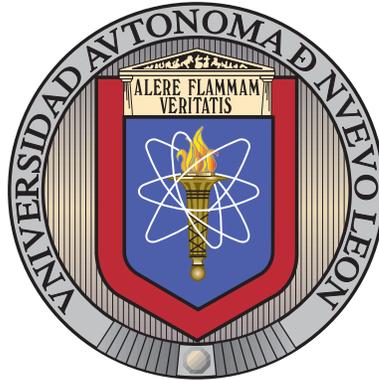


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO



CLASIFICACIÓN DE COMPONENTES CRÍTICOS  
BASADOS EN TÉCNICAS DE CLUSTERIZACIÓN

POR

FÁTIMA ABYGAIL DRAGUSTINOVIS LÓPEZ

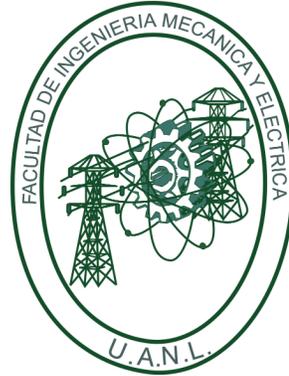
COMO REQUISITO PARA OBTENER EL GRADO DE  
MAESTRÍA EN LOGÍSTICA Y CADENA DE SUMINISTRO

JULIO 2024

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO



CLASIFICACIÓN DE COMPONENTES CRÍTICOS  
BASADOS EN TÉCNICAS DE CLUSTERIZACIÓN

POR

FÁTIMA ABYGAIL DRAGUSTINOVIS LÓPEZ

COMO REQUISITO PARA OBTENER EL GRADO DE  
MAESTRÍA EN LOGÍSTICA Y CADENA DE SUMINISTRO

JULIO 2024

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**  
**Facultad de Ingeniería Mecánica y Eléctrica**  
**Posgrado**

Los miembros del Comité de Evaluación de Tesis recomendamos que la Tesis “Clasificación de componentes críticos basados en técnicas de clusterización”, realizada por el/la estudiante Fátima Abygail Dragustinovis López, con número de matrícula 1545768, sea aceptada para su defensa como requisito parcial para obtener el grado de Maestría en Logística y Cadena de Suministro.

**El Comité de Evaluación de Tesis**

Dr. Leonardo Gabriel Hernández Landa  
Director

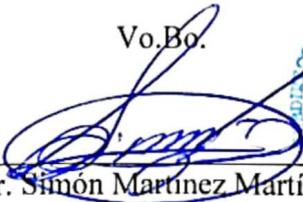
Dra. Carolina Solís Peña  
Co-director

MLyCS Blanca Idalia Pérez Pérez  
Revisor

Dra. Jania Astrid Saucedo Martínez  
Revisor

Dr. Federico Guadalupe Figueroa Garza  
Revisor

Vo.Bo.

  
Dr. Simón Martínez Martínez  
Subdirector de Estudios de Posgrado



Institución 190001

Programa 642597

Acta Núm. 4384

Ciudad Universitaria, a 05 de julio del 2024.

*Quiero dedicar esta tesis a quiénes supieron de mi dedicación, esfuerzo, sacrificio y trabajo. A quiénes me impulsaron a estudiar un posgrado para tener un futuro brillante. Papi & Mami, ¡los amo!*

*A mis hermanos, porque aguantaron esas noches/madrugadas con luz prendida en el cuarto, cuando debía investigar y hacer tareas, gracias por apoyarme y por su incondicional cariño y compañía.*

*A mi esposo, por haber sido paciente en este proceso y por apoyarme cuando lo necesitaba.*

*Y por último a Dios, por ser bondadoso con todas mis peticiones y por brindarme sabiduría, paciencia y madurez para este proyecto, estoy totalmente agradecida.*

# ÍNDICE GENERAL

---

<b>Agradecimientos</b>	<b>x</b>
<b>Resumen</b>	<b>xii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Descripción del problema . . . . .	3
1.2. Objetivo . . . . .	6
1.3. Hipótesis . . . . .	6
1.4. Justificación . . . . .	6
1.5. Metodología propuesta . . . . .	7
1.6. Estructura de la tesis . . . . .	8
<b>2. Marco Teórico</b>	<b>9</b>
<b>3. Revisión de literatura de la metodología</b>	<b>20</b>
3.1. Diseño de la investigación . . . . .	21
3.2. Análisis de información preliminar . . . . .	22

---

3.3. Técnicas de análisis y clasificación comunes en inventarios . . . . .	24
3.3.1. Clasificación ABC. . . . .	25
3.3.2. Clasificación XYZ . . . . .	25
3.4. Método de clasificación de inventarios basado en análisis de clústers .	26
3.4.1. Técnica de comparación de los métodos . . . . .	35
<b>4. Aplicación y resultados de la metodología</b>	<b>36</b>
4.1. Comprobación de eficiencia del modelo con bases de datos existentes .	38
4.2. Implementación con el método de clasificación de inventarios con clustering. . . . .	44
4.2.1. Limpieza y selección inicial de los datos . . . . .	44
4.2.2. Determinación de las variables con mejor desempeño para clustering. . . . .	47
4.2.3. Obtención de categorías mediante clustering . . . . .	48
4.2.4. Comparativa entre métodos . . . . .	58
<b>5. Conclusiones</b>	<b>64</b>
5.1. Contribuciones . . . . .	65
5.2. Trabajo futuro . . . . .	66
<b>A. APÉNDICE</b>	<b>67</b>
A.1. Librerías utilizadas en programación R. . . . .	67
A.1.1. Librerías utilizadas . . . . .	67

# ÍNDICE DE FIGURAS

---

2.1. La cadena de suministro . . . . .	10
4.1. Clasificación de las especies de iris basado en las dos variables con mayor importancia . . . . .	41
4.2. Clasificación de las especies de iris basado en las dos variables con menor importancia . . . . .	41
4.3. Clúster obtenido basado en longitud y amplitud del pétalo. . . . .	42
4.4. Histograma de costo de inventario por componentes sin función logarítmica. . . . .	50
4.5. Histograma de costo de inventario por componentes con función logarítmica . . . . .	52
4.6. Dendrograma final después del tratamiento de los datos . . . . .	55
4.7. Clústers adquiridos considerando costo vs rotación . . . . .	56
4.8. Clúster de k-medias . . . . .	57
4.9. Clúster de k-medias basado en costo vs rotación . . . . .	57
4.10. Distribución de los grupos generados con el método ABC . . . . .	60
4.11. Distribución de los grupos generados con el método XYZ . . . . .	61

