómez, et al., 2020)

De esta manera se espera poder tomar decisiones las cuales puedan aportar a la empresa opciones de mejorar en sus procesos, como es el caso de tener un control más claro sobre la eficiencia que hay en los trabajadores del área de ventas y saber si están cumpliendo con los objetivos y expectativas planteadas por parte de la compañía. A través de los algoritmos que usan las metodologías dentro del Machine Learning permiten que se pueda clasificar y analizar el problema planteado.

A través de la implementación del Machine Learning se podrá tomar decisiones acorde al rendimiento que hay en el área de ventas por el análisis predictivo que posee este concepto dentro de sus procesos, ya que con el uso de la gran cantidad de datos que hay en la empresa, se podrá programar para que la maquina aprenda y arroje resultados claros acerca de la eficiencia que hay, así tomaran acción ante los datos presentados y se

pueda dar una reestructuración para que el área sea más eficaz en sus procesos. (Rabbi, 2019)

Con esta investigación se podrá lograr grandes mejoras en una empresa, ya sea a corto o largo plazo, el implementar estas nuevas tecnológicas en el ámbito empresarial es de vital importancia, ya que, mostrara datos que las personas no siempre podrán tener claro si hacen uso de métodos tradicionales que para la toma de decisiones se convierten en métodos anticuados y poco precisos para realizar un análisis respectivo de los problemas que se planteen.

Alcance

A través de esta investigación se espera llegar a las empresas del Ecuador, para que tengan conocimiento de lo importante que es el manejo de sus procesos en cuanto al desempeño de sus empleadores, este caso en el área de ventas, que no solo almacenen datos, y sean tratados de la forma tradicional, ya que, de toda esa información con el uso de las herramientas que posee el Machine Learning se puede tener una perspectiva precisa de lo que está ocurriendo en el área en el que se haga uso de este concepto. Permitiendo así un análisis eficaz el cual permitirá tomar acciones necesarias para que haya un rendimiento favorable para la empresa.

Como el Machine Learning trabaja por medio de la Inteligencia Artificial esta nos da un enfoque a futuro en donde se podrá tomar decisiones por parte de los gerentes de área, permitiendo así conocer que tan bien está rindiendo cada empleado y su cumple lo esperado dando como resultado un margen de error mínimo ya que se pueden detectar con tiempo ineficiencias, de esta manera se podrá optimizar a futuro el rendimiento de los empleados y si es necesario la búsqueda de nuevos empleados que puedan cumplir con las expectativas que tiene la empresa.

La aplicación de este concepto tiene cada vez más acogida por diferentes empresas a nivel nacional, ya que esto permite tener procesos óptimos y claros. Como es el caso de muchas compañías grandes a nivel

mundial que tratan todos sus datos obtenidos para hacer uso de las herramientas que brinda el Machine Learning para el análisis de patrones de los compradores, así como también análisis de rendimiento de las áreas de la compañía, el cuales les permita tomar decisiones acertadas.

Objetivo General

Analizar el rendimiento de los empleados del área de ventas de la empresa Quimasa S.A a través de la aplicación de conceptos de Machine Learning.

Objetivos Específicos

- (a) Analizar el marco teórico y conceptual mediante la revisión de literatura basada en la recopilación de artículos científicos.
- (b) Analizar la metodología resultante de la revisión de la literatura para la aplicación de modelos análisis supervisado
- (c) Analizar los resultados obtenidos y conocer la eficiencia de los trabajadores del área de ventas.

Hipótesis

A través de los procesos de clasificación que tiene el Machine Learning y no por medios estadísticos convencionales, se puede analizar el desempeño de parte de los trabajadores del área de ventas.

Capítulo 2: Marco Teórico

Para tener el conocimiento adecuado de lo que es el Machine Learning y el gran aporte que le brinda a las empresas, se debe revisar las teorías que abarca este concepto, así como también lo que es la Inteligencia Artificial para la elaboración de esta investigación, dando a conocer el concepto del Machine Learning y sus herramientas que serán usadas como lo es el Árbol de decisión para clasificar, bosques aleatorios.

Machine Learning

El aprendizaje automático es un área relacionada tanto con la cibernética como con la informática, y recientemente ha recibido un gran

interés tanto por parte de los profesionales como del público en general. El origen del aprendizaje automático en el sentido moderno está más relacionado con el nombre del psicólogo de la Universidad de Cornell Frank Rosenblatt. Inició un grupo para construir máquinas que reconozcan letras en el alfabeto Rosenblatt, basándose en ideas sobre cómo funciona el sistema nervioso humano. (Fradkov, 2020)

Con el pasar de los años esto tuvo cada vez más presencia en distintos sectores a nivel mundial, pero aun no siendo considerado de gran utilidad puesto que no se esperaba que una maquina pudiera hacer mejor el trabajo del ser humano, lo cual ha cambiado considerablemente gracias al avance tecnológico que se ha dado.

Ya que la revolución de la tecnología de la información y comunicación está provocando que cada vez haya más acogida a los términos como la Inteligencia Artificial, Machine Learning o ciencia de datos, Big Data, dado que sus algoritmos son capaces de imitar el comportamiento que tiene el ser humano como es la capacidad de reconocer imágenes, el orden, clasificar objetos por sus características, etc. (Álvarez, et al., 2019)

En la última década grandes empresas como Amazon, Microsoft, IBM, Google, etc. Empezaron a dar un paso a la implementación de Machine Learning ya que observaron el gran potencial que este concepto tenía, dando así un gran avance en la era tecnológica. Como ya es de conocimiento, las empresas de gran escala poseen grandes cantidades de datos digitalizados lo cual lleva una complejidad alta para el ser humano en poder realizar análisis de patrones de estos. Pero a través de los algoritmos del Machine Learning se logra conseguir estos patrones automáticamente. (Guadalupe, et al., 2021)

El concepto que tiene el Machine Learning es que a través de la aplicación de algoritmos se crea la capacidad de aprender por medio de grandes cantidades de datos construyendo así algoritmos con un rendimiento más alto que mejoran de forma automática. Esto permite que los

métodos detecten automáticamente patrones en la información que se posee, de esta manera son utilizados para predecir datos futuros en entornos inciertos. (Management Solutions, 2021)

Para las empresas que desean tener un cambio a largo plazo en donde haya un entorno laboral adecuado a lo que se espera es importante que hagan uso de estas nuevas formas tecnológicas que cada vez revolucionan más las industrias, con el concepto del Machine Learning se puede analizar el desempeño de un área de la empresa en el caso de que se tengan las variables y datos necesarios, dando así la opción de tomar decisiones certeras para la mejora del área que por ende dará una mejora a la empresa.

Esta forma innovadora aporta grandes beneficios a las empresas como se puede apreciar en las grandes empresas, al ser referencia del aprendizaje automático que tiene las máquinas aportan importantes soluciones donde se pueda evolucionar y aprender de los datos que se poseen, puede mejorar el servicio al cliente, analizando las preferencias de los compradores para así ofrecer un tipo de producto adecuado a sus preferencias. Hay una gran disminución de errores, ya que mientras más tiempo se haga uso de este concepto, menor será la cantidad de errores que se den en la empresa. Esto facilita totalmente el análisis de datos para obtener mejoras, además de que aporta ventajas competitivas aseguradas.

El Machine Learning tiene diferentes técnicas que pueden ser utilizadas ya que, tiene dos enfoques fundamentales uno de ellos es el caso del Aprendizaje Supervisado el cual parte a tener herramientas importantes en donde esta nos permite realizar el proceso y análisis de la eficiencia de empleadores en una empresa como puede ser en el sector de ventas al tener la cantidad de datos y las variables necesarias para hacer un estudio.

Aprendizaje Supervisado

Dentro del Machine Learning tenemos al Aprendizaje Supervisado el cual al algoritmo se le enseña cómo ordenar un conjunto de datos bajo una

idea o idea en particular para encontrar patrones que se puedan aplicar al análisis y generar resultados conocidos, con esto se refiere a que, quien esté realizando o haciendo uso de este método ya tiene conocimiento de la variable de respuesta. (Rojas, 2020)

Este modelo es el más utilizado ya que posee pares de entrada y salida, es por esto por lo que es implementado cuando se necesita hacer una predicción clara por medio de nuevas entradas que el modelo no las haya visto la cual permita tomar decisiones acertadas. Además, este opera a través del entrenamiento de datos, siempre debe pasar por este entrenamiento para que haya buenos resultados, así se realiza el ajuste del modelo.

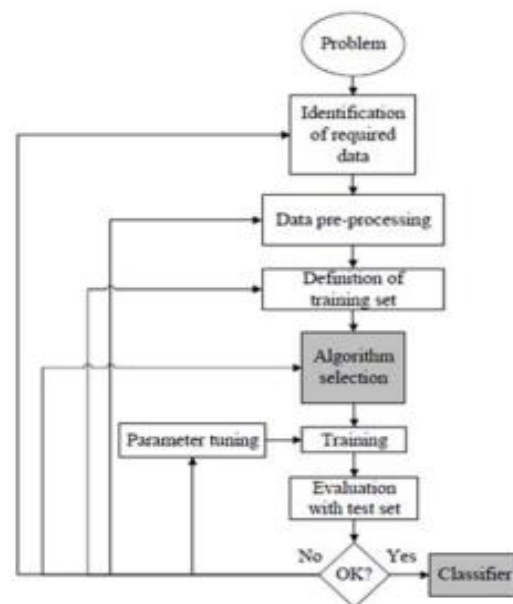


Figura 2 Esquema general del Aprendizaje Supervisado (Dey, 2017)

Para las empresas este tipo de aprendizaje suele ser el más utilizado ya que está mucho más definido, además, el modelo otorga al algoritmo la experiencia que necesita para así realizar clasificaciones o predicciones en datos que son desconocidos, haciéndolo así de gran importancia al aportar una optimización al desempeño de los procesos. Como por ejemplo se identificar el uso de este método al poder anticipar en qué momento es probable que se hagan transacciones fraudulentas con una tarjeta de crédito. (Machine Learning, 2019)

El objetivo fundamental que tiene es el de poder crear funciones que tengan la posibilidad de predecir valores que correspondan a objetos que sean de entrada luego de haberse adaptado con una gran cantidad de ejemplos que son los datos más importantes llamadas datos de entrenamiento.

X – Conjunto de objetos

Y – Conjunto de etiquetas o respuestas

$\gamma: X \rightarrow Y$ – Valores conocidos de la función

$$\gamma^i = \gamma(x^{(i)}), i = 1, \dots, l$$

Figura 3 Función para predecir

El problema vendría a ser el poder encontrar la hipótesis ya que, esta es la que permite acercar el conjunto de objetos al de las respuestas. Para eso nos ayuda el entrenamiento a través de su algoritmo da la posibilidad de crear una hipótesis. Y la prueba, que es donde la hipótesis que se obtuvo genera nuevas predicciones con objetos nuevos.

