



**TRABAJO DE GRADO**  
**Opción Seminario-Diplomado.**

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES**  
**EN DATOS DE VENTAS DE ALMACENES DE LA COMPAÑÍA GLOBAL DE**  
**PINTURAS-PINTUCO, UTILIZANDO ESTRATEGIAS DE MACHINE LEARNING**

Corporación Universitaria Remington.

Facultad de ingeniería

Especialización en dirección de operaciones y mejoramiento continuo

Estudiante:

Sebastián Torres Ramírez

Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.

2024.

## **Dedicatoria**

A Dios por enseñarme y acompañarme en el camino, muchas veces no nos damos cuenta que es Él quien nos guía, a mi madre por apoyarme y motivarme cada día a ser mejor persona y profesional.

**Tabla de Contenidos**

Marco conceptual y contextual .....	5
Pregunta problema .....	6
Acercamiento a los datos .....	6
Descripción de variables .....	6
Objetivos .....	9
Objetivo general.....	9
Objetivos específicos .....	9
Desarrollo e implementación del aprendizaje.....	9
Metodología de trabajo .....	10
Preparación de los datos.....	10
Modelo de toma de decisiones.....	11
Validación del modelo .....	13
Conclusiones y trabajos futuros .....	14
Referencias bibliográficas.....	15

### **Resumen**

En este trabajo se realiza un análisis de los dataset de ventas de diferentes tiendas oficiales de la empresa Pintuco para determinar con la ayuda de Google Colab y el código PYTHON el comportamiento de los mismos, realizar un análisis estadístico y gráfico para posteriormente proceder con una clasificación de los mismos. Después de esto se realiza un modelo computacional que permite realizar una clasificación de nuevos datos basado en la variabilidad anterior. Esto se realiza con el interés de tener criterios para asignación de incentivos económicos y laborales a los empleados de las tiendas basados en el desempeño obtenido, además en un futuro impulsar campañas de ventas y marketing que permitan aumentar las ventas en cada una de las tiendas.

### **Palabras clave**

Ventas, Datasets, Análisis de datos, Machine learning, Google colab, Clasificación, modelo, marketing.

### **Marco conceptual y contextual**

El manejo de los inventarios es de gran importancia en todas las empresas y almacenes, la cadena de suministro debe estar siempre abastecida tanto de materias primas, equipos, personal y productos disponibles para los clientes. Autores como Jara Et al [1] especifican que las empresas que no se adaptan a los cambios del mercado y las tecnologías desaparecen, como fue el caso de blackberry, hifive o blockbuster y otras compañías a través de los últimos 20 años, estas consecuencias no solo ocurren por una desconexión en las tendencias del mercado, sino también, por incumplimientos en abastecimiento de las materias primas y los servicios, manejo excesivo de inventarios o desconocimiento de la rotación de los mismos, lo que representa pérdidas de capital, clientes, credibilidad y gastos innecesarios. [2]

Por otro lado Villacres Et al[3] expone la importancia de la gestión de ventas en los almacenes, llevar los controles de la producción y de los comentarios de los clientes. La gestión de ventas y marketing busca entender el comportamiento en las ventas y los clientes para así determinar estrategias que impulsen el crecimiento de las ventas [6].La utilización de herramientas tecnológicas es muy importante en la actualidad para el control, seguimiento, gestión e impulso de la cadena de suministro, producción, marketing y ventas en cualquier empresa de la cuarta revolución industrial, la implementación de estas herramientas permite ofrecer a los consumidores un mejor servicio, un mejor producto, se fideliza y se llama la atención de los clientes y se generan estrategias de ventas que buscan mantener y mejorar los indicadores de ingresos. La incorporación de tecnologías en las ventas se ha potencializado con el avance en la inteligencia artificial, esto ha permitido la agrupación de los datos y la predicción de resultados basado en el machine learning, la incorporación de la realidad aumentada e inteligencia artificial en la industria y los almacenes, Barrientos Et al [4] especifica que las tiendas que ofrecen servicios tienda a tienda tienen el desafío de controlar los procesos y negocios para tener mejor productividad.