---

4.12. Distribución de los grupos generados con el método de clustering jerárquico . . . . .	62
4.13. Distribución de los grupos generados con el método clustering de k-medias . . . . .	63

# ÍNDICE DE TABLAS

---

4.1. Base de datos iris . . . . .	39
4.3. Niveles de importancia de las variables en términos de calidad de clúster según el coeficiente de Gini . . . . .	40
4.4. Resultado de la matriz de confusión con especies de iris. . . . .	43
4.5. Base de datos <i>clustering</i> . . . . .	45
4.6. Base de datos "CLUSTERING" . . . . .	47
4.7. Variables importantes aplicando el coeficiente de GINI de <i>random forest</i> . . . . .	48
4.8. Tabla comparativa de métodos clustering . . . . .	58

# AGRADECIMIENTOS

---

Quiero empezar agradeciendo a Dios, por haberme dado salud, perseverancia, inteligencia, paciencia y madurez en este episodio de mi vida profesional. Este proyecto definitivamente fue retador, pero gracias a su gracia, me permitió incrementar mis estudios, habilidades y conocimientos en una rama la cuál me gusta y anhelaba obtener.

También quiero agradecer a mi familia, mis padres y mis hermanos. Papá y mamá fueron mis principales alentadores en el deseo de estudiar un posgrado, esto para alcanzar un nivel más alto en mi carrera profesional y para aspirar a un mejor puesto de trabajo en mi vida. Me comentaban que los conocimientos nunca están de más, al contrario nutren y ayudan a uno mismo a razonar, desarrollar habilidades, ser más creativos y competitivos, en otras palabras me harían ver las cosas de diferente manera.

A mis hermanos, quiénes me apoyaron y me motivaban en seguir adelante, cuando sentía que ya no podía. Ellos hicieron más ligero el camino, pues debía combinar el trabajo, y la escuela al mismo tiempo y no se diga también, las levantadas temprano, los tráficos y las desveladas, eran los que más me impactaban en mi cuerpo y salud. Ustedes me escuchaban, me alentaban a seguir adelante, y toleraron las noches de desvelo con luz prendida en el cuarto.

Gracias por entenderme cuando no podía salir los fines de semana a paseos con ustedes, gracias por ayudarme a hacer los lonches para el trabajo y a realizar los deberes en casa, sabía que no tenía el tiempo suficiente para realizar todo durante

el día, pero fueron comprensibles y me ayudaron a cubrir esas actividades.

Inmensamente gracias por todo el apoyo, su comprensión y cariño en estos dos años de maestría.

También quiero agradecer enormemente a mis tutores el Dr. Leonardo Gabriel Hernández Landa y la Dra. Carolina Solis Peña, sin ustedes esta tesis no hubiera salido adelante. Su conocimiento, sus habilidades, su paciencia, su asistencia, su tiempo, su orientación y su responsabilidad conmigo en el desarrollo de este proyecto, fue parte primordial para que yo pudiera cumplir y graduarme, obteniendo el grado de maestría.

Había días que los molestaba mucho y se que estaban muy ocupados, pero agradezco completamente que siempre se hayan dado el tiempo para responderme, regresar las tareas y documentos firmados que les solicitaba y por aceptar las sesiones ya en la noche cuando sabía que ya querían descansar. A todo, ¡muchas muchas gracias!

Y por último, a mis amistades del trabajo quienes fueron comprensivas en apoyarme cuando debía salir temprano. También, a Daniel, empezamos de novios en el tiempo que estudiaba, gracias por tu comprensión y tu apoyo moral en acabar mis estudios y gracias por no romper tu promesa del compromiso matrimonial que te pedí, hasta que acabara la escuela.

A todos, ¡gracias y les deseo las mejores bendiciones!

# RESUMEN

---

Fátima Abygail Dragustinovis López.

Candidato para obtener el grado de Maestría en Logística y Cadena de Suministro.

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: CLASIFICACIÓN DE COMPONENTES CRÍTICOS BASADOS EN TÉCNICAS DE CLUSTERIZACIÓN.

Número de páginas: 73.

**OBJETIVOS Y MÉTODO DE ESTUDIO:** Este trabajo habla de la clasificación de componentes de un caso de estudio de una empresa dedicada al giro automotriz, cuya característica es detectar cuales componentes son los más críticos, tomando en cuenta sus variables más importantes y características similares, con la finalidad de obtener una planeación adecuada de los inventarios, para cuando existen demandas fluctuantes.

El método de estudio usado para este caso fue, principalmente obtener nuestra información mediante una base de datos generales, de donde obtuvimos información de cada componente y adicional, se usan las técnicas de clusterización por medio de herramientas como el *machine learning*, que son aprendizajes automáticos por medio de computadora donde se desarrollan modelos y algoritmos para tomar decisiones usando patrones especiales.

**CONTRIBUCIONES Y CONCLUSIONES:** La contribución de esta investigación favorece a profesionales, investigadores y empresas o almacenes de cualquier giro industrial, ya que les brindará una visión más amplia para considerar variables diferentes a las del precio y la demanda. Este método les permitirá mejorar la planificación y el

control de inventarios, evitando mermas y gastos innecesarios que puedan afectar la cadena de suministro y la rentabilidad de las empresas.

Y como conclusión, los resultados muestran que los métodos de *clustering* jerárquico y k-medias proporcionan una clasificación más precisa y detallada al considerar variables adicionales, en comparación con los métodos ABC y XYZ. El k-medias es efectivo al optimizar la varianza dentro de cada clúster y la distancia entre ellos, y la elección del número adecuado de clústers es crucial. El *clustering* jerárquico también ofrece una clasificación significativa y uniforme, con una similitud del 90 % con los clústers generados por k-medias, lo que refuerza la validez de ambos métodos. En contraste, los métodos ABC y XYZ resultan menos precisos al no considerar variables adicionales importantes.