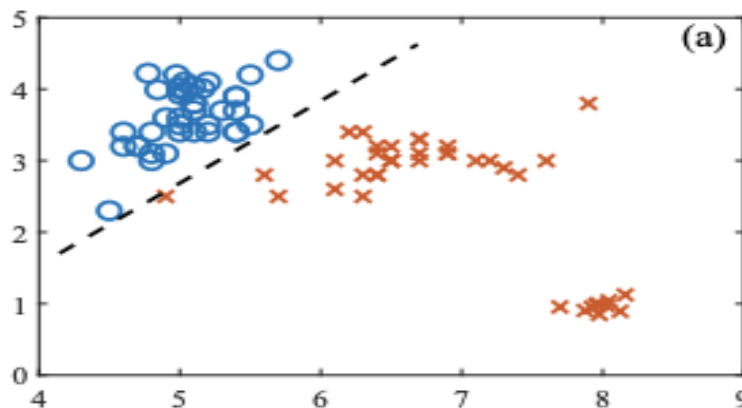


Figura 4 Aprendizaje Supervisado (Simeone, 2018)

En la figura 4 se puede observar un ejemplo de cómo trabaja el conjunto de entrenamiento de los datos con los puertos de entrada y de salida. Estos tienen la finalidad de aprender un mapeo entre los espacios de ambos puertos. En donde los datos de entrada vendrían a ser puntos en el plano bidimensional y los datos de salida son las etiquetas asignadas a cada

entrada, con el objetivo de aprender a clasificar de forma binaria. (Simeone, 2018)

El Aprendizaje Supervisado en función de si los datos de salida son variables continuas o discretas, entran dos subcategorías muy importantes en este modelo. En el caso de que estas variables sean continuas estas se refieren a una regresión, mientras que, si las variables son discretas o también conocidas como categóricas, sería denominado como una clasificación.

Clasificación

En el caso de la clasificación el objetivo de esta es la de predecir las etiquetas que son categóricas de los nuevos datos que se obtengan en base a las observaciones de datos pasados. Depende del tipo de etiqueta se puede determinar si la clasificación es binaria o de multiclase.

La clasificación binaria es en donde solo se puede asignar dos clases diferentes (0,1), esto quiere decir que “0” es “No”, mientras que “1” vendría a ser “Si”. Esto se puede encontrar en la detección de spam que se usan en los correos electrónicos. Y cuando nos referimos a una clasificación multiclase quiere decir que a estas se les otorga múltiples categorías a las variables. (Román, 2019)

Se debe considerar que la variable de salida que hay en los algoritmos de clasificación debe ser categórica, no un valor continuo como el sería el caso de una regresión. Dentro de las empresas es de gran ayuda este tipo de modelo al ser utilizado para la búsqueda de documentos que sean similares, análisis de opiniones, detección de fraudes además de identificación biométrica. Como se puede observar por medio de la clasificación se podrían resolver problemas que para el ser humano son mucho más complejos.

Además, podemos encontrar otro modelo de clasificación muy conocido es el modelo multi-etiqueta. Esto difiere de los dos modelos

anteriores debido al método de etiquetado específico en el que el Algoritmo 1 puede etiquetar datos en diferentes grupos al mismo tiempo al etiquetar múltiples características. Un elemento que comparte otros datos previamente almacenados.

Árbol de Decisión para Clasificar

El árbol de decisión es uno de los algoritmos más utilizados del Machine Learning, ya que, puede tener un análisis de fácil comprensión para las empresas que hagan uso de este método. Se podría decir que este es un diagrama en el cual cada rama tiene una pregunta de respuesta binaria, en donde se va desplazando por sus ramas dependiendo la respuesta que se dio hasta llegar a la solución del problema que se tenía inicialmente. (Mahesh, 2018)

Este método posee una estructura bastante similar a lo que es un diagrama de flujo en donde un nodo interno es quien representa un tipo de característica o también un atributo el cual se va a clasificar y cada rama representa un valor que el nodo mencionado puede tomar.

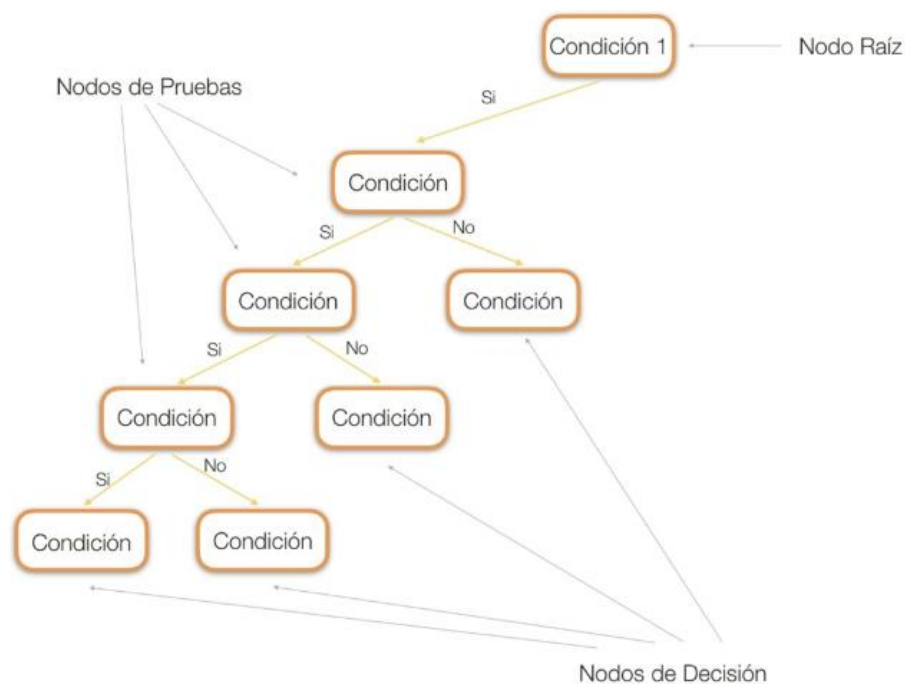


Figura 5 Árbol de Decisión

Como se puede observar en la figura 5, el árbol empieza con un nodo raíz el cual es en donde se encuentra todo el data set completo, esto quiere decir que son todos los datos que se tiene los cuales van a ser analizados. A través de sus algoritmos logra encontrar el mejor atributo lo cual lleva a que el nodo raíz se divida en subconjuntos donde posiblemente contiene los mejores valores para los atributos. Se generan nuevos sub-nodos y se repite este proceso hasta que el árbol de decisión esté listo, con esto se refiere a que ya no habría posibilidad de que la maquina pueda seguir clasificando nuevos nodos.

En el modelo se hace uso de un algoritmo llamado "CART" (Classification and Regression Tree Algorithm), Esto funciona explorando diferentes formas de dividir o subdividir localmente los datos en función de diferentes combinaciones de valores y predictores. (CART, 2022)

Para los gerentes que requieran conocer en un sector específico de su empresa el desempeño de ciertos trabajadores, esta es la mejor manera de hacerlo, ya que, estos algoritmos de clasificación son los indicados además de que, el árbol y el diseño que este posee son visualmente fáciles de entender o interpretar, lo cual permitirá que con la información obtenida del árbol se pueda hacer un análisis y así tomar decisiones en el caso de que fuese necesario para que la empresa siga creciendo.

Para la poder resolver la selección de atributos que hay dentro del árbol de decisión se dieron como opciones viables para la elaboración del árbol algunos criterios como son: Entropía, Ganancia de Información, Índice Gini y Chi-Cuadrado. Es importante conocer que estos métodos forman parte del árbol de decisión para clasificar.

Entropía

La Entropía es una medida la cual es aplicada para cuantificar el desorden que hay en un sistema. En el caso de que el nodo sea puro, su Entropía es de 0, esto quiere decir que solo tiene un tipo de observaciones. Y si es de igual a 1, la misma frecuencia existe para cada clase de observación. Este tiende a crear nodos que estén balanceados en el número de observaciones. La ecuación de la Entropía es la siguiente:

$$H = - \sum_{i=1}^n P_i * \log_2 P_i$$

Una de las grandes ventajas que tiene el árbol por clasificación con relación a las técnicas tradicionales de análisis de datos multivariantes, es el gran comportamiento que presenta ante ciertas situaciones de estructura discriminante que están bastante alejadas a la linealidad. (García, 2020)

Ganancia de Información

La ganancia de información es una propiedad estadística la cual se encarga de medir qué tan bien un atributo en particular separa una muestra de entrenamiento en función de una clasificación objetivo. El propósito de construir un árbol de decisión es poder encontrar los atributos que proporcionen la mayor ganancia de información y la menor entropía.

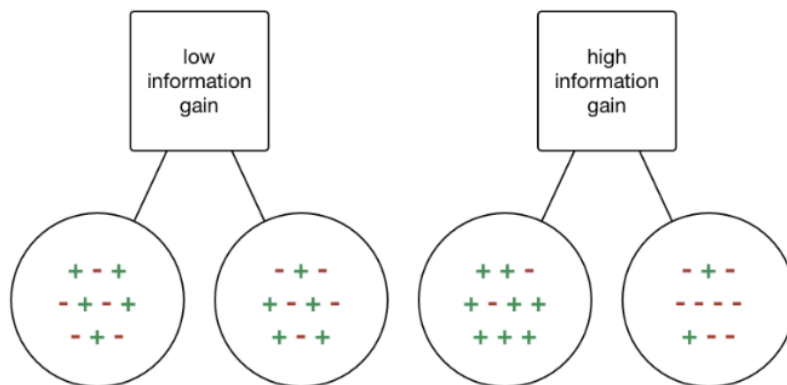


Figura 6 Ganancia de Información

Este calcula la diferencia que hay entre la entropía antes de la división y el promedio después de la división que se haya dado en el conjunto de datos en función de los valores de atributo otorgados. Esta se representa matemáticamente de la siguiente manera:

$$Information\ Gain = Entropy(before) - \sum_{j=1}^K Entropy(j, after)$$

En esta ecuación el “before” es el conjunto de datos antes de la división, “K” es el número de subconjuntos que fueron generados por la división que hubo y (j, after) vendría a ser el subconjunto “j” después de que se haya realizado la división.