La compañía global de pinturas, Pintuco es una empresa colombiana con más de 100 años en el mercado, con influencia en los mercados de centro América y el caribe, especialmente en Colombia, Ecuador, Costa Rica, Panamá y las Antillas, actualmente hace parte de la

multinacional AkzoNobel la cual tiene presencia en todo el mundo [5]. Con más de 1000 productos en el portafolio, abastece el mercado con hasta 6 veces más productos en las diferentes presentaciones. Cuenta con diferentes tiendas distribuidas en el territorio colombiano y en otros países, las tiendas cuentan con un registro de ventas con los cuales se generan algunos de los indicadores más importantes para la compañía.

El presente trabajo se enfoca en clasificar y agrupar los datos recopilados para hacer un análisis del comportamiento de las ventas en las diferentes tiendas oficiales, y poder asignar incentivos o bonificaciones a los empleados basados en el resultado arrojado por machine learning. Es de gran importancia generar una buena clasificación para no tener inconvenientes con los trabajadores y poder fortalecer estos incentivos en un futuro.

### **Pregunta problema**

¿Cómo desarrollar una estrategia computacional para desarrollar planes de incentivos para las tiendas a partir de algoritmos de Machine Learning?

### **Acercamiento a los datos**

Los datos corresponden a los registros de ventas de las tiendas del supermercado Walmart, los cuales fueron convertidos y optimizados para casos prácticos, se modificaron con el objetivo de generar un caso práctico en donde se busca hacer un comparativo con las ventas de las tiendas de la Compañía global de pinturas- Pintuco, relacionando los resultados con las unidades en inventario, el precio del dólar y el número de empleados, para generar criterios válidos para asignar los incentivos económicos y laborales para los empleados de las tiendas. En total se tiene un archivo de Excel con 5760 registros.

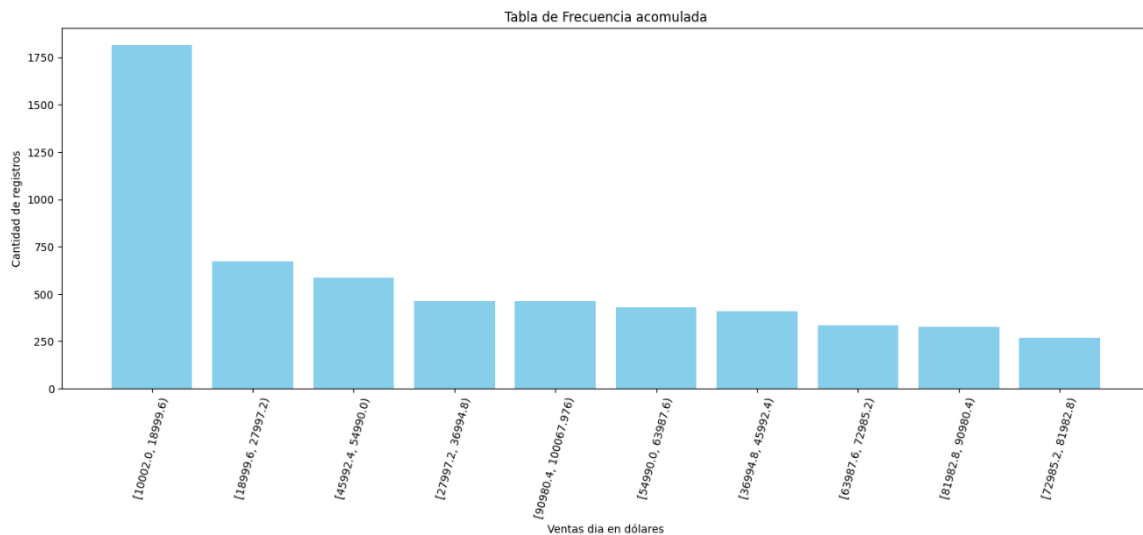
### **Descripción de variables**

#### **En las variables tenemos:**

- **Tienda:** 1-45 corresponden a las 45 tiendas oficiales de la compañía distribuidas por toda la región de centro América, Colombia y Ecuador
- **Fecha:** Fecha en la que se recogieron los datos.

