



TRABAJO DE GRADO
Opción Seminario-Diplomado.

**SISTEMA DE RECOMENDACIÓN PARA COMPRA DE MATERIAL DE FORMA
INTELIGENTE A PARTIR DE DATOS DE CONSUMO, UTILIZANDO ESTRATEGIAS
DE MACHINE LEARNING**

Corporación Universitaria Remington.
Nombre de la facultad: Ingenierías
Nombre del programa académico: Ingeniería Industrial e ingeniería de sistemas

Estudiantes:
ING Jose David Ortega Fuentes
ING Gilberto José Rodríguez Guzmán
ING Eliana yulied Albarracín Cristancho

Tutor: ING Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.
2024.

AGRADECIMIENTOS

A DIOS, A NUESTRAS FAMILIAS A LA UNIVERSIDAD.
A MIS COMPAÑEROS, AL PROFE DEL SEMINARIO, MERO TALENTO; y a todas aquellas personas que suman en mi proceso de formación.

TABLA DE CONTENIDOS

	Resumen	5
1.	Marco conceptual y contextual	6
2.	Objetivos	7
2.1	Objetivo general	7
2.2	Objetivo específico	8
3.	Desarrollo e implementación del aprendizaje	8
3.1	Preparación y análisis de los datos	8
3.2	Modelo de toma de decisiones	9
3.3	Análisis de desempeño	
3.4	Validación del modelo	
4.	Conclusiones y trabajos futuros	
	Referencias bibliográficas	

RESUMEN

Ingeomega SAS es una empresa de ingeniería con sede en Medellín que ofrece soluciones en áreas como telecomunicaciones, electricidad y obras civiles. En el desarrollo de sus proyectos, Ingeomega ha detectado la necesidad de optimizar el proceso de adquisición de materiales para ajustarse a las necesidades específicas de cada grupo de trabajo. Esta estrategia de compras inteligentes se centra en mantener un control preciso sobre los materiales utilizados, con el objetivo de reducir costos por exceso de inventario y minimizar el riesgo de desabastecimiento, lo que podría afectar la continuidad de sus proyectos.

Este estudio busca analizar cómo la implementación de un modelo de Machine Learning basado en los datos históricos de consumo de materiales puede mejorar la eficiencia del proceso de abastecimiento y reducir el riesgo de desabastecimiento en la empresa. La hipótesis plantea que un análisis predictivo de consumo permitirá a Ingeomega predecir con precisión las necesidades de materiales para cada proyecto, optimizando así el proceso de compras y reduciendo los costos asociados.

Para alcanzar estos objetivos, se propone un modelo de análisis de datos que permita prever el consumo de materiales en función de patrones históricos. Los objetivos específicos incluyen el desarrollo de un modelo de Machine Learning para la predicción de consumo, la implementación de un sistema de gestión de inventario basado en patrones de consumo,

la evaluación de la reducción de costos y la implementación de un sistema que evite el desabastecimiento de materiales en proyectos clave de la empresa.

En cuanto a la metodología, el análisis se realizará sobre un conjunto de datos que contiene registros detallados de consumo de materiales de Ingeomega SAS, empresa que brinda servicios a Empresas Públicas de Medellín (EPM) en el área de control de pérdidas en la región del Bajo Cauca antioqueño. Los datos cuentan con 9 columnas y 16,496 registros, lo cual permite realizar un análisis exhaustivo para identificar patrones y tendencias en el uso de los recursos materiales. Con estos datos, se espera obtener una mejora en la planificación y administración de los inventarios de la empresa.

PALABRAS CLAVE

Optimización de inventario, Aprendizaje automático, Consumo de materiales
Predicción de demanda, compras inteligentes

MARCO CONCEPTUAL Y CONTEXTUAL

1.1 CONTEXTO

1.2 SISTEMAS DE RECOMENDACIÓN

Los sistemas de recomendación son herramientas que utilizan técnicas de filtrado y análisis de datos para predecir y sugerir productos o contenidos a los usuarios en función de sus preferencias, comportamientos previos o datos de consumo. Estas herramientas son ampliamente utilizadas en sectores como el comercio electrónico, las plataformas de entretenimiento y las redes sociales, facilitando una experiencia personalizada para los usuarios. Su propósito es ayudar a los usuarios a encontrar de manera rápida y eficiente los elementos que más se ajusten a sus necesidades o intereses.

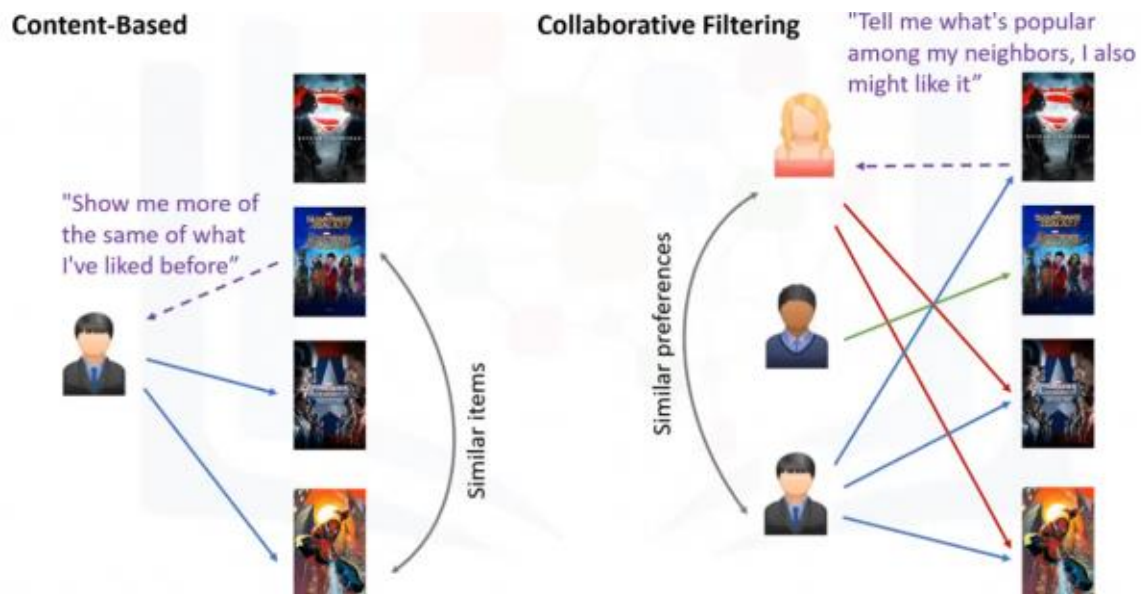
Existen diversos tipos de sistemas de recomendación, entre los que destacan:

- **FILTRADO COLABORATIVO:** Utiliza las preferencias y comportamientos de usuarios similares para realizar recomendaciones, partiendo de la premisa de que las personas con gustos parecidos disfrutarán de opciones similares.

El Filtrado colaborativo y los motores de recomendación son una combinación de técnicas en la que participan algoritmos para encontrar coincidencias y recomendar elementos que tienen alguna relación. Las cosas relacionadas pueden ser artículo, pero podría ser también cualquier cosa como un comportamiento, una dieta, un estilo de vida o cualquier fenómeno en el que existan relación. En resumen, esta es una técnica que de alguna forma logra luego de utilizar el Machine Learning proponer recomendaciones de alguna forma personalizadas las cuales se adaptan de la mejor forma a las necesidades y características de un usuario.

El uso de los motores de recomendación es altamente apreciado por los comerciantes en línea condicionados por sus numerosas ventajas, las que iremos analizando más adelante.

¿QUÉ ES UN SISTEMA DE RECOMENDACIÓN O FILTRADO COLABORATIVO?:



EL FILTRADO COLABORATIVO Y LOS MOTORES DE RECOMENDACIÓN.

FUENTE: STATDEVELOPER

Introducción a los sistemas de recomendación

El Filtrado colaborativo y los motores de recomendación son en términos sencillos, un software que filtra las opciones de los usuarios y les brinda las sugerencias más adecuadas en función de sus requisitos o preferencias.