Índice Gini

La manera de selección de variable para obtener la mejor división se puede considerar el uso del índice Gini (rpart), error de clasificación o la entropía. El coeficiente de Gini mide la limpieza del nodo. Mide la probabilidad de que dos registros de la misma clase no se eliminen del nodo. A mayor índice de Gini menor es la pureza que tiene el nodo. La ecuación que mide el Gini es la siguiente:

$$GINI(t) = 1 - \sum_{i=1}^n (P_i)^2$$

Donde P_i vendría a ser la probabilidad de que un ejemplo sea de tipo 1. Se debe tomar en cuenta que la variable aquí debe ser categórica o binarias, aquí es donde se hace uso de CART, para que se pueda dar esa división binaria. (Chauhan, 2022)

Chi- Cuadrado

Parte importante del Árbol de Decisión para la clasificación es el Chi-Cuadrado ya que este es un algoritmo para encontrar la significancia estadística de las diferencias que hay entre un sub-nodo y un nodo raíz. Funciona con la variable categórica “éxito” y “fracaso”, es importante conocer que mientras mayor sea el valor del Chi-Cuadrado, mayor será su significancia. Este se mide a través de la suma de sus cuadrados. Genera un árbol llamado CHAID (Chi-square Automatic Interaction Detector), a continuación, se podrá observar la ecuación del Chi-Cuadrado:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Where:

χ^2 = Chi Square obtained
 \sum = the sum of
 O = observed score
 E = expected score

Figura 7 Ecuación matemática de Chi-Cuadrado

La forma en que se calcula es para un nodo individual en donde se calcula la desviación de éxito y fracaso que nos los datos obtenidos. (Decisión, 2022)

Existen casos en los que puede darse un sobre ajuste de en el árbol, ya que, este parece que memoriza en su totalidad el conjunto de datos de entrenamiento y si no llegase a haber un límite establecido podrá dar un 100% de precisión, lo cual afecta drásticamente la precisión a predecir muestras que no son parte de los datos de entrenamiento. Para evitar esto, se debe hacer uso de los Bosques Aleatorios (Random Forest).

Bosques Aleatorios

Cuando nos referimos a los boques aleatorios, estos pueden traer grandes beneficios al ser implementados en una empresa, por la creación de varios árboles conlleva a que las predicciones finales sean mucho más sofisticadas. Un ejemplo del algoritmo es que este puede ser usado para identificar en qué momento va a existir el riesgo de que un cliente vaya a dejar la compañía. Así como también se puede analizar que clientes son los que van a tener la mayor cantidad de gastos generados en el año, esto se logra a través de un análisis de una gran cantidad de variables y atributos. (Bautista, 2020)

Esta hace una técnica de clasificación supervisada de datos en donde se forma una combinación de los árboles predictivos, donde se trabaja con

dichos árboles in-correlacionados y los promedia. Esto nos muestra que cada árbol es dependiente de los valores de un vector aleatorio de la muestra de forma independiente y con una misma distribución de todos los árboles que se tenga en el bosque. El error de generalización que se puede encontrar el bosque de aleatorio de clasificación depende de la fuerza que hay dentro de cada árbol que conforma el bosque y la correlación que existe entre ellos. (Merino & Chacón, 2017)

Todos estos métodos tienen en común que se genera un vector aleatorio θ_k para el k-ésimo árbol, independientemente de los últimos vectores aleatorios $\theta_1, \dots, \theta_{k-1}$, pero la distribución es la misma. Luego, el árbol se desarrolla utilizando el conjunto de entrenamiento y de θ_k y se convierte en un clasificador. Donde $h(x, \theta_k)$ es el vector de entrada.

Al basarse este en un conjunto de árboles de decisión, quiere decir que, este es sometida a una prueba binario en cada nodo, conocido como "Split", hasta llegar a una hoja en donde se obtenga una respuesta. Esta es una técnica la cual es utilizada para dividir un problema que tiene un grado de complejidad alta en un conjunto de problemas que sean más simples de analizar.

Para poder crear los datos de entrenamiento es importante definir la cantidad de árboles que habrá dentro de nuestro bosque. En la etapa de entrenamiento se intenta optimizar los parámetros de las funciones "Split" a partir de las muestras de entrenamiento.

$$\theta_k^* = \operatorname{argmax}_{\theta} \sum_{j \in \tau_j} I_j$$

También entra lo que es la ganancia de información en donde se hace uso de la siguiente función:

$$I_j = H(j) - \sum_{i \in 1,2} \frac{|S_j^i|}{|S_j|} H(S_j^i)$$

Donde “S” representa el conjunto de muestras que existen en el nodo que se está dividiendo y “Sⁱ” son los dos conjuntos creados a partir de la división. Aquí se mide la entropía y depende mucho del problema el cual se está analizando.

Se puede definir al bosque como un clasificador el cual consiste en lo que es la colección de clasificadores estructurada de los árboles {h(x, Θ_k), k=1, ...}, en donde {Θ_k} vendrían a ser vectores aleatorios los cuales fueron distribuidos de manera idéntica. Al tener un conjunto de clasificadores, h₁(x), h₂(x), ..., h_k(x), con los datos de entrenamiento se extrae al azar con dicha distribución aleatoria del vector. Esta función de margen está representada de la siguiente forma:

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j)$$

Donde la “I”, es la función de ganancia de información. Es importante conocer que mientras más alto sea el margen, más confianza habrá en la clasificación. La generalización del error viene dada en la función:

$$PE^* = P_{X,Y}(mg(X, Y) < 0)$$

Los subíndices que podemos observar en la función, que son “X, Y” nos muestra que la probabilidad esta sobre el espacio de “X, Y”. Para muchos árboles, hay convergencia de errores. Esto se demuestra por las "muchas leyes fuertes" y la estructura del árbol mismo.

Con los datos listos para el entrenamiento se debe tomar cada subconjunto y de forma recurrente realizar la partición de datos, en donde esto se requiere que sea seleccionado de forma aleatoria pero solo una

parte. Que se podrá observar que es muy parecido a lo que hace el árbol de decisión.

Con el conjunto de datos ya entrenado es mucho más fácil realizar predicciones, ya que se introduce el nuevo dato que se tenga a cada árbol y se realiza una clasificación individual. Dentro de este proceso se hace presente el “bootstrapping”, así como también la agregación de los resultados los cuales permiten realizar la predicción. La combinación de estos dos conceptos es conocida como “bagging”, que es conocido como el nombre genérico que tienen los algoritmos como son los Bosques Aleatorios y otros similares. (Sotaquirá, 2021)

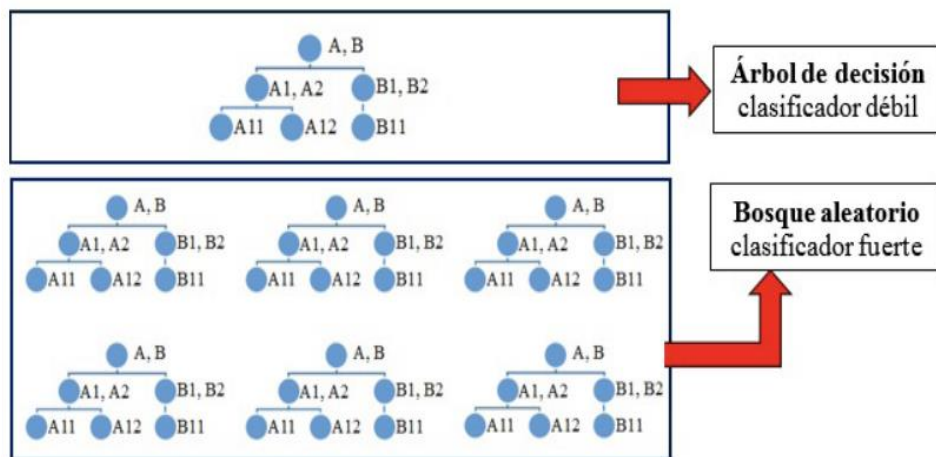


Figura 8 Árbol de decisión vs Bosque Aleatorio (Merino & Chacón, 2017)

Como se puede observar en la Ilustración 8, al tener una gran cantidad de árboles en las cuales existen diferentes significancias entre los sub-nodos hace que el bosque aleatorio sea una herramienta mucho más precisa el momento de predecir. Además, a diferencia del árbol este evita que exista algún tipo de sobreajuste al construir los árboles que sean de diferente tamaño partiendo de los subconjuntos y la combinación de resultados que se presenten. (Medina & Atehortúa, 2019)

Si las empresas optan por realizar este tipo de procedimientos de forma tradicional les tomaría una gran cantidad de tiempo, así como también habrá una gran probabilidad de errores ya que se requiere una limpieza de datos. Pero con el Bosque Aleatorio se logra minimizar esto al tratar los

datos faltantes de una manera adecuada. También por medio de las técnicas que posee este funciona para equilibrar los errores que hay en poblaciones y otro tipo de conjunto de datos los cuales no están bien equilibrados.

El bosque aleatorio claramente puede ayudar a clasificar el rendimiento que tengan los trabajadores de una empresa, como podría ser en el sector de ventas, al tener las variables necesarias para el análisis, se podría a través de esta metodología conocer si están cumpliendo con los objetivos que tiene la empresa, o si su rendimiento bajo con el pasar del tiempo y no han logrado superar esa caída. Todo esto permitirá que se pueda tomar decisiones importantes para que la empresa siga en dirección creciente.

Marco Conceptual

Inteligencia Artificial

Desde hace mucho tiempo atrás se le ha dado a la inteligencia artificial (IA) distintas definiciones las cuales están ligadas a procesos mentales, así como también al razonamiento, la conducta e incluso en la medición de la fidelidad del ser humano en su forma de actuar. Pero hay conceptos la cual la definen como racionalidad, es importante conocer que un sistema es racional cuando este hace lo “correcto”, en base a su conocimiento.

Con estas distintas definiciones que se le da, existen dos formas las cuales definen este término, una lo señala como una herramienta que se encarga de resolver tareas que sean muy específicas y que estas sean casi imposibles realizarlas para los seres humanos. La otra forma de definirla es como un sistema que imita la inteligencia humana, así como también los procesos cognitivos, la manera de interpretar, de aprender. (Enholm, et al., 2021)

Las tecnologías basadas en la inteligencia artificial están siendo utilizadas para dar un gran beneficio a los seres humanos ya que estas están aportando beneficios significativos el cual les brinda una mayor eficiencia en las diferentes tareas o situaciones donde se aplique esta

tecnología. En la actualidad eso está siendo aplicado de distintas formas como es el reconocimiento de imágenes, así como también el poder clasificar y etiquetar información relevante, se hacen muy presentes también en las mejoras estratégicas algorítmicas comerciales, mantenimiento predictivo, entre muchas formas más de ser aplicada.

Esta tecnología también tiene la capacidad de ofrecernos sugerencias y predicciones que tengan relación con temas de salud, impacto en áreas de una empresa, la educación e incluso en nuestras relaciones interpersonales. Una de las partes más importantes la manera en que está cambiando la forma de hacer negocios ya que esta aporta una gran ventaja competitiva para las empresas que opten por implementar diferentes conceptos o métodos de la inteligencia artificial. (Rouhiainen, 2018)

La llegada de la inteligencia artificial en los negocios es muy importante puesto que se espera que las empresas hagan uso de las aplicaciones que brinda IA se den cuenta de los beneficios del valor comercial, con esto se refiere a oportunidad de tener mayores ingresos, que los costos sean reducidos y haya una mayor eficiencia comercial.