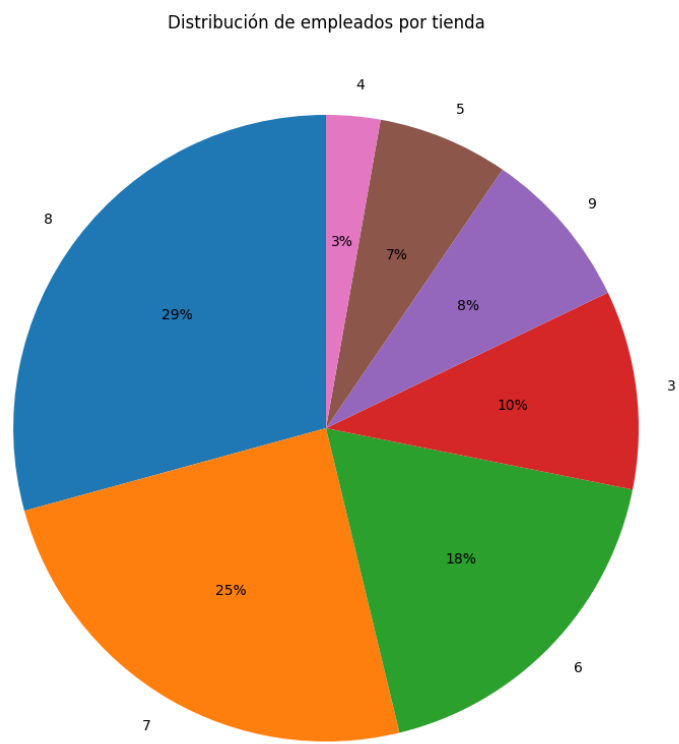
- **Ventas\_dia\_dolar:** Ventas en dólares registradas en el día
- **Turno:** 0 o 1, el turno 0 corresponde si la venta se realizó en semana y el turno 1 si se realizó un fin de semana.
- **Unidades vendidas:** Numero de unidades vendidas de pintura, independientemente del volumen
- **Unidades en inventario:** Unidades disponibles en inventario el día de la venta
- **Precio del dólar:** Tasa representativa del mercado para Colombia el día del registro
- **Ventas\_dia\_peso:** Ventas en pesos colombianos el día del registro.
- **Número de empleados:** Número de empleados en la tienda en el turno registrado.

En el grafico 1 se relaciona la tabla de frecuencias acumulada para las ventas diarias en dólares en donde se puede observar que la mayor cantidad de registros se encuentra entre \$10.000 y \$19.000 lo que, a su vez, es el intervalo de menor valor