El primer sistema de recomendación se creó en la década de 1970, en la comunidad de investigación de la Universidad de Duke; luego fue desarrollado por Xerox Palo Alto Research Center.

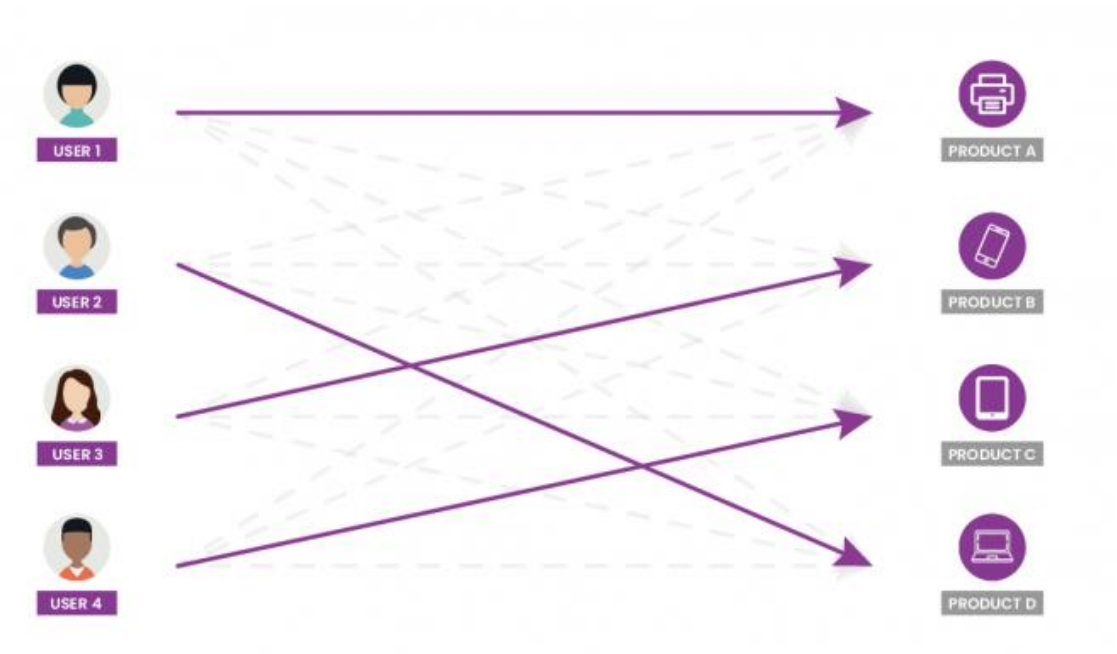
Luego, Cuando apareció Internet en la década de 1990, los sistemas de recomendación se adoptaron de inmediato como la forma de ayudar a las personas a seleccionar los productos más adecuados entre una gran cantidad de opciones disponibles.

Desde entonces, los sistemas de recomendación se han vuelto cada vez más populares y ahora juegan un papel fundamental para las grandes empresas de Internet como Facebook, Amazon, Netflix, Google, YouTube y Tripadvisor, incursionando en los ámbitos de las redes sociales, el entretenimiento, el comercio electrónico, turismo, emparejamiento y más.

Veamos ahora las características de los motores de recomendación, explorando lo que implica construir, uno de alto rendimiento para que esté mejor equipado cuando decida cómo integrar esta funcionalidad en sus aplicaciones.

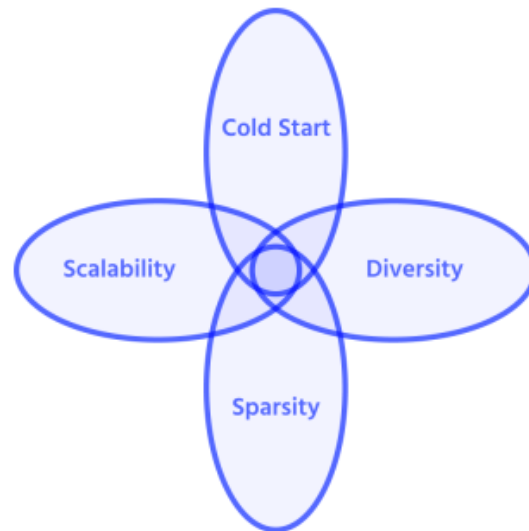
Los cuatro principios rectores de los sistemas de recomendación personalizados

Primero, dentro de El Filtrado colaborativo y los motores de recomendación debemos hacer una distinción entre los sistemas de recomendación personalizados y no personalizados . Con recomendaciones no personalizadas, todos los usuarios reciben las mismas recomendaciones. Algunos ejemplos de recomendaciones no personalizadas son películas populares, canciones exitosas recientes y restaurantes mejor calificados en un lugar determinado.



El objetivo operativo más obvio de usar un sistema de recomendación personalizado es recomendar artículos que sean relevantes para el usuario, ya que es más probable que las personas compren artículos que les parezcan atractivos.

El Filtrado colaborativo y los motores de recomendación deben lograr cuatro objetivos secundarios:



A. ARRANQUE EN FRÍO. Cuando no hay suficiente información o metadatos disponibles, un motor de recomendación no funciona de manera óptima. Hay dos problemas (extremos) de arranque en frío: cuando no hay datos sobre un usuario, acá hay una falta de conocimiento previo sobre el visitante y no hay suficiente información de la sesión actual., Cuando no hay suficientes comentarios o calificaciones de los usuarios. Por ejemplo, supongamos que un visitante por primera vez busca un nuevo teléfono inteligente en un sitio de comercio electrónico. En una semana, compra uno y ya no está interesado en buscar teléfonos. ¿Qué debería mostrar ahora el motor de recomendaciones?

B. SUPERAR LA ESCASEZ DE DATOS. La falta de datos surge del hecho de que los usuarios en un sitio web de comercio electrónico tienden a interactuar (agregar un artículo al carrito, comprar o revisar) con una cantidad limitada de artículos. La mayoría de los motores de recomendación agrupan calificaciones de usuarios similares; sin embargo, la

matriz de usuario-elemento informada tiene hasta un 99 % de calificaciones vacías o desconocidas debido a la falta de incentivos para que el usuario califique los elementos o al conocimiento suficiente sobre un elemento para calificarlo. Como resultado, los usuarios que no brindan comentarios ni calificaciones reciben recomendaciones irrelevantes.

La escasez de datos se produce cuando hay una falta o limitación de datos de entrenamiento etiquetados, o cuando hay un desequilibrio de datos para una etiqueta determinada. Para superar la escasez de datos, se pueden utilizar diversas estrategias, como:

- **UTILIZAR FUENTES DE DATOS EXISTENTES**
- **APLICAR TÉCNICAS DE ESTIMACIÓN**
- **UTILIZAR MÉTODOS CUALITATIVOS**
- **AÑADIR EXPERIENCIA PERSONAL**
- **COMBINAR DATOS REALES Y SINTÉTICOS**
- **UTILIZAR INTELIGENCIA ARTIFICIAL**

Los datos sintéticos son datos que imitan datos del mundo real, pero que no fueron creados por humanos. Se generan mediante algoritmos de computación y simulaciones basadas en tecnologías de inteligencia artificial generativa.

Los datos inexactos pueden causar problemas en la cadena de suministro, aumentar los costos operativos y generar retrasos en los procesos logísticos.

C. ESCALABILIDAD Los problemas de escalabilidad han aumentado significativamente con el rápido crecimiento de la industria del comercio electrónico: se requieren motores de

recomendación modernos para generar resultados en tiempo real para aplicaciones a gran escala. En otras palabras, el rendimiento del modelo de recomendación se mide en términos de rendimiento (número de inferencias por segundo) y latencia (tiempo para cada inferencia).