En un estudio que fue realizado por parte de MIT Sloan Management Review encontró que más del 80 % de las organizaciones ven la IA como una gran oportunidad estratégica, y casi el 85 % ven la IA como una forma en la que se puede lograr una ventaja competitiva (Ransbotham, et al., 2017). Esta ventaja es la que está provocando que las empresas inviertan en las tecnológicas que tiene la inteligencia artificial. (Enholm, et al., 2021)

Para que las organizaciones puedan darle una buena acogida a la implementación de técnicas de la inteligencia artificial, es importante que el líder de la empresa exprese las ventajas que esa otorgaría, así como también el beneficio que les puede dar a largo plazo, ya que, cuando se escucha hablar de este término, suelen sentir miedo al pensar que este los dejara sin su puesto de trabajo lo cual provoca que haya un rechazo por parte de los trabajadores cuando se habla de este tema. (Fountaine, et al., 2019)

Hay que considerar que, así como la aplicación de herramientas de la inteligencia artificial podrá dar grandes resultados a largo plazo, también el comienzo de todo esto es un conjunto de barreras y desafíos. Al unir el

conocimiento entre dominios los cuales sean para desarrollar modelos precisos y significativos, manejo y limpieza de datos, identificación de problemas, son de las muchas formas en que trabaja esta tecnología.

Business Intelligence

La inteligencia de negocios se entiende como un conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas en crear y administrar información que permita a los usuarios de la empresa tomar mejores decisiones. Es importante que las organizaciones hagan uso de este tipo de metodologías para así llevar y tratar de una manera eficiente los datos, permitiéndoles ser competitivos en el mercado al realizar un manejo eficiente de la información, ya sea interna o externa. (Mamani, 2018)

La integración de inteligencia empresarial hace uso de los datos que posee la organización para impulsar un cambio, como puede ser la disminución notable de ineficiencias, lo que le permitiría adaptarse a los cambios que se den dentro del mercado o la demanda. Este término incluye los procesos y métodos para la recopilación, almacenamiento y análisis de datos que se generan por las actividades operacionales de un negocio dando como resultado la optimización del rendimiento en sus procesos.

Muchas empresas aun no toman la iniciativa de hacer uso de estas metodologías, pero en los últimos años se ha podido observar mayor acogida no solo para el tema de negocios, sino también en sectores de salud, educación, servicios financieros las cuales están aprovechando ya que permite la mejor toma de decisiones al ofrecer datos actuales como históricos dentro del contexto en el que se lo esté manejando. (Mora, 2018)

De esta manera las organizaciones podrán manejar de una forma más eficiente, ya que la inteligencia de negocios les permite identificar tendencias de mercado con el fin de incrementar ventas o ingresos. Para esto es necesario hacer buen uso de la información que se tiene, de esta manera a partir de las distintas ramas que tiene la inteligencia de negocios se podrá aplicar diferentes tecnologías dependiendo de la situación en que se los quiera plantear y así hacer los análisis necesarios para una toma de decisión asertiva para el negocio.

De una manera técnica se recopilan datos primarios sobre las operaciones que se realizan dentro de la empresa. Los datos se procesan y estos pasan a ser almacenados. Una vez que se almacenan los datos, los usuarios podrán acceder a ellos, lo que desencadena el proceso de análisis para responder preguntas comerciales, permitiendo extraer conclusiones a partir de dicho análisis, esta se logra a través del uso de estadísticas avanzadas y análisis predictivos, con el fin de encontrar algún tipo de patrón ya sea actual o futuro.

Como es de esperarse la inteligencia de negocios seguirá creciendo y evolucionando de forma continua, dependiendo de las necesidades que haya en el sector empresarial y demás sectores, con esta razón se puede observar las nuevas tendencias que surgen cada año. Así como el B.I, el aprendizaje automático continuara creciendo y siendo parte fundamental de una empresa, a la larga estas tecnologías serán implementadas por el bien de los negocios, el tratamiento y análisis de datos es muy importante a la hora de resolver problemas y tomar decisiones.

Machine Learning

Se considera al Machine Learning (Aprendizaje Automático) como un subconjunto de las técnicas que hay dentro de la inteligencia artificial, este es uno de los conceptos que más está siendo utilizado. Es importante entender que al Machine Learning se lo considera como un concepto, mas no como una metodología o herramienta, ya que existen procesos en este, pero no herramientas.

Existen varias definiciones que se le pueden dar como lo mencionan Wang et al. (2019) “El aprendizaje automático le permite a la maquina aprender sin programación explícita. Este proceso lo logra la maquina mediante la recopilación de datos, el análisis de los datos y la realización de predicciones”. Esto quiere decir que mediante métodos estadísticos se realiza un enfoque inductivo el cual permite que las reglas de decisión se identifiquen en función de los datos que fueron recopilados. (Enholm, et al., 2021)

Los procesos que contiene este concepto cuentan con características de desarrollo de soluciones para lograr evolucionar y aprender de los datos

que se tiene almacenados, dando como resultado la automatización de muchos procesos en los que inclusive no es necesario que haya una compleja interacción de parte del ser humano, ya que este puede generar modelos analíticos para el estudio de los datos.

La implementación del aprendizaje automático se puede dar en varias situaciones siempre y cuando haya datos que sean relevantes para analizar, ya que a través de sus procesos puede ayudar a darle sentido como es la detección de patrones o el poder predecir comportamientos que estén basados en datos históricos. (Norman, 2019)

Para las empresas eso puede generar grandes avances a largo plazo ya que, al hacer uso de sus procesos, se puede determinar si un cliente está próximo a abandonarlo, la disposición a pagar más por parte del cliente por un servicio en específico, el impacto de las ventas en la empresa, el desempeño laboral de un área en específico de una empresa. Eso se logra a partir de las distintas formas o análisis que tiene el aprendizaje automático. (BUSINESS & SCHOOL, 2018)

En la actualidad muchas de las grandes empresas están optando por invertir en los procesos del aprendizaje automático e inteligencia artificial, el adentrarse en este mundo tecnológico aporta una ventaja competitiva bastante alta. Al ser autónomo para realizar tareas brinda este beneficio para las organizaciones al no depender de métodos tradicionales los cuales son propensos a tener errores continuos ya que todo el proceso es elaborado por el ser humano, mientras que este aprendizaje optimiza todo esto bajando de manera considerable las probabilidades de error debido al buen manejo de datos que ofrece.

Este concepto ha demostrado ser importante en el sector empresarial, por el manejo adecuado de datos, el cual permite analizarlos de forma eficaz dada la capacidad recopilarlos y proporcionar información relevante de lo que se está buscando. Permitiendo que las empresas puedan crear nuevas estrategias que puedan cumplir que su misión en el mercado y aprovechar de mejor manera las oportunidades que se presenten.

Aprendizaje Supervisado

Uno de los procesos dentro del Machine Learning es el Aprendizaje Supervisado el cual se define como un conjunto de técnicas que permiten

realizar predicciones futuras en base a comportamientos que son analizados en datos históricos los cuales fueron etiquetados. Esto quiere decir que aprenden funciones, que son asociadas “entradas” con “salidas” dando como resultado que se ajusten a un conjunto. (Calvo, 2019)

Este resuelve problemas que son conocidos al utilizar un conjunto de datos los cuales se encuentran etiquetados para entrenar un algoritmo y de esta manera realice tareas específicas. Sus procesos utilizan modelos de predicción como es el caso de conocer cuantas personas pueden estar dentro de una imagen o en el sector de los negocios, se podría analizar los factores de detección de fraudes o defectos que haya en un producto. Es importante para esto que los datos de entrenamiento estén correctamente etiquetados para la clasificación.

En este caso se supervisa lo que es el entrenamiento del algoritmo a partir de las correcciones que se hagan en los parámetros de este, dependiendo de los resultados que lance. En el caso que la función sea de valor numérico se está hablando de una regresión, mientras que si es una etiqueta se refiere a una clasificación.

El aprendizaje supervisado proporciona un camino directo para transformar los datos en información realista y procesable. El hacer uso de los datos como un recurso proporciona a las empresas el poder comprender e incluso prevenir los resultados que no se esperan. Este aprendizaje es considerado uno de los motores más grandes que hay dentro de los sistemas de la inteligencia artificial al permitir que se tomen decisiones empresariales de una forma muchas más eficaz y precisa que la de los seres humanos. es importante saber que este solo funciona si los datos históricos que van a ser usados contienen valores reales para el resultado el cual se está intentando predecir.

Marco Legal

En el ámbito legal, en el Ecuador existen normativas importantes como son: La Ley Orgánica de Telecomunicaciones, el Código Orgánico de la Economía Social de los Conocimientos, la Ley Orgánica de Protección de Datos y más leyes que se han planteado para el cuidado de datos y uso de tecnologías.

Como podemos observar en según los siguientes artículos de la Ley Orgánica de Protección de Datos Personales (2021) Art. 66, numeral 99:

La Constitución de la República la Constitución de la República reconoce y garantiza a las personas: “19. El derecho a la protección de datos carácter personal, que incluye el acceso y la decisión sobre información y datos de este carácter, así como su correspondiente protección. La recolección, archivo, procesamiento, distribución o difusión de estos datos personales requerirán la autorización del titular o el mandato de ley”.

Además que en el artículo 92 de la Norma Suprema (2021) prescribe que: “Toda persona, por sus propios derechos v como representante legitimado para el efecto, tendrá derecho a conocer de la existencia y acceder a los documentos, datos genéticos, bancos o archivos de datos personales e informes que sobre sí misma, o sobre sus bienes, consten en entidades públicas o privadas, en soporte material o electrónico Asimismo tendrá derecho a conocer el uso que se haga de ellos, su finalidad, el origen y destino de información personal y el tiempo de vigencia del archivo o banco de datos. Las personas responsables de los bancos o archivos de datos personales podrán difundir la información archivada con autorización de su titular o de la ley. La persona titular de los datos podrá solicitar al responsable el acceso sin costo al archivo, así como la actualización de los datos, su rectificación, eliminación v anulación. En el caso de datos sensibles, cuyo archivo deberá estar autorizado por la ley o por la persona titular, se exigirá la adopción de las medidas de seguridad necesarias. Si no

se atendiera su solicitud, ésta podrá acudir a la jueza o juez. La persona afectada podrá demandar por los perjuicios ocasionados”.

Es importante mencionar que, la Resolución 45/95 de 14 de diciembre de 1990 de la Organización de las Naciones Unidas adopta principios rectores para la reglamentación de los ficheros computarizados de datos personales, garantías mínimas que deberán preverse en legislaciones nacionales para efectivizar este derecho.

Así como también uno de los ejes de la Estrategia acordada en el año 2016 de la red Iberoamericana de Datos Personales 2020 consiste en “Impulsar y contribuir al fortalecimiento y adecuación de los procesos regulatorios en la región, mediante la elaboración de directrices que sirvan de parámetros para futuras regulaciones o para revisión de las existentes en materia de protección de datos personales”.

Las leyes en el Ecuador con respecto a la propiedad privada se ven respaldadas por los siguientes artículos puesto que según la Constitución de la Republica del Ecuador (2008) en su Art. 280 menciona que:

...El Plan Nacional de Desarrollo es el instrumento al que se sujetarán las políticas, programas y proyectos públicos; la programación y ejecución del presupuesto del Estado; y la inversión y la asignación de los recursos públicos; y coordinar las competencias exclusivas entre el Estado central y los gobiernos autónomos descentralizados. Su observancia será de carácter obligatorio para el sector público e indicativo para los demás sectores (pág 123).