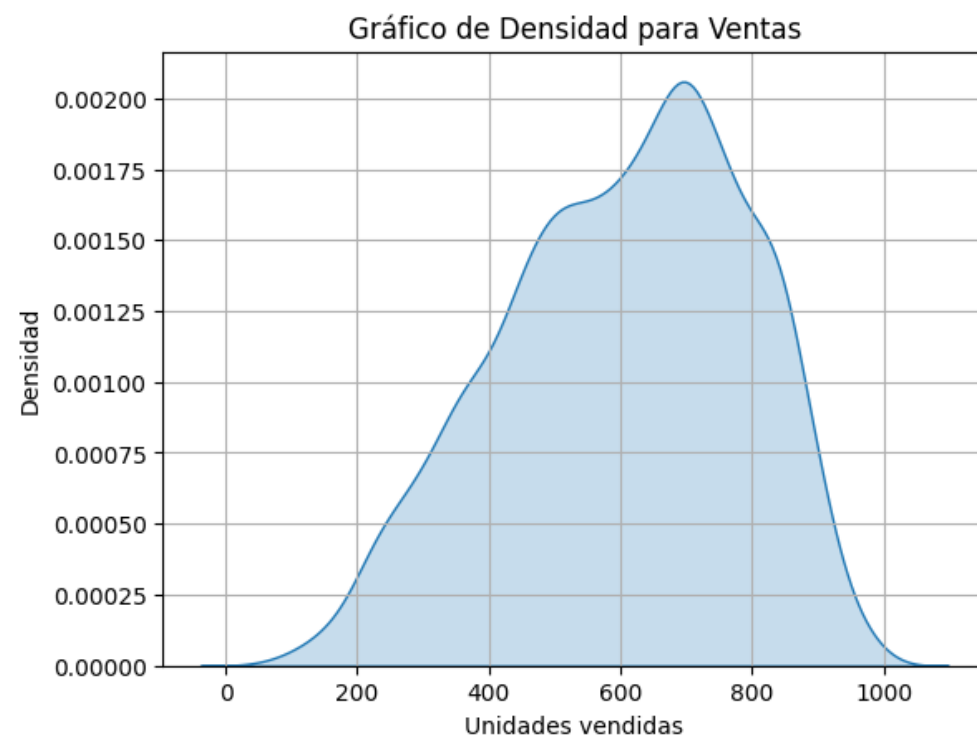


*Grafico 1. Frecuencia acumulada de ventas diarias en dólares*

También se grafica la distribución de numero de empleados en todo el archivo, obteniendo el grafico 2. En el grafico 3 se representa la distribución de las unidades vendidas, con el fin de identificar visualmente la tendencia de ventas en los almacenes de la compañía, se obtiene como resultado que en promedio los almacenes están vendiendo 700 unidades en cada turno registrado.



*Grafico 2. Distribución de empleados por tienda*



*Grafico 3. Representación de unidades vendidas por dia*

## Objetivos

### Objetivo general.

Realizar un análisis de datos de las ventas de las tiendas de Pintuco y crear e implementar un algoritmo para la toma de decisiones utilizando estrategias de machine learning.

### Objetivos específicos

- Analizar, caracterizar y procesar los datos obtenidos, enfocado en el desarrollo del proyecto de toma de decisiones.
- Desarrollar e implementar un algoritmo computacional que permita asignar incentivos a los empleados de las tiendas a partir de la correlación de los datos analizados.
- Realizar una evaluación al resultado obtenido en nuevos datos recopilados

### Desarrollo e implementación del aprendizaje

La metodología consiste en el análisis y clasificación de los datos con la ayuda de Google Colaboratory, el cual es una herramienta gratuita en la nube que permite desarrollar y trabajar con código Python, inicialmente se descargan datos de la Kaggle los cuales son de las ventas de los almacenes Walmart, estos datos fueron convertidos y optimizados con el fin de obtener una base de datos o dataset que represente un caso más real y familiar, es así como se cambian las variables y se genera un conjunto de datos de que pueden representar ventas de las tiendas Pintuco. Gracias a las herramientas obtenidas en el curso se cargan los datos en extensión .xlsx, se realiza un análisis y depuración de datos y de variables que no aportan valor para el estudio posterior, se determina el comportamiento de los datos de manera gráfica y numérica, por medio de códigos, se realiza tabla de frecuencias acumulada para determinar la tendencia de los datos y por ultimo se realiza segmentación de los datos para determinar la mejor cantidad de grupos en los que se pueden clasificar los

datos, para posteriormente realizar un análisis de centroides y una IA que nos permita basado en estos criterios tomar decisiones de manera automática