Firma del asesor: \_\_\_\_\_  
Dr. Leonardo Gabriel Hernández Landa

Firma del co-asesor: \_\_\_\_\_  
Dra. Carolina Solís Peña

## CAPÍTULO 1

# INTRODUCCIÓN

---

Los inventarios son de gran importancia, ya que gracias a su disponibilidad, podemos cumplir con un plan de producción o bien una orden de compra, que satisfará a un cliente.

Sin embargo, el control de este conlleva un rol muy importante en el ámbito que sea. Considera un almacén de una empresa, la cantidad de inventario que puede este llegar a soportar, puede variar con base a su modelo de planificación y control de inventarios.

Como por ejemplo, si los encargados del almacén empiezan a observar que la cantidad de materiales que reciben, ya no caben en su localización, comentarán que se está pidiendo un exceso y no hay buena planificación. Y, si el caso fuese contrario a este, que cuando el materialista de la línea de producción pida que se le surta un componente y no le entreguen nada, dirán que no hay y que es una mala planificación y control de inventario.

Con estos dos escenarios, las posibilidades de pérdidas monetarias son altas, debido a que se resume en que no, se cuenta con una adecuada planificación y control de inventarios.

Por lo que esta práctica, de tener un control sobre los inventarios, es nece-

saria y primordial para las empresas. Algunas de las razones por las cuáles no se tienen control sobre los inventarios, es debido a que no existe una clasificación de los componentes que manejan.

Algunas de las razones que las empresas reconocen, por la cuál no clasifican adecuadamente los inventarios son, debido a las altas fluctuaciones de clientes, los nuevos productos que pueden introducirse, bajos niveles de servicio por proveedores que surten, el tamaño de lote y/o el costo de los componentes.

El control de los inventarios es una práctica necesaria en todos los ámbitos, por dar un ejemplo, en nuestro hogar el saber cuanto tengo de alimentos para que me cubra mi semana; hasta para un pequeño local que venda insumos, necesita saber cuánto tiene de cada producto, para evitar que se le acabe y saber cuando pedir más para que siga vendiendo.

En las empresas grandes cuando hablamos de fabricantes de productos, ya sean textiles, eléctricos, plásticos, o hasta automotrices, es común que ellos cuenten con sistemas de control de inventarios debido a la gran cantidad de materias primas y productos que reciben y fabrican para su venta. Sin embargo, existen problemas para identificar la gran cantidad de componentes con los que trabajan, ya que son extensos y no se tiene un buen control de inventario.

El control de inventarios según a como lo especifica Douglas (2000), es una forma de planificar y controlar los materiales y productos con la que una empresa cuenta. Sin embargo, en ocasiones el no tener un control sobre los inventarios de un almacén, puede llegar a perjudicar a la empresa y así mismo al cliente. Según a como lo describe Mora García (2000) algunos aspectos negativos que puede llegar a tener el mal control y gestión de inventarios es la pérdida de materiales por falta de flujo, también puede generar la falta de material y esto causaría el incumplimiento con las órdenes del cliente, lo cuál generaría perder al cliente y que busque otro proveedor que pueda cumplir; y otro aspecto negativo de la falta de control de inventarios sería el robo o extravío de mercancía dentro de la empresa, por mencionar algunos.

Este trabajo tiene como objetivo principal, el demostrar como clasificar inventarios e identificar cuando tenemos componentes críticos, para tomar decisiones correctas en nuestras empresas.

Según Ehrhardt y Brigham (2007) el inventario debe ser administrado eficientemente y sus dos objetivos fundamentales son:

1. *Garantizar con producto o material disponible, la operatividad de la empresa.*
2. *Conservar los niveles necesarios que permitan minimizar los costos totales (de pedido y de mantenimiento).*

¿Y porque es necesario tener un control de inventarios? Algunas de las consecuencias que se pueden recalcar es que se evitarían las mermas, adicional el consumo de almacenamiento innecesario, obsolescencias, re-trabajos, costos de fletes y costos financieros por mencionar algunos.

## 1.1 DESCRIPCIÓN DEL PROBLEMA

La cadena de suministro es el conjunto de todas aquellas actividades donde están incorporados proveedores, transportistas, manufactureros, almacenes y el cliente final, así lo cita Gómez Hernández (2022).

Las empresas manufactureras trabajan con base a órdenes de ventas para la producción de los ensambles. Es por eso que las empresas tienen establecidos sus procesos de manufactura y en este se abordará el caso de una empresa que trabaja bajo un sistema *pull*, o bien denominado en español como jalar, el cuál se adapta en producir solo lo que se va a vender, así lo cita Limón DelosSantos (2015).

Así mismo, la producción *pull* se esfuerza por eliminar la sobreproducción y es uno de los tres componentes principales de un sistema esbelto, así como la producción de justo a tiempo. La forma de ejecutar dicho sistema, es a través de una tarjeta

*kanban*, el cuál es colocado en un pizarrón donde después, logística estará escaneando para dar la señal a producción del modelo y la cantidad que tendrá que producir, así lo menciona Limón DeloSantos (2015).

Cuando es emitido un reajuste a la producción por parte del cliente, se generan fluctuaciones, ya sea por incrementos o reducciones de la demanda y es necesario realizar el ajuste a los planes de producción de inmediato, para ver el impacto que puede tener, tanto en productividad de personal, capacidad de producción e inventarios de materias primas/ componentes y/o ensambles finales.

Estos cambios drásticos en los planes de producción generan problemas a corto plazo, por mencionar falta de componentes ó exceso y/u obsolescencias de inventarios. Esto no solo afecta localmente a la empresa, si no también, a toda la cadena de suministro que está integrada por otros proveedores, con los cuales se tienen que negociar cambios, cancelar contratos, órdenes de compra y pagar penalizaciones.

Cuando tenemos un mercado volátil y las fluctuaciones existen de por medio, trabajando bajo un sistema de producción tipo *pull*, genera una gran pérdida económica, sí, no se mantienen planes eficientes de abastecimiento y compra.