Capítulo 3: Metodología

En esta investigación se hará uso de Árbol de Decisión y Bosques Aleatorios para la clasificación, el cual a través de sus procesos nos permitirán llegar a lo esperado. El uso de árbol de decisión es un método muy utilizado al ser un modelo de predicción. Teniendo en consideración que la variable la cual va a ser usada es discreta, el modelo será de clasificación. Con este algoritmo se trata de encontrar la variable independiente la cual parta en grupos los datos de una mejor manera. (Marcano, 2017)

La base de datos fue proporcionada por la empresa la cual se está realizando el análisis, al ser una empresa importadora y distribuidora de productos químicos posee una gran cantidad de datos los cuales son importantes para el tipo de análisis que se realizara.

Operacionalizacion de Variables

Tabla 1 Tabla de Variables

Variable	Métrica	Indicador	Ítem
CODVENDEDOR	Cualitativo	Código de los trabajadores del área de ventas	No pasa
DIVISION COMERCIAL	Cualitativo	Áreas de los diferentes tipos de producto	No pasa
UNIDADES	Cualitativo	Unidades de medida de los productos	No pasa
KILOS FACTURADOS	Cuantitativo	Cantidad de kilos que se facturaron del producto	No pasa
PRECIO UNITARIO	Cuantitativo	Precio de los distintos tipos de producto	Pasa
VENTA TOTAL	Cuantitativo	Precio final por las cantidades que se hayan requerido	Pasa

PESO	Cuantitativo	Peso del producto	Pasa
TONELADAS	Cuantitativo	Cantidad de toneladas vendidas	Pasa
PROVINCIA	Cualitativo	Provincia a la que se hace entrega el producto	No pasa
CANTON	Cualitativo	Cantón al que se dirige el producto	No pasa
PARROQUIA	Cualitativo	Parroquia en que se hará entrega del producto	No pasa
VIAENTREGA	Cualitativo	Si el producto llega sin problema o hacen devolución	Pasa

Pasos Por Realizar

A través del uso del programa de R-Studio, que es un interfaz que tiene 4 partes subdivididas se podrá hacer uso de los procesos antes mencionados, como es el caso del Árbol de Decisión. Para empezar la creación del modelo, procedemos a llamar a la base de datos con la función de “read.csv2” esta función lee un archivo en formato de tabla y crea un marco de datos a partir de él. Además, procedemos a delimitar y determinar que variables se van a usar y cuales no, por ende, a través de una función se podrá descartar las variables que no serán parte del proceso. A continuación, se podrá observar las funciones en el programa R-Studio:

```
Data = read.csv2("../DATA/TESIS/DATOS VENTAS 2020 - 2021 - HASTA MAYO 2022.csv")
DataMN = Data[,c(11,12,13,14,15,16,19,23)]
```

Figura 9 Captura de R-Studio del inicio de creación del modelo

Con la base de datos lista, en este caso se procede a realizar el cambio de valores a valores numéricos, se empieza convirtiendo la variable a factor, esta función se utiliza para codificar un vector en factor. Una vez convertida en factor, se debe tomar la misma variable ya transformada a

factor y hacer uso de la función “as.numeric”, para que las variables sean interpretadas como numéricas, en la siguiente ilustración podremos observar la escritura de las funciones mencionadas:

```
#Cambio a valores numericos  
DataMN$Cantidad.Facturada = factor(DataMN$Cantidad.Facturada)  
DataMN$Cantidad.Facturada = as.numeric(DataMN$Cantidad.Facturada)  
DataMN$Precio.Unitario = factor(DataMN$Precio.Unitario)  
DataMN$Precio.Unitario = as.numeric(DataMN$Precio.Unitario)  
DataMN$Total.Venta = factor(DataMN$Total.Venta)  
DataMN$Total.Venta = as.numeric(DataMN$Total.Venta)  
DataMN$peso = factor(DataMN$peso)  
DataMN$peso = as.numeric(DataMN$peso)  
DataMN$Ton = factor(DataMN$Ton)  
DataMN$Ton = as.numeric(DataMN$Ton)
```

Figura 10 Captura de R-Studio/Cambio a valores numéricos de las variables a usar

También es muy importante observar que las variables las cuales van a ser usadas estén bien definidas, con esto se refiere a que, si hay que hacer un cambio de etiqueta o no, al ser un árbol de decisión para la clasificación es importante que la variable sea categórica. Esto se logra llevando la variable categórica a “Factor”, esto nos permitirá realizar la clasificación de dichos datos, se lo hace de la misma manera en que se hizo, solo que esta no se la convierte a numérico, como podemos observar en esta función “***DataMN\$viaentrega = factor(DataMN\$viaentrega, levels=c(0,1))***”, se la transforma en factor y se añade la función “levels”, el cual permite pasar el factor como argumento.

Con las variables listas podemos empezar programar las funciones que son para la creación del árbol de decisión, para que este pueda funcionar hay que partir por la descarga de paquetes los cuales poseen una colección de funciones, datos y códigos de “R”. a continuación, se encuentran los paquetes necesarios para realizar el árbol (Mendoza, 2018):

- Caret: Este paquete contiene utilidades que nos sirven para la clasificación y la regresión. Con esto se hará trabajar la matriz de confusión. Además, nos permite partir los datos en entrenamiento y prueba.

- Rpart: En este paquete se encuentra la implementación de árboles de clasificación los cuales van a ser usados.
- Rpart.plot: Nos permitirá realizar las gráficas de árbol.
- Ggplot2: Es otro paquete el cual permite realizar los gráficos.

Primero se lleva a cabo la instalación de los paquetes en caso de que no se los tenga, una vez se haya finalizado con la instalación, se procede a llamar a las librerías de estos paquetes como se puede observar en la Ilustración 11.

```
#Paquetes a usar
install.packages("caret")
install.packages("ggplot2")
install.packages("rpart")
install.packages("rpart.plot")
library(caret)
library(ggplot2)
library(rpart)
library(rpart.plot)
```

Figura 11 Captura de R-Studio mostrando los paquetes para Árbol de Decisión

Para partir este proceso o cualquier proceso derivado del árbol se debe plantar semilla, hay diferentes clases, pero al ser este un árbol para la clasificación, la mejor semilla a usar es a través de la función “set.seed(123)”, esto es lo que permitirá la creación del árbol en la continuación del proceso.

```
#Para un arbol o cualquier derivado de un arbol de decision, se planta una semilla
set.seed(123)
```

Figura 12 Captura de R-Studio, Función de plantar semilla

Con estos pasos elaborados, se podrá ya avanzar con la creación del árbol de decisión, para esto es importante tener un conjunto de datos de entrenamiento para así realizar el modelo predictivo, adicional al entrenamiento, se debe realizar con set de prueba el cual permitirá

comprobar la eficacia del modelo al momento de hacer las predicciones correctas.

Para iniciar con el entrenamiento de los datos, es necesario partir los datos los cuales vamos a clasificar, como se puede observar en la Ilustración 13 estamos haciendo uso de la función “*createDataPartition*” la cual nos permitirá realizarlo, depende de la cantidad de datos que se tenga se hace la partición, en este caso por la gran cantidad de datos que se posee se optó por una partición de 80% - 20%.

```
#Entrenamos los datos
dataentrenamiento= createDataPartition(DataMN$viaentrega,p=0.80, list = FALSE)
|
```

Figura 13 Captura de R-Studio, Creación de Entrenamiento de datos

Para realizar el árbol, se hace uso del paquete “rpart”, en esta función se requiere especificar la variable objetivo de la clasificación. En la Ilustración 14 se podrá observar un ejemplo de cómo se realiza el árbol y nos podemos dar cuenta que contiene los símbolos (~.), en donde, este símbolo quiere decir que la variable clasificatoria va en relación con el resto del conjunto de datos.

```
#Creamos el arbol de clasificacion
arbol= rpart(viaentrega~., data = DataMN[dataentrenamiento,],
            method = "class", control = rpart.control(minsplit = 300, cp=0.01))
arbol
```

Figura 14 Captura de R-Studio, Creación del árbol de decisión

Podemos llamar al árbol para observar los datos que lanzo, como se muestra en la siguiente Figura, aunque después se lo graficara para una mejor comprensión de este.

```
n= 63389

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

1) root 63389 15493 1 (0.2444115 0.7555885)
2) Cantidad.Facturada< 18.5 2250 481 0 (0.7862222 0.2137778) *
3) Cantidad.Facturada>=18.5 61139 13724 1 (0.2244721 0.7755279)
6) Precio.Unitario< 114.5 1852 755 0 (0.5923326 0.4076674) *
7) Precio.Unitario>=114.5 59287 12627 1 (0.2129809 0.7870191) *
```

Figura 15 Captura de Resultados del Árbol

Una vez hecho el entrenamiento, se procede a realizar el diagrama de árbol en el cual se hace uso del paquete “rpart. plot2”, el cual nos permitirá crear el gráfico de árbol para el análisis de este. Podremos observar en la siguiente ilustración la función necesaria.

```
rpart.plot(arbol, type = 1, digits = -1, extra = 0, cex = 0.8, nn= TRUE,
          fallen.leaves = TRUE )
```

Figura 16 Captura de R-Studio, Función del grafico del Árbol

Una vez ejecutada esta esta función, nos aparecerá el grafico del árbol el cual lo podemos observar en la Figura 17.