## Metodología de trabajo

### Preparación de los datos

Lo primero que se realiza, es el descargue de los datos de una de las plataformas de dataset sugeridas (<https://www.kaggle.com/datasets/yasserh/walmart-dataset/data>) estos datos se analizan, se procesan y se convierten para tener un archivo de 5760 datos, en .xlsx que nos ayuden a contar una historia relacionada con el numero de ventas y valores recaudados en diferentes tiendas de una empresa de pinturas colombiana, posteriormente se realiza análisis de los datos en la herramienta, tanto de características como estadístico como se puede ver en las tablas 1 y 2.

```

↔ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 5761 entries, 0 to 5760
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Tienda                5761 non-null   int64
1   Fecha                 5761 non-null   datetime64[ns]
2   Ventas_dia_dolar     5761 non-null   int64
3   Turno                 5761 non-null   int64
4   Unidades vendidas    5761 non-null   int64
5   Unidades en inventario 5761 non-null   int64
6   Precio del dólar    5761 non-null   int64
7   Ventas_dia_peso      5761 non-null   int64
8   Numero de empleados  5761 non-null   int64
dtypes: datetime64[ns](1), int64(8)
memory usage: 405.2 KB

```

*Tabla 1. Características de los datos*

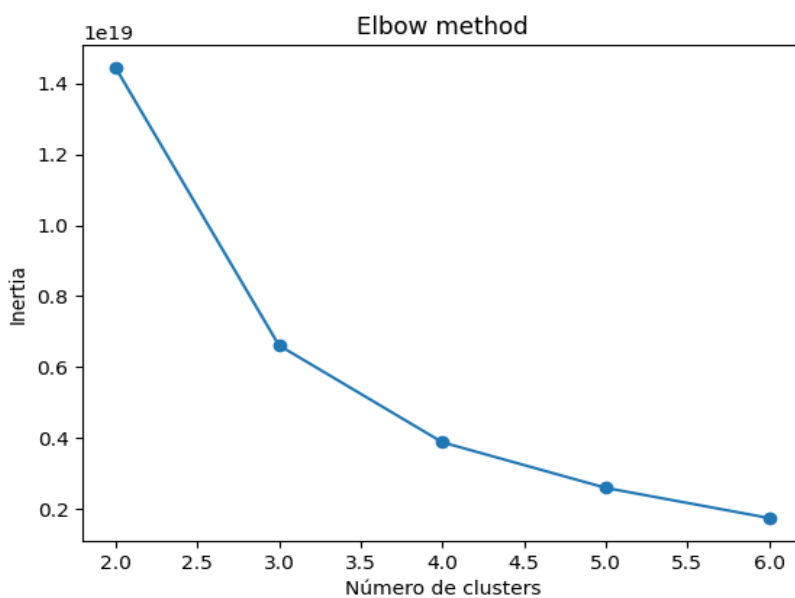
	Tienda	Ventas_dia_dolar	Turno	Unidades vendidas	Unidades en inventario	Precio del dólar	Ventas_dia_peso	Numero de empleados
<b>count</b>	5761.000000	5761.000000	5761.000000	5761.000000	5761.000000	5761.000000	5.761000e+03	5761.000000
<b>mean</b>	23.459990	42103.144072	0.496442	603.353584	1898.250651	2993.324423	1.276794e+08	6.649019
<b>std</b>	13.002731	28037.280890	0.500031	185.174565	829.332374	1076.893950	1.015255e+08	1.687285
<b>min</b>	1.000000	10002.000000	0.000000	60.000000	850.000000	3.000000	4.194800e+04	3.000000
<b>25%</b>	12.000000	16062.000000	0.000000	470.000000	1173.000000	2773.000000	4.190550e+07	6.000000
<b>50%</b>	24.000000	35499.000000	0.000000	620.000000	1580.000000	3283.000000	9.519035e+07	7.000000
<b>75%</b>	35.000000	62806.000000	1.000000	750.000000	2762.000000	3697.000000	1.999023e+08	8.000000
<b>max</b>	45.000000	99978.000000	1.000000	1000.000000	3580.000000	4468.000000	4.275222e+08	9.000000

*Tabla 2. Análisis estadístico de los datos*

Se realiza una depuración de variables, en la cual se elimina la columna de fecha, y se analiza gráfica y matemáticamente los datos para tener un comportamiento de los mismos y poder desarrollar un algoritmo que se acomode a los datos.