El problema que se establece en este caso de estudio del giro automotriz, es tan frecuente que, alrededor de un 30% a 35% de fluctuación es lo que sucede semanalmente con las órdenes del cliente, tanto como puede reducir también puede incrementar, pero el mayor impacto que sobresale cuando sucede esto, es con las reducciones, ya que genera una tensión en toda la cadena de suministro cuya solución no es inmediata.

Es por eso que, el problema que se establece es identificar cuáles son los componentes críticos de todos los ensambles finales que fabrica una empresa, para surtir las órdenes de los clientes. Ya que, las materias primas y componentes que se usan, no están clasificados, o identificados para un mejor control y planificación de inventarios, cuando las demandas son fluctuantes frecuentemente.

A continuación se ejemplifica lo antes mencionado, el productor recibe un requerimiento por la cantidad de 50,000 piezas, ellos fabrican esta cantidad de ensamblajes, una vez fabricadas lo embarcan hacia un centro de distribución en el cuál, es descargado y almacenado hasta que el cliente final solicite esta cantidad. Al momento de que el centro de distribución reciba la orden del cliente por las 50,000 piezas, ellos embarcan a cliente final y consecutivamente así como embarcan, nuevamente colocan una orden al productor para rellenar el inventario que acaban de vender.

Sin embargo, sucede que por alguna razón el cliente revisa su tendencia de ventas y prefiere reducirle un 50 % a su cantidad inicial, ya que el mercado no está solicitando lo que su pronóstico le mostraba .

El cliente final le notifica al centro de distribución que ya solamente va a requerir 25,000 piezas, por lo que esto le generará al centro de distribución reducir la orden que ya le había pedido al fabricante y como el fabricante ya se había preparado con materia prima y personal para producir las 50,000 piezas, al recibir esta reducción generará en su producción y en almacenamiento cambios imprevistos, que le afectarán de manera que se quede con inventario, y reduzca su plan de trabajo, hasta que nuevamente le sea solicitada una orden para consumir lo que previamente ya había fabricado y en consecuencia reducir espacio.

Estos cambios en las órdenes finales, afectan los pedidos y las entregas de los componentes y materias primas solicitadas, ya que, existen circunstancias como el tamaño de lote, el tiempo de vida de resinas que se caducan, e inclusive el *lead time*, el cuál es el tiempo que transcurre desde que una orden es introducida hasta el momento en el que el cliente desea recibir su pedido, además de esto también afecta, el tiempo de tránsito que tarda cierto medio de transporte ya sea marítimo, terrestre, aéreo o ferroviario. Por todas estas razones, difícilmente pueden reducirse y/o cancelarse órdenes de componentes que vienen en tránsito.

Una vez que llegan a nuestro almacén, como no serán utilizados inmediatamente debido a una reducción en el plan de producción, tendremos por consecuencia que

la política de inventario se excederá, debido a que ya no será el mismo consumo que se requería para las 50,000 piezas; por lo tanto al realizar la clasificación y agrupamientos de aquellos componentes que son críticos contemplando variabilidad en las demandas y variables como precio, tamaño, lote, y modo de transportación, etc. podremos mejorar la planificación y el control de inventarios, para reducir los impactos económicos que genera a la empresa estos cambios drásticos del mercado.

## 1.2 OBJETIVO

El objetivo de esta investigación es identificar los componentes de los ensambles finales que maneja una empresa, con el propósito de clasificar y agrupar con base a mayor criticidad, para mejorar la planificación y el control de los inventarios con el motivo de reducir costos logísticos de la operación y el inventario.

## 1.3 HIPÓTESIS

Si identificamos mediante técnicas de clasificación y agrupamientos, componentes con diferentes variables y que demuestren si son críticos, podremos mejorar la planeación de materiales y así mismo, tomar decisiones importantes que ayudarán a la empresa a tener un mejor control y gestión de sus inventarios.

## 1.4 JUSTIFICACIÓN

De acuerdo al objetivo señalado con anterioridad, de no tener una clasificación de los componentes dentro de una empresa, puede llegar a originar efectos negativos según lo ha mencionado Ruíz (2018), entre ellos se destacan los altos costos financieros y logísticos innecesarios, la obsolescencia de materiales, el sobre inventario, el

bajo nivel de servicio, las elevadas pérdidas y daños de material.

Por lo que, una gestión inadecuada de inventarios no solo pone en riesgo la estabilidad económica de la empresa, si no que, también afecta la administración y el control de sus procesos. La falta de conocimiento e identificación de las variables independientes que pertenecen al círculo de cada uno de los componentes y/o materiales de una empresa, conducen a pérdidas derivadas de una mala planificación de abastecimiento.

Es por eso que, el identificar y clasificar los componentes más críticos, resulta fundamental para mejorar el control sobre los inventarios, lo que contribuye a reducir problemas como excedentes, faltantes de material y obsolescencias. Y además de esto, permite a la empresa a ajustarse mejor a los cambios drásticos en los planes de producción, actualizando las políticas de inventario y tomando decisiones importantes basadas en variables identificadas.

## 1.5 METODOLOGÍA PROPUESTA

De acuerdo al caso de estudio, se pretende desarrollar una clasificación de inventarios críticos en contraste a la técnica de clasificación del ABC, con la cuál será desarrollada y comparada mediante una técnica de programación con un logaritmo, que pueda demostrar los agrupamientos de los componentes, tomando en cuenta diferentes variables.

La manera de ejecutar esto, es iniciando por el planteamiento del problema, después se hace una revisión de literatura, le continúa con la selección de la herramienta para ejecutar la investigación, después se define las hipótesis y variables, se ejecuta mediante la experimentación y por último, se presentan los resultados.

## 1.6 ESTRUCTURA DE LA TESIS

La siguiente tesis esta dividida de la siguiente manera.

El capítulo 1, empieza por la introducción en la cual describimos el problema del caso de estudio y lo integra el objetivo, la hipótesis, la justificación, la metodología propuesta y la estructura de la tesis. En el capítulo 2, se encuentra el marco teórico, donde se proporciona la base conceptual y teórica sobre la cual se apoya la investigación.