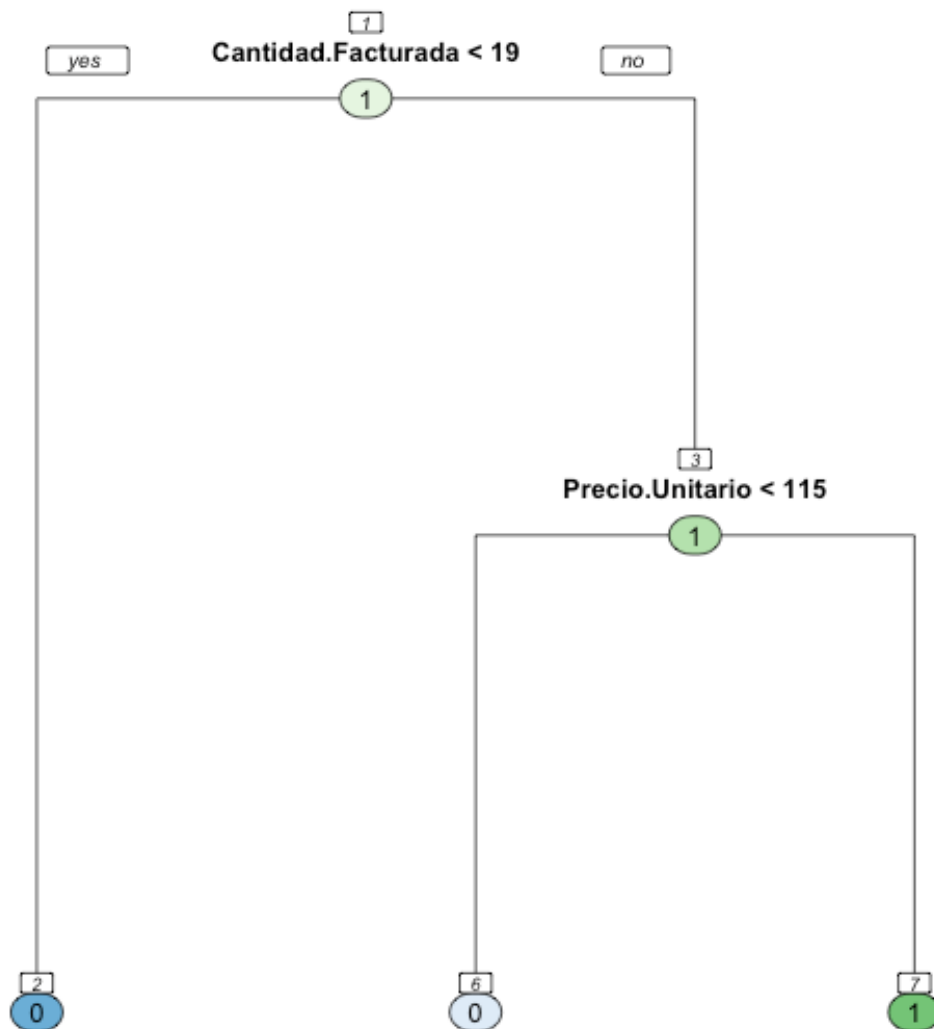


Figura 17 Captura de grafico del árbol

Con todo esto listo se puede apreciar una gráfica clara del árbol de decisión los cuales pueden darnos una idea de que tan preciso fue el modelo al realizar las predicciones permitiéndonos así saber que la probabilidad de

devolución crece en aquellos paquetes menores a 19 unidades y con precios unitarios menores a 115 dólares.

Para tener un mejor análisis de datos a partir del árbol se hace uso del Bosques Aleatorio, este es una simulación de este el cual da resultados más eficientes ya que de-correlaciona mucho más los árboles que serán generados en el proceso. La metodología por usar en R-Studio, parte también con paquetes los cuales permitirán realizar las funciones necesarias para llegar a los resultados esperados.

En este caso al ser una simulación del árbol, se hará uso de paquetes que ya se utilizaron antes como nuevos paquetes que en este caso son:

- randomForest: En este paquete se encuentran los algoritmos más importantes para la creación del modelo de Bosques Aleatorios.
- ROCR: Este paquete es el que nos permitirá realizar diagramas, además, contiene la estructura de la matriz de confusión.
- Caret: Contiene utilidades para la clasificación.

Podemos observar a continuación como llamamos a las librerías que van a ser usadas para la creación del bosque aleatorio.

```
library(randomForest)
library(ROCR)
library(caret)
```

Figura 18 Captura de las librerías a usar en el bosque aleatorio

Con las librerías listas, y la base de datos lista se podrá empezar con el proceso a realizar el bosque. Además, como en el caso del árbol y en cualquier derivado de un árbol de decisión es necesario plantar semilla. En este caso seguimos haciendo uso del mismo tipo de semilla que se usa en el árbol de decisión.

Con los datos que ya fueron entrenados se procede a armar el bosque aleatorio, donde habrá que especificar las variables de los puntos “X” y “Y”, en este caso en la “y” va la variable que estamos clasificando. También, hay que especificar la cantidad de árboles que queremos que haya en nuestro bosque. En la siguiente figura podremos observar la función para armar el bosque aleatorio para clasificar.

```
#Armamos el bosque aleatorio
RandomTreeModel = randomForest(x=DataMN[dataentrenamiento, 1:7],
                                y=DataMN[dataentrenamiento,8],
                                ntree = 1000, keep.forest = TRUE)
```

```
RandomTreeModel
```

Figura 19 Captura de función para realizar el bosque aleatorio

Podemos llamar al bosque aleatorio en donde podremos observar los datos que nos arrojó, en la figura 20 podemos observar que hay un error del 18.43% en el modelo, este es un porcentaje mínimo.

```
Call:
  randomForest(x = DataMN[dataentrenamiento, 1:7], y = DataMN[dataentrenamiento, 8], ntree = 1000, keep.forest = TRUE)

Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 2

OOB estimate of error rate: 18.43%
Confusion matrix:
  0  1 class.error
0 4753 10740 0.69321629
1  941 46955 0.01964673
```

Figura 20 Captura de los datos del bosque aleatorio

Con esto pasamos a la predicción, esta es la media de las predicciones de todos los árboles que lo conforman. Es recomendable realizar primero la predicción y después la matriz de confusión, por ende, se realiza la predicción del bosque que creamos, en este caso la función “predict” que se usa, es una función genérica para realizar predicciones a partir de resultados de varias funciones de ajuste de modelos, en donde podremos observar en la Figura 21 de cómo se la realiza.

```
#Realizo la prediccion
Prediccion = predict(RandomTreeModel, DataMN[-dataentrenamiento,])
Prediccion
```

Figura 21 Captura de función para la predicción

Siguiendo con el proceso, se procedió a realizar la matriz de confusión ya que, en esta se encuentran los valores predichos por el modelo contra los datos observados.

```
#Armo la matriz de confusion
Matriz =table(DataMN[-dataentrenamiento,"viaentrega"], Prediccion, dnn = c("Actual", "Predicho"))
Matriz
```

Figura 22 Captura de la función de matriz de confusión

Con la matriz hecha pasamos a realizar la probabilidad de éxito que hay en el bosque, con esto se obtiene la probabilidad con la que el modelo considera que cada dato puede ser parte de cada una de las clases. En este caso el valor de probabilidad que se podrá observar para cada predicción es la fracción de las observaciones de cada tipo en los nodos terminales que alcanza la observación pronosticada dentro del conjunto de los árboles.

```
#Probabilidad de éxito  
Probabilidad = predict(RandomTreeModel,DataMN[-dataentrenamiento,], type = "prob")  
Probabilidad
```

Figura 23 Captura de la función de probabilidad de éxito

Para finalizar se procede a añadir los valores clasificados que nos dieron al dataframe para poder observar los resultados que nos arrojó el modelo y darnos cuenta de los errores que se pudieron haber dado. Esto se lo hace por medio de la siguiente función:

```
#Valores clasificados  
DataMN[dataentrenamiento,"Prediccion"]= predict(RandomTreeModel,DataMN[dataentrenamiento,])
```

Figura 24 Captura de función para observar los valores clasificados en el dataframe

Resultados

Los resultados que pudimos encontrar en esta investigación muestran que el modelo de árbol de decisión y bosques aleatorios es bueno por el bajo porcentaje de error que tuvo. Además, podemos observar en la siguiente imagen la matriz la cual nos da datos importantes de la variable que estamos tratando.

```
> Matriz
      Predicho
Actual  0     1
  0  1145 2728
  1   229 11745
```

Figura 25 Resultados de la matriz

Se procede a realizar operaciones matemáticas para determinar los porcentajes, en este caso sumamos las primeras cantidades, que son **1145+2728= 3872**. Ahora tomamos el primer número y lo dividimos para el resultado **1145/3873= 0.2956**, que en valor porcentual sería **29.56%**, esto quiere decir que de todo lo que se entrega de producto el 29.56% se suele devolver por algún tipo de error. Ahora con los datos siguientes hacemos el mismo procedimiento, realizamos la suma de **229+11745= 11974**, en este caso tomamos el segundo valor y lo dividimos para el resultado **11754/11974= 0.9808**, que en valor porcentual es **98.08%**. Esto quiere decir que de la buena producción el 98.08% de producto que se entrega, no se regresa.

Ahora también observaremos en la siguiente ilustración datos importantes acerca de las devoluciones de producto:

viaentrega	Prediccion
0	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1

Figura 26 Resultados de la predicción

Como podemos observar en este fragmento de la tabla con los resultados, en la primera fila tenemos el valor “0”, esto quiere decir que el producto que se envió fue regresado por algún tipo de error o problema, pero con el proceso que se realizó podemos verificar que en la predicción nos muestra que el producto no debió ser regresado. Mostrándonos así que hubo fallas en la toma de decisiones con respecto a ciertos envíos.

Análisis de resultados

Se encontró que dentro de los procesos no hay supervisión previa de entrega lo que se debe reestructurar es una clasificación primaria como la que arrojo el análisis del árbol de clasificación dado dos premisas:

- La probabilidad de devolución crece en aquellos paquetes menores a 19 unidades y con precios unitarios menores a 115 dólares por lo tanto debe hacerse una clasificación en la que se tome en cuenta que hay probabilidad de que esos paquetes retornen a la empresa.
- Se debe otorgar estos paquetes a un grupo de entrega especializado en la cual se le dé la orden de reclasificar los paquetes y supervisar nuevamente que estos estén en buenas condiciones para su entrega, de esta manera se bajara la probabilidad de que estos sean devueltos.

Todo esto se hace con la finalidad que estos nuevos procesos discrimen a los antiguos en forma de nuevas órdenes y de esa manera no se incurra en costo alguno por lo tanto se prueba la ayuda del machine learning en correcciones mediante toma de decisiones sin que afecte económicamente a la empresa.

Conclusión

Hoy en día se hace uso de machine Learning para adquirir ventajas ante las demás empresas, ya que nos permite a través del análisis de los datos que se poseen encontrar errores y de esta manera llevarlos a una reestructura para que no se vuelvan a cometerlos y de esta manera se permita tomar decisiones acertadas para que la empresa nos pierda.

Las técnicas que fueron utilizadas nos permiten tener una visión clara de cuál es el rendimiento que tienen los vendedores, a través del árbol de decisión y bosques aleatorios pudimos darnos cuenta de que es necesario reestructurar ciertos puntos, así como también a través de estas técnicas la empresa Quimasa S.A podrá tener un mejor control de sus datos. Y tener conocimiento de los acontecimientos que se dan pero que muchas veces los jefes de área no saben.

También nos podemos dar cuenta del gran aporte que nos dan estas técnicas utilizadas que hacen uso de la inteligencia artificial, al ser estas más eficaces que los métodos tradicionales que las empresas usan y que en muchos casos ni siquiera hacen un buen uso de los históricos de sus datos para un análisis y saber si la eficacia en sus trabajadores o procesos es buena.

La implementación adecuada de los procesos que se realizaron en R-Studio logran garantizar un análisis más eficiente en base a lo que estábamos analizando, por lo que se logró obtener los resultados esperados para esta investigación.

Recomendaciones

En el caso de realizarse un tipo de investigación en donde se haga uso de alguno de los procesos de Machine Learning, es importante que se tenga conocimiento previo de cómo funciona R-Studio, así como también es necesario tener bien planteada la problemática, de esta manera se podrá obtener un buen análisis. Además, parte fundamental es la base de datos que se va a usar, es necesario tener una base de datos que cumpla con los aspectos básicos para aplicar procesos de Machine Learning, también esta base debe ser lo más actualizada posible para un análisis más eficaz.