### **Modelo de toma de decisiones**

Para este estudio se decidió realizar el modelo de toma de decisiones basado en la segmentación de datos o clustering, para esto se realizó un análisis de los datos para determinar y gestionatr la mejor cantidad de beneficios o incentivos a entregar basado en el comportamiento de los datos, para tal resultado se realiza un análisis de codo obteniendo la gráfica 4.



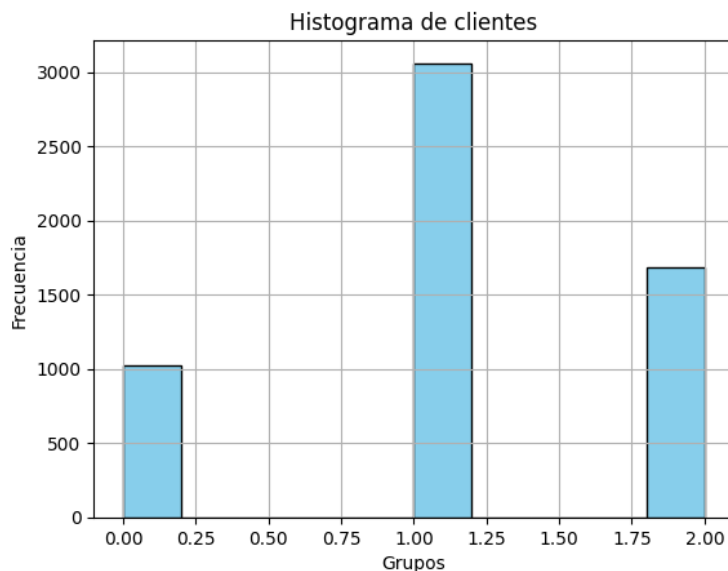
*Gráfico 4. Calculo de agrupaciones de los datos*

Con el resultado del grafico anterior se determina que la mejor cantidad de grupos en los que se pueden clasificar los datos es 3, esto quiere decir que se pueden generar tres incentivos para las tiendas, los cuales están enfocados en remuneraciones económicas y en tiempo de descanso para los trabajadores, dicho esto se procede con la generación de los centroides con los cuales se procederá a determinar la variable mas representativa para la clasificación o agrupación de los datos. El resultado de este procedimiento se muestra en la tabla 3.

	Tienda	Ventas_dia_dolar	Turno	Unidades vendidas	Unidades en inventario	Precio del dólar	Ventas_dia_peso	Numero de empleados
<b>0</b>	25,70	85180,91	0,52	577,19	1881,83	3536,56	299742383,74	6,41
<b>1</b>	21,56	22284,83	0,49	602,17	1901,38	2645,09	47553960,30	6,73
<b>2</b>	25,57	52113,42	0,49	621,38	1902,50	3298,92	169425669,03	6,65