Continuando con el capítulo 3, se encuentra la revisión de literatura de la metodología, en donde se encuentra información de varias herramientas existentes y de la metodología que se usará para llevar acabo el caso de estudio y asegurar que los resultados sean válidos. Después, le sigue el capítulo 4 el cual es la aplicación y resultados de la metodología y en ella puedes encontrar los resultados obtenidos de la metodología y por último, en el capítulo 5 se encuentra la conclusión en donde también se encontrarán las contribuciones y el trabajo a futuro.

También contamos con el Apéndice, en la cual se describen las librerías utilizadas en la herramienta seleccionada, que fue por medio de programación R.

Y ya por último, también encontrarás la parte de la bibliografía.

## CAPÍTULO 2

# MARCO TEÓRICO

---

Describiendo un poco el proceso de la cadena de suministro, Chopra y Meindl (2008), la describen como todas aquellas partes involucradas de manera directa o indirecta, en la satisfacción de una solicitud de un cliente. La cadena de suministro incluye no solamente al fabricante y al proveedor, sino también a los transportistas, almacenistas, vendedores al detalle (o menudeo) e incluso a los mismos clientes, así como se muestra en la figura 2.1.

Con base a lo anterior, el primer eslabón es el de los inventarios, pero existen varias maneras de clasificarlas, de acuerdo a Nahmias (2007), los inventarios se dividen en:

- Materias primas. Son los recursos necesarios para la producción o procesamiento.
- Componentes. Pueden ser materias primas o subgrupos que se incluirán posteriormente en un producto final.
- Trabajo en proceso, por sus siglas en inglés (*WIP*). Son los inventarios que se encuentran en la planta esperando ser procesados.
- Artículos terminados. Son artículos que han terminado su proceso de producción y esperan ser embarcados.

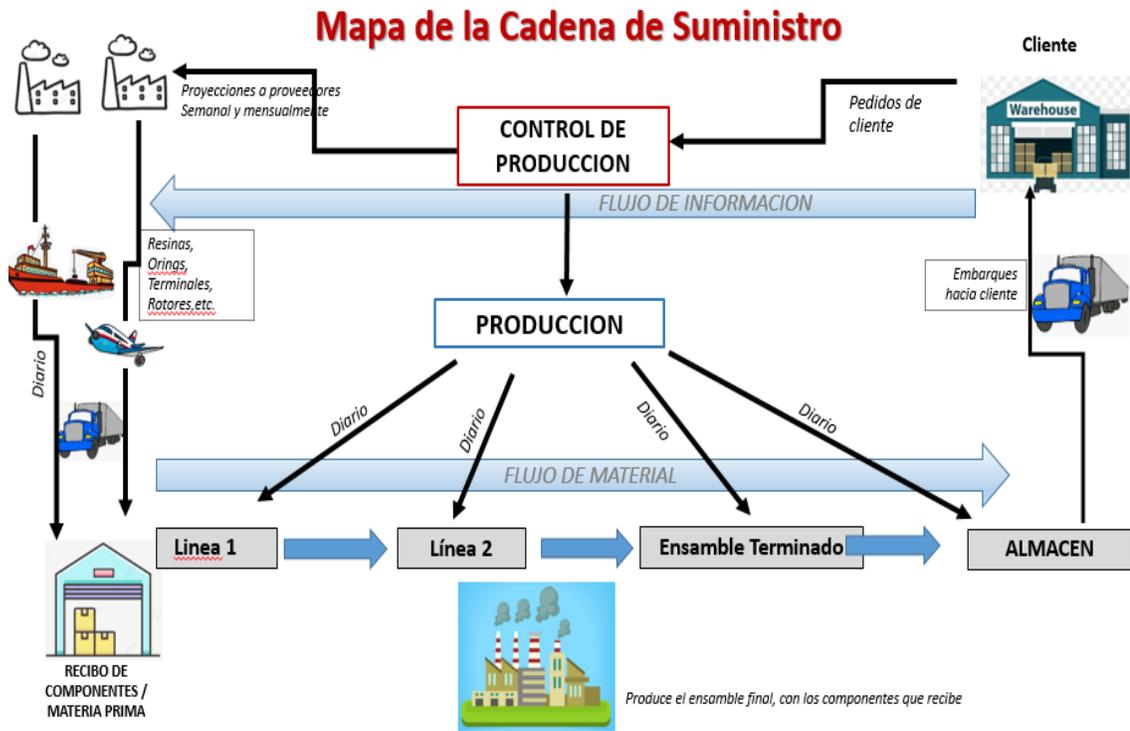


FIGURA 2.1: La cadena de suministro

Ya habiendo mencionado las clasificaciones anteriores, este caso de estudio se realiza dentro del sector automotriz, el principal objetivo es el de clasificar los inventarios que maneja una empresa dedicada a este sector, con el fin de detectar los componentes críticos y se pueda realizar una mejor planeación de estos.

Los inventarios en una empresa de cualquier ámbito industrial son los recursos más importantes, estos harán que toda la cadena de suministro se mueva, de lo contrario si no tenemos inventario de materia prima y componentes, no será posible la producción de las órdenes del cliente. Sin embargo, es necesario que estos estén clasificados correctamente para tener las cantidades necesarias y solo las correctas en tiempo y en forma, para evitar mermas u obsolescencias en una empresa.

A través de los años, las cadenas de suministro han ido evolucionando, primero empezaron por hacer la fabricación de las piezas en talleres, después paso a ser la producción en masa y por último y en la actualidad fue la producción esbelta, la

cuál es la más conocida y la que la mayoría de las empresas manufactureras del giro automotriz utilizan.

La producción esbelta es conocida por haber emergido de la empresa japonesa Toyota, la cuál se caracteriza por producir flexiblemente y en lotes pequeños, paros de línea por operarios cuando se detecta un defecto, comunicación con el personal, trabajadores capacitados para realizar multitareas y habilidades, visión a largo plazo, trabajo en equipo, una administración visible, y tener una producción mixta, según lo cita López Sánchez (2007).

Una cadena de suministro esbelta requiere mantener los inventarios controlados para satisfacer la demanda del cliente.