Se debe tener en consideración el análisis de la base de datos, ya que esta puede contener ciertos errores o información que no ayude a la investigación, o si la base de datos no contiene la cantidad de datos necesarios para un análisis.

Con respecto a los procesos que se realicen, es importante que se lo haga con tiempo, ya que pueden existir errores a la hora de aplicar los algoritmos, causando que el modelo no sea bueno y puede alterar la investigación causando así el no poder llegar a los resultados esperados. Como se mencionó anteriormente, la base de datos debe cumplir con los aspectos básicos para aplicar procesos de Machine Learning, mientras la base de datos tenga una extensa cantidad de datos más eficiente será el modelo que se realice, dando así resultados más claros.

Referencias

- Alvarez-Jareño, J. A., Badal-Valero, E., & Pavía, J. M. (2019). Aplicación de métodos estadísticos, económicos y de aprendizaje automático para la detección de la corrupción . *Revista Internacional- Transparencia e Integridad*, 1-11.
- Athey, S., & Imbens, G. W. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*.
- Bautista, J. I. (2020). Métodos de análisis estadístico con datos dispersos. *Revista Ingeniería y Región Vol. 24* .
- BUSINESS, E., & SCHOOL, M. (2018). El aprendizaje automático en el ámbito empresarial .
- Calvo, D. (2019). *Diego Calvo*. Obtenido de <https://www.diegocalvo.es/aprendizaje-supervisado/>
- CART. (2022). *Minitab*. Obtenido de Minitab: <https://www.minitab.com/es-mx/predictive-analytics/cart/#:~:text=Para%20quienes%20no%20est%C3%A1n%20familiarizados,valores%20y%20combinaciones%20de%20predictores>.
- Chauhan, N. S. (2022). *KDnuggets*. Obtenido de KDnuggets: <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
- Decisión, A. d. (2022). *Bookdown*. Obtenido de Bookdown: <https://bookdown.org/content/2031/arboles-de-decision-parte-i.html>
- Dey, A. (2017). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*.
- Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Inteligencia artificial y valor comercial: una revisión de la literatura. *Frente del sistema de infantería*.
- Fountaine, T., McCarthy, B., & Saleh, T. (2019). Construyendo la organización impulsada por IA. *Harvard Business Review*, 62-73.
- Fradkov, A. L. (2020). Early History of Machine Learning. *IFAC-PapersOnLine*, 1385-1390.

- García, N. (2020). Qué son los árboles de decisión y para qué sirven. *Maxima Formacion*.
- Gómez, K., Magallan, M., Méndez, G., Ramírez, G., Guarda, T., Coronel, M., Quirumbay, D. (2020). Aplicación del aprendizaje automático como solución en los sistemas de negocios. *Revista Ibérica de Sistemas e Tecnologías de Informação*, 137-148.
- Hansen, S. (2018). *Funcas*. Obtenido de Papeles de Economía Española: https://www.funcas.es/wp-content/uploads/Migracion/Articulos/FUNCAS_PEE/157art12.pdf
- Landriscini, G. S. (Diciembre de 2017). *Scielo*. Obtenido de http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-42222017000200004&lang=es
- Machine Learning, e. m. (2019). *Agenda Profesionales & Negocios* . Obtenido de Revistaagenda: [https://www.revistaagenda.net/blog/machine-learning-el-motor-de-la-innovacion-en-las-empresas/#:~:text=Los%20algoritmos%20de%20aprendizaje%20supervisado,%E2%80%9CR%E2%80%9D%20\(corridas\)](https://www.revistaagenda.net/blog/machine-learning-el-motor-de-la-innovacion-en-las-empresas/#:~:text=Los%20algoritmos%20de%20aprendizaje%20supervisado,%E2%80%9CR%E2%80%9D%20(corridas)).
- Mahesh, B. (2018). Machine Learning Algorithms - A Review . *International Journal of Science and Research (IJSR)* .
- Mamani, Y. (2018). *Researchgate*. Obtenido de Researchgate: https://www.researchgate.net/profile/Yonatan-Mamani-Coaquira/publication/323993348_Business_Intelligence_herramientas_para_la_toma_de_decisiones_en_procesos_de_negocio/links/5ab6bc4ba6fdcc46d3b6b9ee/Business-Intelligence-herramientas-para-la-toma-de-decis
- Marcano, A. J. (2017). Algoritmo Adaptativo para la Estimación por Series de Tiempo Usando Árboles de Decisión . *Sesión de Tecnologías y Herramientas Computacionales*.
- Medina, J. A., & Atehortúa, B. E. (2019). Comparison of maximum likelihood, support vector machines, and random forest techniques in satellite images classification. *Tecnura*, 13-26.
- Mendoza, J. (2018). *Rpubs*. Obtenido de https://rpubs.com/jboscomendoza/arboles_decision_clasificacion

- Merino, R. F., & Chacón, C. I. (2017). Bosques aleatorios como extensión de los árboles de clasificación con los programas R y Python. *INTERFASES*, 165-189.
- Mora, G. (2018). Siglo XXI economía de la información: gestión del conocimiento y Business Intelligence, el camino a seguir hacia la competitividad. *SIGNOS-Investigación en Sistemas de Gestión*, 161-174.
- Norman, A. (2019). Aprendizaje Automático En Acción. En A. Norman, *Aprendizaje Automático En Acción*. Litres.
- Pinho, C. M., Gaspar, M. A., & Sassi, R. J. (2022). *Scielo*. Obtenido de <https://doi.org/10.1590/SciELOPreprints.3825>
- Rabbi, F. (2019). A review of the recent trends in the use of machine learning in business. *International Journal of Artificial Intelligence and Machine Learning*.
- Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). Reformar los negocios con inteligencia artificial: cerrar la brecha entre la ambición y la acción. *MIT Sloan Management Review*.
- Rebollo, & Sánchez. (2012). Renovación y reestructuración de destinos turísticos en áreas costeras: marco de análisis, procesos, instrumentos y realidades (Vol. 11). Universitat de Valencia.
- Rojas, E. M. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologías de Informação*, 586–599.
- Román, C. & Quezada, M. (2017). Incidencia de clusters en las exportaciones de banano de Ecuador hacia países de medio oriente durante el periodo 2012-2015. *VI Congreso Internacional de Ciencias Administrativas*, 99-119. Obtenido de https://www.researchgate.net/profile/Mauricio-Noblecilla-Grunauer/publication/319086491_Las_caracteristicas_del_lider_perspectivas_de_los_profesionales_y_poblacion_activa_de_la_Provincia_de_El_Oro/links/598f24e2aca2721d9b684223/Las-caracteristicas-del-lider-perspectivas-de-los-profesionales-y-poblacion-activa-de-la-Provincia-de-El-Oro.pdf

- Roman, V. (2019). *Ciencia y Datos*. Obtenido de Medium:
<https://medium.com/datos-y-ciencia/aprendizaje-supervisado-introducci%C3%B3n-a-la-clasificaci%C3%B3n-y-principales-algoritmos-dadee99c9407>
- Romero Bravo, G. J., Macgluf Issasi, A., Rodríguez Rodríguez, L. A., Espinoza Maza, J. d., & Suárez Álvarez, Á. (2021). Aplicación de Machine Learning en la industria 4.0 en tiempos de pandemia. *Interconectando Saberes*, 1-6.
- Rouhiainen, L. (2018). Inteligencia artificial- 101 cosas que debes saber hoy sobre nuestro futuro. *Alienta Editorial*.
- Sánchez, Z., & Andrés, X. (2018). *Análisis comparativo del otorgamiento de créditos de la Corporación Financiera Nacional en el sector de alimentos frescos y procesados en Guayaquil (2012-2016)*.
<http://repositorio.ucsg.edu.ec/handle/3317/9923>
- Sandoval, L. J. (2019). Algoritmos de aprendizaje automático para análisis y predicción de datos. *REDICCES*.
- Simeone, O. (2018). *Repositorio Universidad de Cornell*. Obtenido de <https://arxiv.org/pdf/1808.02342.pdf>
- Solutions, M. (2021). *ManagmentSolutions-Making things happen* . Obtenido de <https://www.managementsolutions.com/es/publicaciones-y-eventos/informes-sectoriales/white-papers/machine-learning-una-pieza-clave-en-la-transformacion-de-los-modelos-de-negocio>
- Sotaquirá, M. (4 de Junio de 2021). *codificandobits*. Obtenido de [codificandobits: https://www.codificandobits.com/blog/bosques-aleatorios/](https://www.codificandobits.com/blog/bosques-aleatorios/)

ANEXOS

Anexo 1 Script completo en R-Studio

```
1 Data = read.csv2("../DATA/TESIS/DATOS VENTAS 2020 - 2021 - HASTA MAYO 2022.csv")
2 DataMN = Data[,c(11,12,13,14,15,16,19,23)]
3
4 #Cambio a valores numericos
5 DataMN$Cantidad.Facturada = factor(DataMN$Cantidad.Facturada)
6 DataMN$Cantidad.Facturada = as.numeric(DataMN$Cantidad.Facturada)
7 DataMN$Precio.Unitario = factor(DataMN$Precio.Unitario)
8 DataMN$Precio.Unitario = as.numeric(DataMN$Precio.Unitario)
9 DataMN$Total.Venta = factor(DataMN$Total.Venta)
10 DataMN$Total.Venta = as.numeric(DataMN$Total.Venta)
11 DataMN$peso = factor(DataMN$peso)
12 DataMN$peso = as.numeric(DataMN$peso)
13 DataMN$Ton = factor(DataMN$Ton)
14 DataMN$Ton = as.numeric(DataMN$Ton)
15
16 #Cambio a factor la variable a predecir
17 DataMN$viaentrega = factor(DataMN$viaentrega, levels = c(0,1))
18
19 library(caret)
20 install.packages("caret")
21 library(caret)
22 library(ggplot2)
23 library(rpart)
24 library(rpart.plot)
25 install.packages("rpart.plot")
26 library(rpart.plot)
27 library(randomForest)
28 library(ROCR)
29 library(caret)
```

```

#Para un arbol o cualquier derivado de un arbol de decision, se planta una semilla
set.seed(123)

#Entrenamos los datos
dataentrenamiento= createDataPartition(DataMN$viaentrega,p=0.80, list = FALSE)

#Creamos el arbol de clasificacion
arbol= rpart(viaentrega~., data = DataMN[dataentrenamiento,],
            method = "class", control = rpart.control(minsplit = 300, cp=0.01))
arbol
#forma bonita
rpart.plot(arbol, type = 1, digits = -1, extra = 0, cex = 0.8, nn= TRUE,
           fallen.leaves = TRUE )

#Armamos el bosque aleatorio
RandomTreeModel = randomForest(x=DataMN[dataentrenamiento, 1:7],
                               y=DataMN[dataentrenamiento,8],
                               ntree = 1000, keep.forest = TRUE)

RandomTreeModel

#Realizo la prediccion
Prediccion = predict(RandomTreeModel, DataMN[-dataentrenamiento,])
Prediccion

#Armo la matriz de confusion

Matriz =table(DataMN[-dataentrenamiento,"viaentrega"], Prediccion, dnn = c("Actual", "Predicho"))
Matriz

#Probabilidad de exito
Probabilidad = predict(RandomTreeModel,DataMN[-dataentrenamiento,], type = "prob")
Probabilidad

#Valores clasificados
DataMN[dataentrenamiento,"Prediccion"]= predict(RandomTreeModel,DataMN[dataentrenamiento,])