*Tabla 3. Calculo centroides de los datos agrupados*

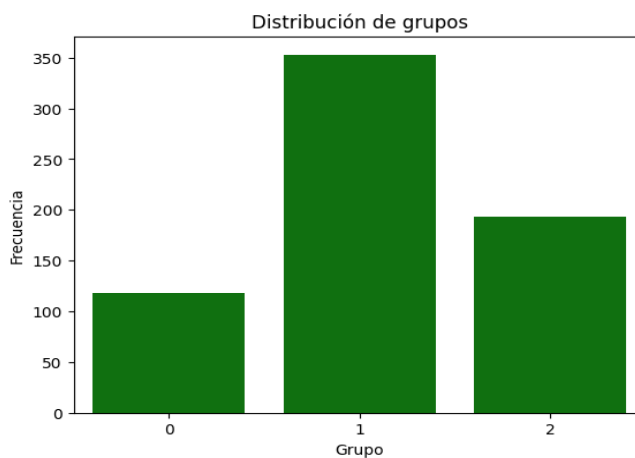
Del resultado anterior se puede observar que la variable con mayor representación en los datos son las ventas, tanto en dólares como en pesos ya que son directamente proporcionales dado que son los resultados con mayor variación en los tres grupos seleccionados. En la grafica 5 se pude observar la distribución de los datos en los tres grupos, confirmando así, la cantidad de datos estudiada y la variabilidad de los mismos, la mayor cantidad de datos se encuentra en el grupo 1.



*Grafico 5. Representación grafica de la agrupación de datos*

### **Validación del modelo**

El modelo de datos generado se evalúa en un grupo de datos diferente el cual no ha sido analizado, para determinar por medio de la IA que tipo de beneficio o incentivo se puede asignar, el código asigna un grupo al archivo de datos el cual tiene un total de 65° registros, los resultados se muestran en la grafica 6.



*Grafico 6. Agrupación de datos nuevos*

## Conclusiones y trabajos futuros

- Se puede concluir que los datos obtenidos presentan una variabilidad importante lo que permite tener una mejor clasificación de los mismos, además se pudo realizar un análisis estadístico y gráfico lo generó visualizaciones de la cantidad y clasificación de los mismos.
- Se pudo realizar una clasificación de los datos, en el cual se obtuvieron tres grupos principales, por lo que se puede sugerir a la compañía que genere tres incentivos económicos o laborales para los empleados de la compañía.
- Con la generación del modelo de inteligencia artificial se pudo realizar un análisis de nuevos datos obteniendo una clasificación de los datos en tres grupos establecidos, lo que significa que pueden repartir los incentivos de manera equitativa basados en el comportamiento histórico de los datos.
- Se pudo observar y determinar la importancia y relevancia de las técnicas de machine learning en la actualidad para la clasificación, agrupación y predicción de datos y situaciones en casos de la vida real, esto nos permite tener una nueva herramienta para el futuro.

### Referencias bibliográficas

[1] Jara-Cordero, S., Sánchez-Partida, D., & Martínez-Flores, J. L. (2017). Análisis para la mejora en el manejo de inventarios de una comercializadora. *Revista de Ingeniería Industrial*, 1(1), 1-18.

[2] Arroba Salto, J. E., Angulo Rosales, Y. A., & Naula Valla, S. M. (2018). Control de inventarios y su incidencia en los estados financieros. *Observatorio de la economía latinoamericana*, (noviembre).

[3] Villacres-Guato, G. A., Romero-Fernández, A. J., & Gallegos-Riofrío, R. X. (2022). Gestión de proceso de ventas para el mejoramiento de la rentabilidad. *CIENCIAMATRIA*, 8(15), 154-171.

[4] Barrientos-Avendaño, E., Coronel-Rojas, L. A., Cuesta-Quintero, F., & Rico-Bautista, D. (2020). Sistema de administración de ventas tienda a tienda: Aplicando técnicas de inteligencia artificial. *Revista ibérica de sistemas e tecnologías de informação*, (E27), 677-689.

[5] Pinturas Pintuco, El color de la calidad. (2024, 4 junio). Historia - Pintuco. Inturas Pintuco | el Color de la Calidad. <https://www.pintuco.com.co/historia/?section=hitos-mas-significativos-en-la-historia-de-pintuco>

[6] Núñez Garcés, D., Parra Cruces, M., & Villegas Pinuer, F. (2011). Diseño de un modelo como herramienta para el proceso de gestión de ventas y marketing—Financiado por Proyecto Fondecyt N. 1110013.