La escalabilidad es la capacidad de un sistema o empresa para adaptarse a nuevas demandas y requisitos, sin perder valor ni rendimiento:

- **EN SISTEMAS INFORMÁTICOS**

La escalabilidad es la capacidad de un sistema para cambiar su tamaño o configuración para adaptarse a nuevas circunstancias. Para escalar un sistema, se puede agregar hardware adicional o actualizar el hardware existente.

- **EN EMPRESAS**

La escalabilidad es la capacidad de una empresa para adaptarse a los cambios del mercado sin perder valor. El objetivo es aumentar el volumen de producción sin aumentar los costos empresariales.

- **EN EL DESARROLLO DE SOFTWARE**

La escalabilidad es la capacidad de un sistema para adaptarse a un aumento en la carga de trabajo sin comprometer su rendimiento.

La escalabilidad es un concepto fundamental en el desarrollo de software, ya que es crucial para el éxito a largo plazo de cualquier proyecto tecnológico.

Para medir la escalabilidad de un negocio, se puede comparar los ingresos y los gastos durante un periodo determinado de tiempo.

La escalabilidad en software es la capacidad de un sistema para seguir funcionando de manera eficiente a pesar de un aumento en el número de usuarios o solicitudes.

Para escalar un sistema, se puede:

- Agregar hardware adicional
- Actualizar el hardware existente sin modificar mucho la aplicación
- Mejorar el rendimiento de un componente ya creado

La escalabilidad es un aspecto importante del diseño de software y está relacionada con la capacidad de mantener la aplicación y la consistencia de sus datos.

Para asegurar la escalabilidad en el desarrollo de software, se pueden seguir algunas mejores prácticas, como: Planificar anticipadamente, Utilizar metodologías ágiles, Mantener una documentación detallada, Automatizar tareas repetitivas, Evaluar continuamente la arquitectura.

Existen dos tipos de escalabilidad: horizontal y vertical. La escalabilidad vertical consiste en añadir más potencia a los recursos actuales, mientras que la escalabilidad horizontal significa añadir más recursos para dividir la carga.

D. DIVERSIDAD Y NOVEDAD Los resultados más precisos que pueden obtener los motores de recomendación se basan en la similitud de usuarios u objetos.

Esto, sin embargo, expone al usuario a una selección más limitada de elementos (los populares), mientras que los elementos de nicho altamente relacionados pueden pasarse por alto. La diversidad de recomendaciones permite a los usuarios descubrir elementos que no encontrarían fácilmente.

Creación de una arquitectura escalable para un recomendador



- **FILTRADO BASADO EN CONTENIDO:** Analiza las características de los productos o materiales consumidos previamente por el usuario para recomendar elementos que compartan atributos similares.

Es un método para recuperar información que utiliza las características de los elementos para seleccionar y devolver los más relevantes para una consulta.

En este tipo de filtrado, las predicciones se basan en el producto, en lugar del usuario. Esto significa que se utilizan las características del artículo, como la marca, el precio, las calificaciones, el tamaño, la categoría, etc., para hacer las recomendaciones.

Un ejemplo de filtrado basado en contenido es un sistema de recomendaciones de un servicio de música en streaming.

En este caso, las canciones son el "producto" y los datos de los que se dispone son el grupo, el cantante, la discográfica y el género.

Los sistemas de recomendación o recomendadores son flujos de trabajo con IA que sugieren productos, contenido o servicios a los usuarios. Los recomendadores de contenido son el tipo de sistema de recomendación de IA más común y se utilizan en plataformas de streaming y de comercio electrónico.

- **FILTRADO HÍBRIDO:** Combina enfoques de filtrado colaborativo y basado en contenido para mejorar la precisión de las recomendaciones.

En el caso de la adquisición de materiales para proyectos de ingeniería, un sistema de recomendación puede ofrecer sugerencias precisas sobre los materiales necesarios en función del consumo histórico y los patrones de uso en proyectos anteriores. Este tipo de sistema permite optimizar el inventario, reducir los costos de almacenamiento y minimizar el riesgo de desabastecimiento, al anticiparse a las necesidades específicas de cada proyecto. Al aplicar técnicas de Machine Learning, estos sistemas se vuelven aún más efectivos, dado que pueden ajustar las recomendaciones en tiempo real y adaptarse a las variaciones en la demanda o en los patrones de consumo.

En Ingeomega S.A.S, un sistema de recomendación basado en el análisis de datos históricos y patrones de consumo de materiales puede ser clave para implementar una estrategia de compras inteligentes.

1.2 DESCRIPCIÓN DE CASO DE ESTUDIO.

La contratista Ingeomega S.A.S Ubicada en la ciudad de Medellín como sede principal, se dedica a ofrecer soluciones en diversas áreas de la ingeniería como lo son las telecomunicaciones, eléctricas y sibil. En su operación diaria ha visto la necesidad de optimizar el proceso de adquisición de los materiales para los diferentes proyectos, de manera que esta compra se pueda a justar a los diferentes grupos de trabajo, esta estrategia de compras inteligentes se enfoca en llevar un control preciso de los materiales empleados en los distintos proyectos lo que les permitiría reducir costos por compras excesivas y posibles desabastecimientos que podrían afectar la continuidad de los proyectos.

El estudio de caso se caracteriza por:

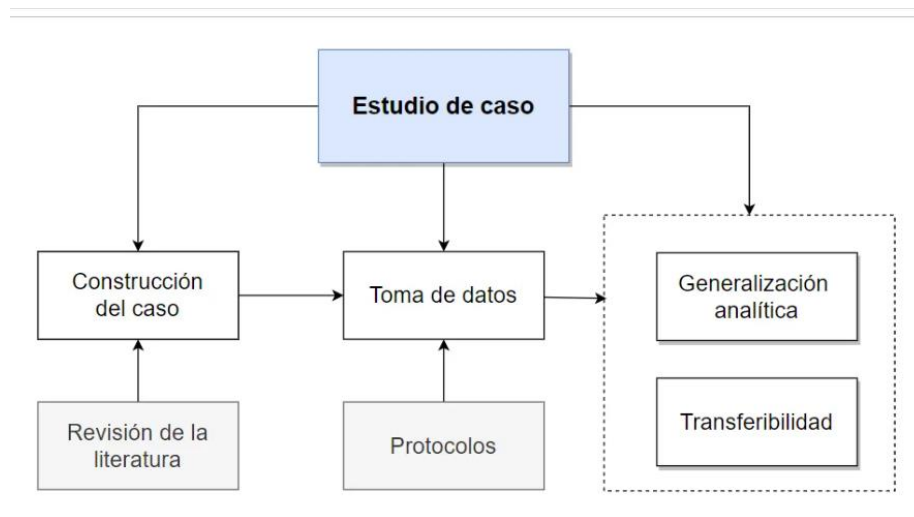
- Ser un método cualitativo, aunque a veces se utilizan métodos cuantitativos
- Centrarse en un caso específico, que puede ser una persona o un grupo
- Identificar las características comunes a otras situaciones, así como las que hacen que el caso sea particular
- Demostrar cómo estas características influyen en un sistema, conjunto de personas o eventos

Para escribir un estudio de caso, se puede seguir el siguiente procedimiento:

- Definir el objetivo del caso
- Preparar los datos
- Identificar a las personas involucradas

- Escribir el primer borrador
- Solicitar retroalimentación
- “Probar” el caso en clase

En un estudio de caso descriptivo, se describen las percepciones de los personajes principales, las soluciones que consideraron, el proceso de implementación, los resultados, y el estado actual del problema y la solución.