Para eso se requiere clasificar los inventarios y de acuerdo a como López Sánchez (2007) cita, las clasifica en 5 condiciones:

1. Por el tipo de distribución de la demanda, la cuál puede ser estocástica o determinista.
2. Por el número de productos o componentes a controlar por la política, uno o varios productos.
3. Por el número de instalaciones de distribución o suministro a controlar, es decir, una sólo localización o varias en la cadena de suministro.
4. Por el tipo de producto que controla de acuerdo a una clasificación ABC la cual está basada en la demanda uso dólar o peso anual.
5. Por la frecuencia de revisión de los niveles de inventario en el tiempo, que puede ser periódica o continua.

Por otro lado, Singh (2014) habla sobre la conducta del análisis *ratio*, el método ABC y VED análisis de los inventarios y menciona como el departamento de compras, sugiere las mejores técnicas para la compañía sobre tener un mejor control

del inventario. El cuál, si se identifica y se ponen en rango los componentes dentro del método ABC, los encargados de compras podrían negociar con proveedores el precio, cantidad de pedido, y la forma de envío, para que satisfaga las necesidades y la capacidad de los requerimientos que nuestros clientes nos demanden.

Las empresas manufactureras en la actualidad han ido cambiando debido al dinamismo del mercado existente. Antes de que sucediera la pandemia del COVID-19, los mercados se mantenían estables o existía dicha estabilidad, ya que el flujo económico de la sociedad existía. Sin embargo, la pandemia trajo consigo algunos aspectos negativos para el ámbito manufacturero, ya que diversos factores como la escasez de materias primas y componentes generó una reducción, si no es que un paro a varios sectores del mercado (M. *et al.*, 2020).

Por mencionar en el mercado del giro automotriz hoy en día, son de los más dinámicos a nivel global, esto debido a la falta de unidades disponibles para su venta según lo mencionan, Ewing y P (2021). Aunado a que los precios ahora son más caros de lo que normalmente estaban, la cadena de suministro de este sector incrementó todos sus precios, ya que la pandemia retrasó los tiempos de producción de todos los materiales con los que se fabrican los autos.

Recordemos que los vehículos son bienes que se devalúan sin embargo, siempre tienen algo nuevo que ofrecer, ya que año tras año van cambiando y actualizándose debido a las constantes tendencias, tecnologías, e innovaciones que existen en la actualidad.

Considerando que este sector trabaja a gran escala y maneja muchos materiales para fabricar un vehículo, las empresas automotrices deben tener un control sobre los inventarios de sus materias primas, para evitar sobre inventarios y pérdidas que los lleven nuevamente a una crisis económica.

La actualidad, demanda a las empresas automotrices y a cualquier empresa en particular, a cuidar su flujo económico. Parte de este flujo económico se encuentran principalmente en los inventarios que maneja una empresa.

Es por eso que las empresas necesitan manejar adecuadamente sus inventarios de producto terminado y materia prima, siempre que se administre y controle correctamente sus insumos.

Según Pandya y Thakkar (2016), definen a la gestión de inventarios como la práctica de desarrollar y gestionar los niveles de inventario de materias primas, productos en proceso y productos terminados, de manera que haya suministros adecuados y minimice el costo de las existencias. El control de inventario no siempre está relacionado con el ahorro financiero de los inventarios, sino que también afecta al ahorro de espacio, al ahorro de mano de obra, a la simplificación y transparencia del proceso, al aumento del flujo de producción y a la mejora de la imagen de la empresa.

Algunos de los modelos existentes según Pandya y Thakkar (2016) hacen referencia, a los modelos de análisis ABC, XYZ, HML, VED, FSN, SDF, GOLF y SOS. Y, si nos enfocamos en los de ABC y XYZ, son técnicas comunes de control de inventario utilizadas para reducir los costos de inventario y mejorar la gestión de existencias. El análisis ABC clasifica los productos en categorías A, B y C según la frecuencia de uso y su valor, mientras que el análisis XYZ evalúa las fluctuaciones en la demanda o el consumo de los artículos y los clasifica en categorías X, Y y Z según la variabilidad de la demanda.

Por otro lado, el trabajo de Flores y Whybark (1987) mencionan que uno de los métodos para controlar los inventarios es mediante el análisis ABC, que define a los artículos definidos como -A-, los que deben de tener una revisión más cercana que los artículos -C-, estos son los que no se les debería de invertir mucho tiempo en revisar, ya que son los menos requeridos.

Así mismo, Flores y Whybark (1987) denominan este criterio el más utilizado para identificar inventarios, y el cuál dirige a reducir inventarios y mejorar en las entregas programadas de los productos terminados. Adicional observaron que algunos importantes criterios que no son necesariamente monetarios surgieron, como son

los tiempos de tránsito, obsolescencias, disponibilidad, la sustitución y la criticidad, también son factores que afectan para clasificar de cierta manera un inventario.

Flores *et al.* (1992) extiende un poco su trabajo anterior a que, el método de proceso analítico jerárquico, integra variedad de criterios para categorizar ciertos artículos de inventarios. Ya que, si se consideran otras maneras de criterios, estos influyen en la administración de dichos inventarios. No necesariamente pueden ser clasificados como A, B y C, si no que dependiendo del tipo del giro de la empresa pueden categorizar a los componentes como -AA-, -AB- o -AC- o así hasta -BA-, -BB-, -BC-, -CA-, -CB- o -CC-, esto con la finalidad de determinar una importancia para cada uno que sea establecido.

Paredes Rodriguez *et al.* (2019) realizan una investigación en donde analizan una propuesta de controlar los inventarios de un almacén de repuestos para identificar los más importantes y considerando varias variables. Ellos comentan que, dos de las variables más importantes que se consideran para tomar en cuenta que política deben asignar para cada repuesto son, el costo y la criticidad de la pieza. Sin embargo, otras variables como el tamaño, la forma, y el impacto que puede tener cada repuesto, pueden ser de vital importancia para definir una política de inventario.