```


Anexo 2 Captura de la predicción

	Cantidad.Facturada	Precio.Unitario	Total.Venta	peso	Doc.Type	Ton	Mes	viaentrega	Prediccion
2839	5	56	2808	1	1	1	2	1	1
1068	5	109	1	10	1	1	12	1	1
1069	5	109	1	10	1	1	12	1	1
1071	5	131	1	10	1	1	12	1	1
1067	5	209	1	10	1	1	12	1	1
1143	5	247	11	10	1	1	4	1	1
1093	5	302	1	10	1	1	9	1	1
1086	5	330	1	10	1	1	6	0	1
1062	5	330	1	10	1	1	6	1	1
1063	5	352	1	10	1	1	6	1	1
1092	5	372	1	10	1	1	7	1	1
1090	5	386	1	10	1	1	6	1	1
1091	5	395	1	10	1	1	7	1	1
1070	5	453	1	10	1	1	12	1	1
1095	5	492	1	10	1	1	3	1	1
1144	5	654	12	10	1	1	7	1	1
1094	5	1299	1	10	1	1	10	1	1
1643	5	1790	9352	3	1	1	2	1	1
2307	5	1910	1173	4	1	1	6	1	1
12519	5	1934	11612	2	1	1	6	1	1
12520	5	1934	11612	2	1	1	6	1	1
34485	5	2038	6836	1	1	1	4	1	1
41733	5	2250	9108	1	1	1	6	1	1
42837	5	2320	9399	1	1	1	4	1	1
48844	5	2538	11250	1	1	1	8	1	1
50318	5	2555	11702	1	1	1	6	1	1
50833	5	2559	11903	1	1	1	5	1	1
1101	6	1725	2	10	1	1	2	1	1
1129	9	243	5	10	1	1	4	1	1
1132	9	265	5	10	1	1	5	1	1

1130	9	273	5	10	1	1	5	1	1
11964	12	1892	11310	10	1	1	9	1	1
13181	12	2040	12050	10	1	1	11	1	1
1615	13	1440	9343	4	1	1	1	1	1
1160	14	643	23	10	1	2	10	1	1
1168	14	691	25	10	1	2	10	1	1
1167	14	694	25	10	1	2	10	1	1
1175	14	702	26	10	1	2	4	1	1
1181	14	791	29	10	1	2	4	1	1
2805	14	2003	2454	10	1	2	4	1	1
33293	15	2173	6481	10	1	2	1	1	1
2493	16	1540	1650	10	1	2	2	1	1
15718	17	1353	129	9	1	2	8	1	1
15720	17	1353	129	9	1	2	11	1	1
16204	17	1355	403	9	1	2	6	1	1
16387	17	1357	406	9	1	2	4	1	1
2781	17	1563	2451	23	1	5	1	1	1
2835	17	1586	2462	23	1	5	11	1	1
3260	17	1624	3484	23	1	5	3	1	1
3911	17	1844	4644	10	1	2	1	1	1
3912	17	1844	4644	10	1	2	2	1	1
3908	17	1844	4644	10	1	2	7	1	1
3909	17	1844	4644	10	1	2	7	1	1
3910	17	1844	4644	10	1	2	8	1	1
3916	17	1846	4646	10	1	2	1	1	1
3926	17	1848	4649	10	1	2	6	0	1
3925	17	1848	4649	10	1	2	6	1	1
4000	17	1853	4902	10	1	2	4	1	1
4211	17	1859	5157	10	1	2	6	1	1
31660	17	1881	5935	10	1	2	1	1	1
Cantidad.Facturada	Precio.Unitario	Total.Venta	peso	Doc.Type	Ton	Mes	viaentrega	Prediccion	
31662	17	1881	5935	10	1	2	1	1	1
31663	17	1881	5935	10	1	2	2	1	1
31667	17	1881	5935	10	1	2	2	1	1
31664	17	1881	5935	10	1	2	3	1	1
31668	17	1881	5935	10	1	2	4	1	1
31669	17	1881	5935	10	1	2	7	1	1
31665	17	1881	5935	10	1	2	9	1	1
31670	17	1881	5935	10	1	2	9	1	1
31671	17	1881	5935	10	1	2	9	1	1
31672	17	1881	5935	10	1	2	9	1	1
31673	17	1881	5935	10	1	2	10	1	1
5247	17	1883	6035	10	1	2	3	0	1
32900	17	1907	6255	10	1	2	8	1	1
32916	17	1908	6274	10	1	2	8	1	1
32917	17	1908	6274	10	1	2	8	1	1
33051	17	1909	6305	10	1	2	8	1	1
33281	17	2005	6481	10	1	2	4	1	1
33282	17	2005	6481	10	1	2	5	1	1
33284	17	2005	6481	10	1	2	5	1	1
33285	17	2005	6481	10	1	2	6	1	1
33286	17	2005	6481	10	1	2	6	1	1
33271	17	2005	6481	10	1	2	7	1	1
33272	17	2005	6481	10	1	2	7	1	1
33287	17	2005	6481	10	1	2	7	1	1
33289	17	2005	6481	10	1	2	8	1	1
33273	17	2005	6481	10	1	2	9	1	1
33290	17	2005	6481	10	1	2	9	1	1
33292	17	2005	6481	10	1	2	10	0	1
33274	17	2005	6481	10	1	2	10	1	1

↕	Cantidad.Facturada ↕	Precio.Unitario ↕	Total.Venta ↕	peso ↕	Doc.Type ↕	Ton ↕	Mes ↕	viaentrega ↕	Prediccion ↕
33291	17	2005	6481	10	1	2	10	1	1
33277	17	2006	6481	10	1	2	1	1	1
33278	17	2006	6481	10	1	2	2	1	1
33279	17	2006	6481	10	1	2	3	1	1
33280	17	2006	6481	10	1	2	4	1	1
33276	17	2006	6481	10	1	2	12	1	1
33542	17	2011	6487	10	1	2	10	1	1
33564	17	2012	6507	10	1	2	10	1	1
33600	17	2013	6518	10	1	2	7	1	1
33601	17	2013	6518	10	1	2	8	1	1
33701	17	2014	6571	10	1	2	10	1	1
5861	17	2016	6637	10	1	2	10	0	1
5862	17	2016	6637	10	1	2	11	0	1
5882	17	2024	6639	10	1	2	4	1	1
34484	17	2037	6835	10	1	2	7	0	1
34496	17	2039	6852	10	1	2	4	1	1
34497	17	2039	6852	10	1	2	5	1	1
34499	17	2039	6852	10	1	2	6	1	1
34500	17	2039	6852	10	1	2	9	1	1
34495	17	2039	6852	10	1	2	10	1	1
34630	17	2040	6898	10	1	2	10	1	1
34845	17	2048	6998	10	1	2	1	1	1
34862	17	2048	6998	10	1	2	1	1	1
34847	17	2048	6998	10	1	2	3	1	1
34850	17	2048	6998	10	1	2	5	1	1
34851	17	2048	6998	10	1	2	6	1	1
34852	17	2048	6998	10	1	2	6	1	1
34864	17	2048	6998	10	1	2	6	1	1
34865	17	2048	6998	10	1	2	6	1	1
34854	17	2048	6998	10	1	2	7	1	1



DECLARACIÓN Y AUTORIZACIÓN

Yo, **Mena Freire, Nicolas Humberto** con C.C: # 1726010596 autor/a del trabajo de integración curricular: **Reestructuración de procesos en una empresa importadora y distribuidora de fertilizantes a nivel nacional** previo a la obtención del título de **Licenciado en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

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

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

Guayaquil, **19 de septiembre del 2022**

f. _____
Nombre: **Mena Freire, Nicolas Humberto**
C.C: **1726010596**



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TRABAJO DE INTEGRACIÓN CURRICULAR

TEMA Y SUBTEMA:	Reestructuración de procesos en una empresa importadora y distribuidora de fertilizantes a nivel nacional		
AUTOR(ES)	Mena Freire, Nicolas Humberto		
REVISOR(ES)/TUTOR(ES)	Carrera Buri, Félix Miguel		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Facultad de Economía y Empresa		
CARRERA:	Negocios Internacionales		
TÍTULO OBTENIDO:	Licenciado en Negocios Internacionales		
FECHA DE PUBLICACIÓN:	19 de septiembre de 2022	No. DE PÁGINAS:	54
ÁREAS TEMÁTICAS:	Ventas, inteligencia artificial, aprendizaje automático		
PALABRAS CLAVES/ KEYWORDS:	Inteligencia Artificial, machine Learning, aprendizaje supervisado, árbol de decisión, bosques aleatorios, inteligencia de negocios		

RESUMEN/ABSTRACT

Con un mundo globalizado existen nuevas maneras y oportunidades a las cuales nos tenemos que adaptar, más que nada a las nuevas herramientas tecnológicas. Para las empresas es importante que realicen un análisis de estas para que así la toma de decisiones sea acertada a largo plazo, dando como resultados se puedan reestructurar procesos para que estos sean óptimos. En este caso, la inteligencia artificial es de los conceptos más llamativos e innovadores para las empresas en los últimos tiempos por el gran aporte que este genera en el ámbito empresarial.

Dentro de los conceptos de inteligencia artificial, tenemos al Machine Learning, este brinda los recursos necesarios para realizar un análisis de lo que una empresa necesite por medio del manejo de datos que esta posee. El objetivo que se tiene en esta investigación es el analizar la eficiencia de los vendedores de la empresa Quimasa S.A, por medio de distintos procesos de Machine Learning, ya que esta empresa no ha implementado este tipo de análisis. Se obtuvo los datos de ventas de la empresa del último año y se hizo uso de procesos de Aprendizaje Supervisado, los cuales fueron Árbol de Decisión y Bosques Aleatorios, para así dar a conocer a la empresa el rendimiento que hubo en las ventas que se realizaron y se pueda mejorar a largo plazo para que se cometan los mismos errores.

ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO
CONTACTO CON AUTOR/ES:	Teléfono: +593-987915687	E-mail: nicolas.mena2108@gmail.com
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UIC):	Nombre: Román Bermeo, Cynthia Lizbeth	
	Teléfono: +593-984228698	
	E-mail: cynthia.roman@cu.ucsg.edu.ec	

SECCIÓN PARA USO DE BIBLIOTECA

Nº. DE REGISTRO (en base a datos):	
Nº. DE CLASIFICACIÓN:	
DIRECCIÓN URL (tesis en la web):	