1.3 PREGUNTA PROBLEMA:

¿Cómo la implementación del Machine Learning para el análisis del consumo de materiales puede mejorar la eficiencia en el abastecimiento y reducir el riesgo de desabastecimiento en Ingeomega S.A.S

1.4 HIPÓTESIS:

El realizar comparaciones de consumo de los años anteriores y el uso de Machine Learning permitiría que Ingeomega S.A.S pudiera predecir con mayor precisión las necesidades futuras para cada una de sus proyectos, optimizando las compras y reduciendo los costos

1. OBJETIVOS

2.1 OBJETIVO GENERAL.

Implementar un modelo de análisis de datos utilizando Machine Learning para optimizar la estrategia de compras de Ingeomega S.A.S mediante el análisis de consumos históricos de material con el fin de reducir los costos y mejorar la planificación de inventario y evitar el desabastecimiento de los diferentes proyectos de la empresa

2.2 OBJETIVOS ESPECÍFICOS.

- Desarrollar un modelo de análisis de datos con Machine learning para predecir el consumo de materiales
- Implementar un sistema de gestión de inventario basado en el patrón de consumo.
- Evaluar la reducción de costos tras la implementación del modelo
- Implementar un sistema para evitar el desabastecimiento de material en los proyectos de la empresa.

2. DESARROLLO E IMPLEMENTACIÓN DEL APRENDIZAJE

3.1 PREPARACIÓN Y ANÁLISIS DE LOS DATOS

El documento a analizar hace referencia al consumo de material de la empresa INGEOMEGA S.A.S una empresa ubicada en la ciudad de Medellín que presta servicios de ingeniería y que tiene como cliente a Empresas públicas de Medellín EPM , en la región del bajo cauca antioqueño en el área de control de perdidas, la información del archivo está conformado por 9 columnas y 16496 datos, lo que permite realizar un análisis del uso del material lo que facilita la identificación de patrones y tendencias en la utilización de los recursos.

Parte 1.



The screenshot shows a Google Colab notebook with the following Python code in the first cell:

```
#Para cargar los datos
import pandas as pd
from google.colab import files
uploaded = files.upload()
for filename in uploaded.keys():
    Conjunto_Datos = pd.read_csv(filename, sep=',')
Conjunto_Datos.head()
```

The second cell shows the file upload process and a preview of the CSV data:

Elegir archivos: consumo-cpn3.csv
 • consumo-cpn3.csv(text/csv) - 2154410 bytes, last modified: 2/11/2024 - 100% done
 Saving consumo-cpn3.csv to consumo-cpn3.csv

Item	Encargado	Cuadrilla	Fecha Ejecución	Municipio	Descripción Material	Código	Unidad	Cantidad	
0	1.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CABLE COBRE AISLADO CON NEUTRO CONCENTRICO 1X8...	200410	Metro	4
1	2.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CAJA HERMETICA SENCILLA ALOJAR MEDIDOR ENERGIA...	211829	Unidad	2
2	3.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CIERRE ACERO INOXIDABLE 1/2 PG PARA CINTA META...	211374	Unidad	2
3	4.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CINTA DE ACERO INOX DE 1/2" RLL DE33M	211370	Unidad	0.06
4	5.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CONECTOR DERIVACION COMPR (10/10) O 4 A 4 ACSR...	212846	Unidad	2

Próximos pasos: [Generar código con Conjunto_Datos](#) [Ver gráficos recomendados](#) [New interactive sheet](#)

La imagen presenta una estructura detallada de los datos registrados en el archivo de consumo de materiales correspondiente al contrato CPN3, utilizado en la gestión de materiales de Ingeomega S.A.S. Cada campo incluido en el archivo cumple una función específica para facilitar el análisis y seguimiento del consumo de materiales en los proyectos de la empresa.

- **Artículo:** Este campo identifica el nombre o tipo de material utilizado, permitiendo una fácil clasificación y búsqueda dentro del inventario.
- **Encargado:** Designa a la persona responsable de supervisar y gestionar el uso de los materiales, lo que facilita la trazabilidad y el control de las operaciones.
- **Cuadrilla:** Se refiere al grupo de trabajo asignado para la ejecución de tareas específicas, lo que permite asociar el consumo de materiales a diferentes equipos dentro del proyecto.

- **Fecha de ejecución:** Indica el día en que se utilizó el material, siendo un dato clave para el análisis temporal del consumo, la planificación y la anticipación de futuros requerimientos.
- **Municipio:** Proporciona información sobre la localización geográfica del uso de materiales, útil para proyectos distribuidos en diferentes regiones, en este caso en el bajo Cauca antioqueño.
- **Descripción del material:** Ofrece un detalle sobre las características específicas del material, aportando información adicional que ayuda a evitar errores en la selección de insumos.
- **Código:** Es un identificador único que facilita la búsqueda rápida y la clasificación precisa de cada material en el sistema de gestión.
- **Unidad de medida:** Define la medida en la cual se contabiliza el material (por ejemplo, kilogramos, metros), asegurando precisión en la cuantificación.
- **Cantidad:** Indica el volumen o número de unidades de cada material utilizado, esencial para el cálculo de inventario y el monitoreo del gasto.

Esta estructura de datos es fundamental para implementar un modelo de análisis de Machine Learning que optimice el inventario y las compras de materiales. Al analizar estos campos, el sistema puede identificar patrones de consumo, predecir la demanda futura y recomendar cantidades óptimas para cada proyecto, lo que contribuye a reducir costos y minimizar riesgos de desabastecimiento.

The screenshot shows a Google Colab notebook with the following Python code:

```
#Para cargar los datos
import pandas as pd
from google.colab import files
uploaded = files.upload()
for filename in uploaded.keys():
    Conjunto_Datos = pd.read_csv(filename, sep=',')
Conjunto_Datos.head()
```

Below the code, a file named 'consumo-cpn3.csv' is listed. The table below represents the data loaded from this file:

Item	Encargado	Cuadrilla	Fecha Ejecución	Municipio	Descripción Material	Código	Unidad	Cantidad	
0	1.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CABLE COBRE AISLADO CON NEUTRO CONCENTRICO 1X8...	200410	Metro	4
1	2.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CAJA HERMETICA SENCILLA ALOJAR MEDIDOR ENERGIA...	211829	Unidad	2
2	3.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CIERRE ACERO INOXIDABLE 1/2 PG PARA CINTA META...	211374	Unidad	2
3	4.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CINTA DE ACERO INOX DE 1/2" RLL DE33M	211370	Unidad	0.06
4	5.0	GEIDER DE JESUS TORO TAPIAS	DIAGNOSTICO	1/3/2022	NECHI	CONECTOR DERIVACION COMPR (10/10) O 4 A 4 ACSR...	212846	Unidad	2

At the bottom of the notebook, there are navigation buttons: "Generar código con Conjunto_Datos", "Ver gráficos recomendados", and "New interactive sheet".

IMAGEN 1:

INFORMACIÓN CONTENIDA POR ARCHIVO CONSUMO DE MATERIAL CPN3

La imagen muestra una transformación clave en el archivo de consumo de materiales, donde ciertos campos de datos —como encargado, cuadrilla, código y cantidad— se han convertido a formato numérico. Esta conversión facilita el procesamiento y análisis automatizado de la información, permitiendo que herramientas de análisis de datos y modelos de Machine Learning puedan interpretar y manipular los datos de manera eficiente.

- **Encargado:** Representado numéricamente, este campo asigna un valor único a cada responsable, lo que simplifica la identificación y permite un análisis rápido de patrones de consumo asociados a distintos encargados.
- **Cuadrilla:** La conversión de las cuadrillas a valores numéricos permite identificar cada grupo de trabajo de manera uniforme. Esto es útil para analizar la demanda y

el consumo de materiales a nivel de cuadrilla y permite optimizar la asignación de recursos.

- **Código:** Al convertir los códigos de materiales en números, el sistema puede organizar y procesar estos datos de manera más eficaz, lo cual es esencial para realizar comparaciones y agrupaciones de materiales en el análisis.
- **Cantidad:** Mantener la cantidad en formato numérico permite la ejecución de cálculos automáticos, como el total de materiales utilizados o el promedio de consumo por proyecto. Este campo es crítico para establecer patrones de demanda y proyectar necesidades futuras.