Otra investigación por parte de, Balaji y Senthil Kumar (2014), hablan sobre como la clasificación ABC de inventario multicriterio, es esencial para toda industria. Su investigación está basada en la industria del caucho dedicada a la fabricación de neumáticos y mencionan que el análisis ABC, está dividido por la clase A, la cuál está contribuida por la mayoría, en términos numéricos sería responsable del 70 % al 80 % del valor total del inventario de los elementos. La clase B se la atribuye del 10 % al 15 % y a la clase C solo el 5 % del valor total del inventario de los elementos. Sin embargo, cuando se habla del multicriterio, las empresas usan varios criterios, como el precio unitario, la tasa de demanda anual, la naturaleza crítica, la escasez, la durabilidad, etc.

El trabajo que Balaji y Senthil Kumar (2014) implementaron, habla del Proceso

Jerárquico Analítico (AHP), para estimar el valor del sistema de inventario, donde la clasificación se basa en la asignación de contenedores de los elementos del inventario.

También, Altay Guvenir y Erel (1998), en su investigación de inventarios multicriterios optan por usar un algoritmo genético al cuál le llaman GAMIC (Algoritmo Genético para la Clasificación Multicriterio de Inventario), su finalidad es optimizar los parámetros que representen un peso de criterio y que a final todos los criterios den un mismo valor de importancia. Para lograr obtener los pesos de los criterios, calculan las puntuaciones ponderadas de los elementos del inventario de manera similar al enfoque con el AHP. Luego, los elementos con puntuaciones superiores al valor de corte AB se clasifican como A, aquellos con valores entre AB y BC como clase B, y los elementos restantes se clasifican como clase C.

Por otro lado, Lollia *et al.* (2017) proponen un enfoque de clasificación de inventario de múltiples criterios (MCIC) basado en clasificadores supervisados, como los son los árboles de decisión y bosques aleatorios. Este método realiza una muestra de artículos que ha sido previamente clasificada mediante la simulación exhaustiva de un sistema de control de inventario predefinido. Su objetivo principal es la clasificación en automático de todos los elementos.

Al igual que el investigador anterior, quién también uso los nuevos métodos no supervisados por medio del *machine learning*, para mejorar los sistemas de clasificación de inventarios de diferentes puntos de vista es, Zowid *et al.* (2019), quiénes proponen usar el modelo de mezcla gaussiana (GMM), el cuál es un algoritmo utilizado en estadística y aprendizaje automático para, representar distribuciones de probabilidad que consisten en la combinación de varias distribuciones gaussianas individuales. Cada componente en el modelo, se asocia con una mezcla de pesos que determinan la contribución relativa de cada componente a la distribución final.

Este modelo es utilizado para realizar tareas de clasificación y agrupamiento de datos. Es por eso que este modelo tiene como primer lugar, el clasificar los *SKUs* (*stock keeping unit*) por sus siglas en inglés, el cuál en español significa el número

de referencia único de un producto, sin procesos subjetivos y como segundo objetivo garantizar un costo total de inventario competitivo junto con un nivel de servicio máximo. Los resultados con este modelo fueron prometedores y de carácter competitivo a comparación de otros ya existentes, así lo declararon en su investigación.

Se ha podido observar que varios de los autores anteriores usan modelos automatizados, y los más tradicionales como la clasificación ABC de inventarios en diferentes ámbitos industriales, por eso mismo, es que el trabajo realizado por Mills-tein *et al.* (2014) hablan de un proceso de optimización para determinar al azar grupos de inventario dependiendo del servicio y la familia a la que pertenece. Ya que, a diferencia de los modelos antes mencionados por otros autores, este mejora el enfoque de agrupamiento de inventario ABC al ofrecer soluciones integradas, automatizadas y optimizadas mediante el equilibrio entre el costo de inventario y la ganancia y asigna óptimamente el presupuesto de inventario a cada unidad de referencia (*SKU*). Aparte de que la herramienta propuesta, puede ayudar a los gerentes de las empresas a decidir, qué unidades de referencia serían mejor mantener fuera del inventario.

Complementando un poco la parte de la justificación de este trabajo, se espera resolver o mitigar los aspectos negativos que conlleva el sobre inventario de piezas que no se mueven o que, al contrario, se mueven mucho. Muchas empresas no están actualizadas a los nuevos tiempos y por eso, es que siguen considerado políticas de inventario con solo contemplar el aspecto monetario.

Sin embargo, esto va más allá de solo ver una variable. También lo citan Paredes Rodriguez *et al.* (2019), que las políticas deben de designarse con base a la necesidad y la criticidad de la pieza; en muchas ocasiones no tenemos información que asegure que las piezas van a durar por mucho tiempo o al contrario, que vayan a ser sustituidas por otras. También, no se tienen claros los tiempos de vida de uso y adicional los precios y costos de cada pieza, año tras año estarán cambiando debido a la inflación que los materiales puedan tener y así mismo, por la calidad que puedan

ofrecer.

Una vez ya hablado de las características que son importantes a considerarse para la clasificación de los inventarios, hablemos un poco de los modelos que usan los investigadores, profesionales y prácticos en materia.

Así lo cita Prasad (1994), en su investigación sobre los modelos y sistemas de clasificación de inventarios. Los modelos existentes, mejoran la rentabilidad de las empresas y organizaciones; además les ayuda a que se mantengan y sigan funcionando en el ambiente económico. Sin embargo, hace énfasis en que los gerentes de las empresas tienden por usar viejos y básicos métodos de clasificación de inventarios, el cuál contrae aspectos negativos como son los altos costos de retener mucho inventario y largos tiempo de entrega.

Pues bien, menciona que dependiendo de la cantidad de componentes en analizar, ya sea uno o múltiples, es necesario escoger el método apropiado para clasificar los inventarios, ya sean por sus características como el ambiente de la planeación y producción del producto y/o del mercado, los clientes y/o proveedores, son los más considerados. Los más comunes, son el de la cantidad económica de pedido, también conocido en inglés como *EOQ* y/o la planeación de materiales *MRP*, por sus siglas también en inglés.

Siguiendo con la clasificación del inventario y los nuevos métodos que ayudan a las empresas a considerar diferentes variables para clasificar inventarios, Babai *et al.* (2014) mencionan que la mayoría de las compañías que tienen una larga cantidad de números de parte y componentes, utilizan la clasificación ABC pero, que normalmente no funciona del todo cuando se administran demasiados números de parte, ya que, no hay un control adecuado.