Esta transformación a formato numérico es fundamental para el análisis estadístico y predictivo, ya que simplifica el proceso de detección de patrones y el desarrollo de modelos de Machine Learning. Al tener estos datos en un formato numérico, el sistema puede aplicar técnicas avanzadas de análisis y predicción, optimizando así la gestión de inventario, la planificación de compras y la reducción de costos asociados al consumo de materiales en los proyectos de Ingeomega S.A.S.

Analizando los datos:

```

#Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16497 entries, 0 to 16496
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   Item                                  16496 non-null  float64
1   Encargado                             16496 non-null  object
2   Cuadrilla                             16496 non-null  object
3   Fecha Ejecución                       16496 non-null  object
4   Municipio                             6972 non-null   object
5   Descripción Material                  16496 non-null  object
6   Código                               16496 non-null  object
7   Unidad                               16496 non-null  object
8   Cantidad                              16497 non-null  object
dtypes: float64(1), object(8)
memory usage: 1.1+ MB

```

IMAGEN 2:

ESTRUCTURA DE LOS DATOS

La imagen nos muestra como está conformado los datos del archivo consumo de material CPN3, la cual se divide en 9 columnas y 16497 datos, también podemos evidenciar que los datos son tipo Object y para poder realizar un buen análisis debemos convertirlos en número.

Elegir archivos | transformed_...mo_cpn3.csv

- transformed_consumo_cpn3.csv(text/csv) - 1433443 bytes, last modified: 2/11/2024 - 100% done

Saving transformed_consumo_cpn3.csv to transformed_consumo_cpn3 (3).csv

Item	Encargado	Cuadrilla	Fecha Ejecución	Municipio	Descripción Material	Código	Unidad	Cantidad
0	1.0	7	0	1/3/2022	NECHI CABLE COBRE AISLADO CON NEUTRO CONCENTRICO 1X8...	200410.0	Metro	4.00
1	2.0	7	0	1/3/2022	NECHI CAJA HERMETICA SENCILLA ALOJAR MEDIDOR ENERGIA...	211829.0	Unidad	2.00
2	3.0	7	0	1/3/2022	NECHI CIERRE ACERO INOXIDABLE 1/2 PG PARA CINTA META...	211374.0	Unidad	2.00
3	4.0	7	0	1/3/2022	NECHI CINTA DE ACERO INOX DE 1/2' RLL DE33M	211370.0	Unidad	0.06
4	5.0	7	0	1/3/2022	NECHI CONECTOR DERIVACION COMPR (10/10) O 4 A 4 ACSR...	212846.0	Unidad	2.00

Próximos pasos: [Generar código con Conjunto_Datos](#) [Ver gráficos recomendados](#) [New interactive sheet](#)

IMAGEN 3:

DATOS CONVERTIDOS A NUMERO

En la siguiente imagen se evidencia que los datos como el encargado, la cuadrilla, el código y la cantidad están en formato número.

```

#Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16497 entries, 0 to 16496
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Item                  16496 non-null  float64
1   Encargado              16497 non-null  int64
2   Cuadrilla              16497 non-null  int64
3   Fecha Ejecución       16496 non-null  object
4   Municipio              6972 non-null   object
5   Descripción Material   16496 non-null  object
6   Código                 16454 non-null  float64
7   Unidad                 16496 non-null  object
8   Cantidad                16496 non-null  float64
dtypes: float64(3), int64(2), object(4)
memory usage: 1.1+ MB

```

IMAGEN 4:

NUEVA ESTRUCTURA DE LOS DATOS

La imagen presenta la estructura actualizada del archivo de consumo de material CPN3, en la cual se observa que los datos están organizados en 9 columnas y un total de 16,497 registros. Esta disposición permite un manejo eficiente y preciso de la información para el análisis y procesamiento de datos, preparándola para técnicas avanzadas de Machine Learning y análisis predictivo.

La estructura de datos se divide en tres tipos principales:

- **3 DATOS DE TIPO FLOAT:** Este tipo de dato, con decimales, es útil para registrar cantidades o mediciones que requieren precisión numérica, como las cantidades de material utilizado.
- **2 DATOS DE TIPO ENTERO:** Los datos enteros se emplean para valores que no necesitan decimales, como códigos de identificación o números de cuadrilla, permitiendo una representación simple y eficaz para el procesamiento.
- **4 DATOS DE TIPO OBJECT:** Este tipo de dato representa cadenas de texto o categorías, como nombres de artículos, descripciones de materiales o asignaciones de municipio. Estos datos permiten identificar y clasificar los materiales y ubicaciones, manteniendo un control descriptivo sobre cada registro.

Esta estructura de datos es clave para el análisis ya que permite realizar operaciones matemáticas precisas sobre los campos numéricos y facilita la categorización en los campos de texto. Al contar con datos bien organizados y en el tipo adecuado, el sistema puede aplicar técnicas de Machine Learning para predecir necesidades de inventario, optimizar compras y reducir riesgos de desabastecimiento en los proyectos de Ingeomega **S.A.S.**

```
[34] #Análisis de los datos
Conjunto_Datos.describe()
```



	Item	Encargado	Cuadrilla	Código	Cantidad
count	16496.000000	16497.000000	16497.000000	16454.000000	16496.000000
mean	8248.500000	11.408014	1.326544	216867.040537	5.080452
std	4762.129356	8.810156	0.655125	33836.482545	12.023414
min	1.000000	0.000000	0.000000	376.000000	0.000000
25%	4124.750000	3.000000	1.000000	200410.000000	1.000000
50%	8248.500000	9.000000	1.000000	211374.000000	2.000000
75%	12372.250000	21.000000	2.000000	219404.000000	4.000000
max	16496.000000	27.000000	3.000000	298239.000000	426.000000

IMAGEN 5:

RESUMEN ESTADÍSTICO DE LA INFORMACIÓN

Análisis estadístico de los datos del archivo de consumo de material del contrato

La imagen presenta un análisis estadístico de los datos contenidos en el archivo de consumo de material asociado al contrato. Este análisis permite obtener una visión general de la distribución y cantidad de datos, lo cual es útil para evaluar el comportamiento y uso de materiales. Los resultados más relevantes incluyen:

- **COUNT (CANTIDAD TOTAL DE ENTRADAS):** Representa el número total de registros en el archivo, con 16,496 datos. Este valor indica la cantidad de instancias que contienen información sobre el consumo de materiales en distintos momentos o proyectos, proporcionando una base sólida para el análisis de patrones de consumo.

- **MEAN (CANTIDAD PROMEDIO):** Indica la cantidad promedio de los valores, situándose en 8,242. Este valor representa el promedio del uso de materiales o de otros datos cuantitativos en el archivo, lo cual es útil para establecer una referencia en el consumo habitual.
- **MIN (CANTIDAD MÍNIMA):** Muestra el valor más bajo en los registros, que es de 1 dato. Este dato mínimo puede representar el menor consumo de un material en una entrada específica, ayudando a identificar casos de bajo uso o posibles excepciones.
- **MAX (CANTIDAD MÁXIMA):** Señala el valor más alto registrado, que es de 16,496 datos. Este dato máximo se refiere al total de entradas en el archivo y proporciona un límite superior para cualquier análisis.

Este análisis estadístico básico permite comprender la variabilidad y el rango en los datos de consumo de materiales, lo cual es esencial para el desarrollo de modelos predictivos. Estos resultados ayudan a Ingeomega S.A.S a identificar patrones de uso, optimizar la planificación de compras y establecer estrategias de inventario adecuadas para evitar desabastecimientos y reducir costos en sus proyectos.

```
[35] #Quitando columnas indeseadas
Conjunto_Datos=Conjunto_Datos.drop(['Item','Fecha Ejecución','Municipio','Unidad','Descripción Material'],axis=1)
#resumen de los datos
Conjunto_Datos.head()
```

	Encargado	Cuadrilla	Código	Cantidad
0	7	0	200410.0	4.00
1	7	0	211829.0	2.00
2	7	0	211374.0	2.00
3	7	0	211370.0	0.06
4	7	0	212846.0	2.00

IMAGEN 6:

ELIMINACIÓN DE DATOS INDESEADOS

En la imagen se evidencia la eliminación de datos indeseados como lo fueron: Ítem, Fecha Ejecución, Municipio, Descripción del material, Unidad, ya que no son relevantes para la toma de información del material consumido en el contrato.

Tabla de Frecuencia:			
	Intervalo	Frecuencia Absoluta	Frecuencia Acumulada
0	[0.0, 42.6)	16247	16247
1	[42.6, 85.2)	180	16427
2	[85.2, 127.8)	48	16475
3	[127.8, 170.4)	13	16488
4	[170.4, 213.0)	3	16491
5	[213.0, 255.6)	2	16493
6	[255.6, 298.2)	1	16494
7	[340.8, 383.4)	1	16495
8	[383.4, 426.426)	1	16496
9	[298.2, 340.8)	0	16496

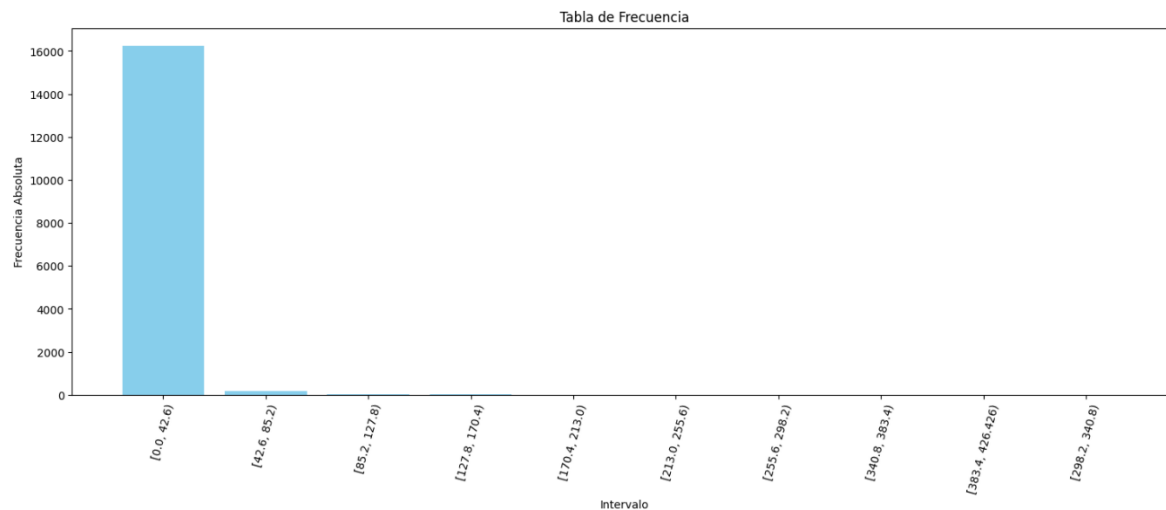


IMAGEN 7:

TABLA DE FRECUENCIA

En la tabla se muestra cómo se distribuyen las cantidades de material consumidos.

Donde la mayoría de los consumos son pequeños y se encuentran en un rango de 0 a 42.6 lo que demuestran que en general se consumen pocas cantidades por cada labor a realizar.

Mientras que los consumos más altos son muy pocos y se encuentran en un rango de 42.6 y 85.2 lo que evidencia que son pocas las labores que requieren gran cantidad de material.

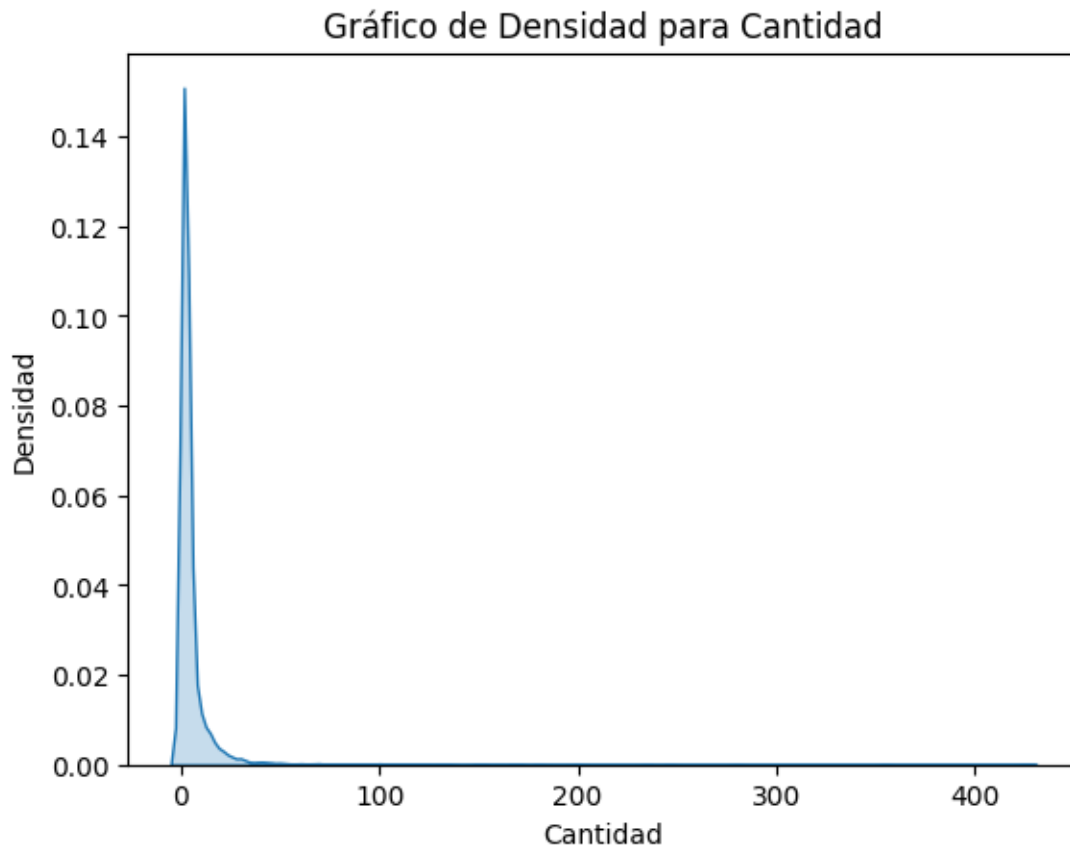


IMAGEN 8:

GRAFICO DE LA DENSIDAD DE CONSUMO DE MATERIAL

El grafico nos muestra el consumo de materiales en este proyecto, el cual este concentrado en pequeñas cantidades y son pocas las actividades que consumen grandes cantidades de material.

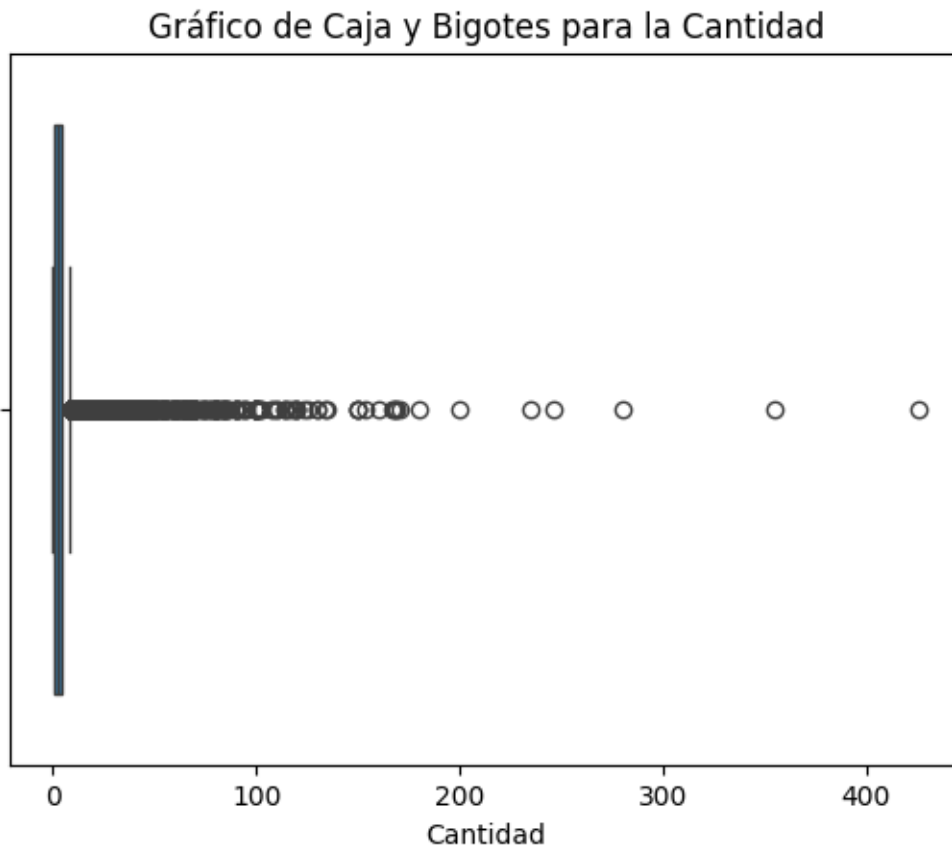


IMAGEN 9:

GRÁFICO DE CAJA Y BIGOTES

El gráfico muestra la distribución del consumo de materiales en el proyecto. La mayoría de los materiales se utilizan en pequeñas cantidades; sin embargo, también se observan algunos puntos alejados que indican un consumo elevado, aunque de forma esporádica. Esto sugiere que, aunque el consumo usual es bajo, en ciertos momentos se requieren grandes cantidades de material para necesidades específicas del proyecto.

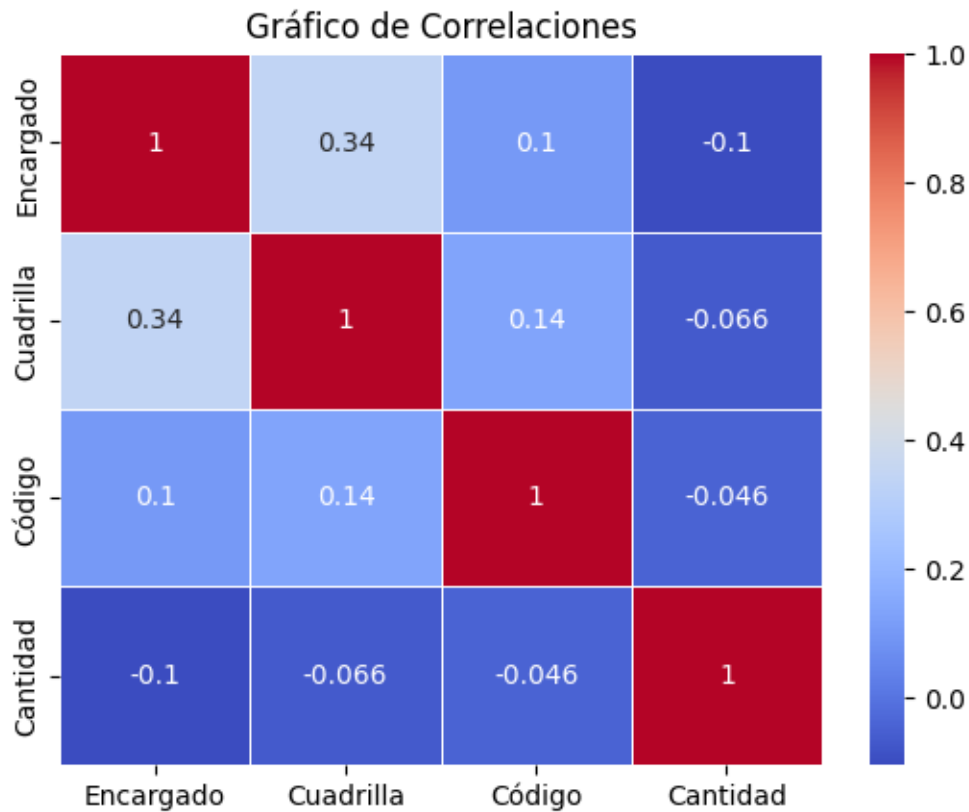


IMAGEN 10:

MATRIZ DE CORRELACIÓN

La matriz indica que existen relaciones débiles entre las variables, lo que sugiere que la cantidad de material consumido no está significativamente relacionada con el encargado o la cuadrilla asignada. Esto implica que el consumo de materiales parece depender de otros factores externos, más allá de quién los gestiona o del equipo de trabajo, y destaca la necesidad de explorar otras variables para comprender mejor los patrones de uso en los proyectos.

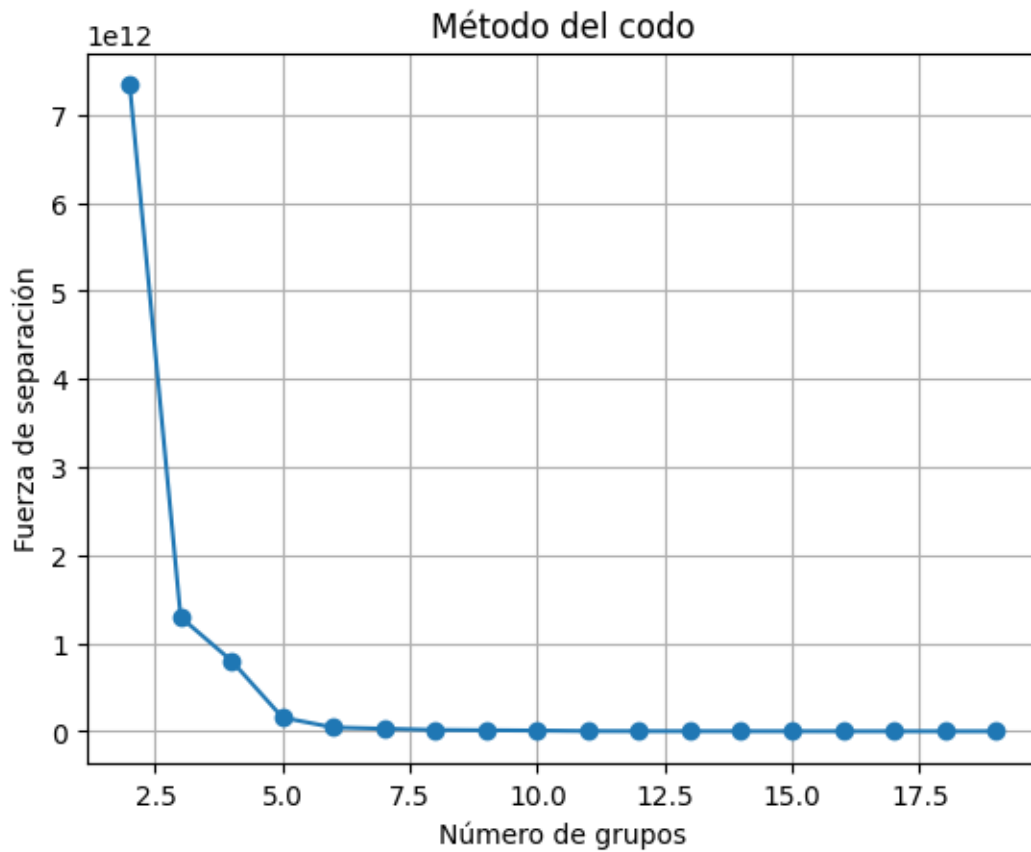


IMAGEN 11:

MÉTODO DEL CODO

El gráfico nos muestra que al dividir los datos en 3 o 4 grupos obtenemos una buena separación de los datos dentro de cada grupo, esto nos permitiría capturar las principales diferencias entre los datos sin hacer clasificaciones complejas.

Con la información del consumo de materiales se pueden implementar varias estrategias para la optimización de la compra de los materiales en la empresa, enfocándonos en los

patrones de consumo, asignación de recursos y agrupación de datos para mejorar la eficiencia, algunas de estas estrategias podrían ser.

IDENTIFICAR LOS MATERIALES DE CONSUMO FRECUENTE: la mayoría de los materiales se consumen en pequeñas cantidades con algunos picos de consumo en materiales específicos, en la empresa se podría establecer un inventario mínimo de estos materiales ya que son de uso recurrente, estas compras se podrían realizar trimestral o semestralmente con el fin de disminuir costos y evitar desabastecimiento de material.

COMPRAS SEGÚN EL CONSUMO HISTÓRICO: Analizando la distribución y densidad de consumo de materiales puede ayudar a predecir la demanda, los picos en el gráfico de densidad muestran una fuerte concentración de consumo, estos datos se pueden usar en un sistema de prevención para realizar las compras en relación al histórico de consumo.

OPTIMIZACIÓN DEL ALMACÉN: en la empresa se podría establecer diferentes zonas de almacenamiento según el consumo de material consumo alto, medio y bajo, con un movimiento constante para los materiales de alto consumo.

REALIZAR COMPRAS BASADAS EN EL MÉTODO CODO: según el gráfico del método codo los datos de consumo los podríamos dividir en tres grupos alta, media y baja demanda, esto permitiría una política de compra diferenciada, realizando compras

frecuentes de los materiales de alta demanda, compras periódicas para los de media y compras bajo pedido para los de baja demanda.

3.3 ANÁLISIS DE DESEMPEÑO

El análisis de desempeño del modelo evalúa su precisión en la predicción de consumo de materiales y su eficacia en la optimización de inventario y compras. Se utilizan métricas como el error medio absoluto (MAE), el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE) para medir la desviación entre las predicciones del modelo y los datos reales. Un bajo valor en estas métricas indica que el modelo está capturando los patrones de consumo de manera adecuada.

Además, se identifican materiales de consumo frecuente, caracterizados por picos ocasionales en el uso de ciertos materiales específicos. Esto sugiere que establecer un inventario mínimo y realizar compras trimestrales o semestrales para estos materiales podría reducir costos y evitar desabastecimientos. La distribución de consumo y los picos en el gráfico de densidad también respaldan el uso de datos históricos para anticipar la demanda y prevenir el agotamiento de existencias.

3.4 VALIDACIÓN DEL MODELO

La validación del modelo asegura su robustez y capacidad de generalización, y se realiza mediante validación cruzada o dividiendo los datos en conjuntos de entrenamiento y prueba. Consistentemente buenos resultados de validación indican que el modelo puede predecir el consumo de manera confiable para nuevos datos.

Basándose en el método del "codo", los datos de consumo se dividen en tres grupos (alta, media y baja demanda), lo que permite políticas de compra diferenciadas. Los materiales de alta demanda se compran con frecuencia, los de demanda media de forma periódica, y los de baja demanda solo bajo pedido. Además, se podría optimizar el almacén estableciendo zonas de almacenamiento según el nivel de consumo (alto, medio y bajo), manteniendo una rotación constante de los materiales de alta demanda.

3. CONCLUSIONES Y TRABAJOS FUTUROS

Este estudio evidencia el impacto positivo de aplicar modelos de Machine Learning y técnicas de análisis predictivo en la optimización del inventario y las estrategias de compra de materiales en Ingeomega SAS. La implementación de un sistema de recomendación basado en consumo histórico permite identificar patrones clave en el uso de materiales, lo que facilita una gestión de inventario más precisa y una política

de compras diferenciada. Este enfoque no solo reduce los costos asociados a las compras excesivas, sino que también mitiga el riesgo de desabastecimiento, garantizando la continuidad de los proyectos de ingeniería. Además, la segmentación de materiales en función de su demanda (alta, media y baja) y la optimización del espacio de almacenamiento contribuyen a una mayor eficiencia en la operación de la empresa, alineándose con sus objetivos de sostenibilidad y rentabilidad.

El análisis de desempeño y validación del modelo confirma que los algoritmos de Machine Learning pueden predecir el consumo de materiales con un alto grado de precisión. Sin embargo, aún existen áreas que podrían beneficiarse de una investigación adicional. El presente trabajo se limita al análisis de patrones históricos sin considerar factores externos que puedan afectar la demanda, como variaciones estacionales, cambios en los costos de los materiales o condiciones de mercado. La incorporación de estas variables podría mejorar la capacidad del modelo para responder a fluctuaciones inesperadas en el consumo.

TRABAJOS FUTUROS

Para futuros estudios, se sugiere explorar enfoques avanzados de Machine Learning, como redes neuronales y modelos de aprendizaje profundo, que podrían captar relaciones no lineales y mejorar la precisión de las predicciones en entornos dinámicos. Asimismo, sería valioso integrar el análisis de datos en tiempo real para ajustar el inventario de forma adaptativa, considerando factores externos como variaciones en el

mercado y condiciones climáticas que puedan afectar los proyectos. Finalmente, la creación de un sistema de monitoreo continuo y de autoaprendizaje permitiría que el modelo evolucione con los cambios en las operaciones de la empresa, optimizando aún más la gestión de recursos y el control de inventario a largo plazo.

REFERENCIAS BIBLIOGRÁFICAS

Chigara, B., y Smallwood, J. (2014). Estrategias de gestión de costos empleadas por contratistas de construcción en proyectos en Zimbabwe. *Revista internacional de ingeniería y tecnología de construcción sostenible*

Gallego-García, S., y García-García, M. (2020). Sistemas de aprovisionamiento optimizados para reducir roturas de stock y minimizar costes en fabricación. *Revista Europea de Ingeniería y Ciencias Formales* , 10(4), 8734.

Aggarwal, CC (2016). *Sistemas de recomendación: el libro de texto*. Saltador. Este libro proporciona una base completa sobre sistemas de recomendación, a incluir técnicas de aprendizaje automático aplicables en la predicción de demandas.

Russell, S. y Norvig, P. (2021). *Inteligencia artificial: un enfoque moderno* (4^a ed.). Pearson. Un recurso fundamental sobre inteligencia artificial que explica los modelos predictivos aplicables en la optimización de inventario.

Chopra, S., Meindl, P. y Kalra, DV (2016). *Gestión de la cadena de suministro: estrategia, planificación y operación* (6^a ed.). Pearson. Este libro explora estrategias de gestión de inventarios y compras, analizando técnicas para reducir costos y evitar desabastecimientos.

Geron, A. (2019). *Aprendizaje automático práctico con Scikit-Learn, Keras y TensorFlow: conceptos, herramientas y técnicas para construir sistemas inteligentes* (2^a ed.). Medios O'Reilly. Una referencia práctica para la

implementación de modelos de Machine Learning en proyectos reales, ideal para entender cómo aplicar algoritmos en la predicción de consumo.

□ Plata, EA, Pyke, DF y Thomas, DJ (2016). *Gestión de inventario y producción en cadenas de suministro* (4ª ed.). Prensa CRC. Este texto profundiza en métodos avanzados de gestión de inventarios, incluyendo pronósticos de demanda y estrategias para la optimización de almacenes y políticas de compra.

Las tablas y figuras debes ponerlas en la página donde las mencionas por primera vez en el texto.

Si tu tabla o figura es muy grande debes ponerla en una página separada.

No debes dejar espacios en blanco en las páginas de texto, pero es posible dejar espacio en blanco en páginas que solo contienen tablas y figuras.

Los títulos de las tablas deben ser puestos sobre las mismas y en las figuras deben ponerse debajo. Todas las tablas deben tener al menos dos columnas y la fila de títulos.

TABLA 1.

MODELO DE DATOS

Id	Género	Edad	Ingreso