Ellos hacen referencia a otros autores y mencionan, que cuando se trata de una clasificación de inventario para una pieza en particular, se toman en cuenta valores como la demanda y el volumen de esta misma. Ellos consideran que varios métodos muy eficientes en la actualidad como lo es el modelo de R, ayudan a desarrollar una

clasificación interesante y original, el cuál trata de maximizar los niveles de servicio y minimizar los costos de inventario. El modelo R, considera variables diferentes y tiene por resultado una optimización lineal.

La gestión adecuada de inventarios es un aspecto crucial para asegurar el funcionamiento eficiente de la cadena de suministro en el sector automotriz y cual sea el giro industrial. Los inventarios representan los recursos más valiosos para las empresas, ya que son los componentes esenciales para satisfacer las demandas de los clientes y mantener la producción en marcha.

Una mala gestión de inventarios puede llevar a costos innecesarios, mermas y problemas en la cadena de suministro. Se ha identificado la necesidad de clasificar los inventarios con el objetivo de detectar los componentes críticos y mejorar la planificación. La clasificación de inventarios se ha abordado a través de diversos métodos, desde los tradicionales como la clasificación ABC, hasta enfoques más avanzados basados en algoritmos genéticos y aprendizajes automáticos.

El análisis ABC ha sido una de las técnicas más comunes utilizadas para clasificar los inventarios según su importancia y valor en la cadena de suministro. Sin embargo, debido a la complejidad y el dinamismo del mercado actual, varios investigadores han propuesto enfoques más sofisticados que toman en cuenta múltiples criterios como la demanda, el costo, la criticidad, la durabilidad y la disponibilidad de los componentes.

El uso de modelos automatizados y algoritmos de aprendizaje automático, han demostrado ser efectivos para optimizar la clasificación de inventarios y ayudar a las empresas a tomar decisiones más informadas y precisas. Estos modelos permiten identificar patrones y tendencias en los datos, lo que facilita la identificación de los componentes más críticos y la asignación adecuada del presupuesto de inventario.

En conclusión, el estudio y la implementación de técnicas avanzadas de clasificación de inventarios pueden brindar una ventaja competitiva a las empresas, permitiéndoles enfrentar los desafíos del mercado actual de manera más efectiva y

asegurando una producción eficiente y rentable.

## CAPÍTULO 3

# REVISIÓN DE LITERATURA DE LA METODOLOGÍA

---

El clúster es una técnica de clasificación que sirve para poder detectar y describir subgrupos de sujetos o variables homogéneas en función de los valores observados dentro de un conjunto aparentemente heterogéneo, así lo citan Vilá Baños *et al.* (2014), en su artículo. Ellas describen que en el proceso de aplicación de esta técnica, el investigador debe tomar una serie de decisiones. Primero, se debe seleccionar las variables relevantes para identificar los grupos; como segundo paso, debe de establecer las medidas de similitud para controlar las medidas de proximidad entre las unidades y, como tercero, debe fijar qué procedimiento o método de agrupación se utilizará para configurar los conglomerados.

Con base en esto anterior, es que en esta sección, se propone una metodología basada en el uso de clústers, para desarrollar una clasificación robusta que permita manejar las variables de mayor interés para el experto. Esta metodología con base al caso de estudio se denomina « Clasificación de inventarios basada en análisis de clústers » .

Para llevar a cabo este proceso, se recomienda que el experto tenga un dominio y conocimiento profundo de las variables relacionadas con el inventario. En primer

lugar, se debe realizar un análisis de la información preliminar para determinar la población de datos y si es necesario, según la naturaleza de la población establecer una muestra representativa. Esto implica determinar el tamaño de muestra adecuado y utilizar técnicas de muestreo técnico.

Tomando como referencia lo anterior, la población usada corresponde a todos los números de parte que se manejan en la empresa donde se está llevando a cabo el caso de estudio, estos números de parte están en una base de datos conformada por el número de parte, su descripción, la unidad de medida, el precio, el tamaño de lote, por mencionar algunas; en total son 156 números de parte que son usados para diferentes ensambles finales. Sin embargo, el 15% de su totalidad no fueron usados, ya que, una de sus características no permitía que la técnica funcionara adecuadamente. La característica que genera inconformidad es la unidad de medida. La mayor parte de los componentes tenían la unidad de medida en piezas, pero las descartadas tenían la unidad de medida en kilogramos, gramos, o metros; por lo tanto, para la muestra solo se tomaron en cuenta los componentes cuya unidad de medida era en piezas.

Una vez obtenida la muestra, se procede a la preparación de la información y se aplican las técnicas de agrupamiento propuestas en la metodología. Finalmente, se lleva a cabo la validación y se realiza una comparación con otras metodologías conocidas, para evaluar la clasificación obtenida.

A continuación, se presenta de manera detallada la parte metodológica.

### 3.1 DISEÑO DE LA INVESTIGACIÓN

Dentro de este apartado, podemos definir la población de estudio, como todos aquellos números de parte que pertenecen a las listas de materiales de ensambles finales específicos, en un proceso de manufactura de partes automotrices, o bien de cualquier otro proceso cuyo interés sea el estudio.

Y por otro lado, tenemos la muestra que consiste en los datos históricos de años recientes, en relación con el nivel de inventario y rotación de los números de parte, en el proceso de manufactura que se está analizando. Se consideran únicamente las listas de componentes que se manufacturan actualmente, excluyendo aquellas que ya no se fabrican.

Como técnicas de muestreo, se puede definir a la técnica de muestreo estratificado, donde se seleccionan intencionalmente las listas de componentes actuales para su análisis.

Y en el tamaño de la muestra, será igual al número total de componentes relevantes, dado que la metodología propuesta requiere analizar todos los componentes que pertenecen a las listas seleccionadas y los datos históricos.

## 3.2 ANÁLISIS DE INFORMACIÓN PRELIMINAR

El análisis de información preliminar es un proceso esencial en cualquier investigación o proyecto que requiera el examen de datos y fuentes relevantes. Proporciona una visión general de los datos disponibles y establece las bases para una investigación más detallada y exhaustiva. En esta sección, se exploran los fundamentos del análisis de información preliminar, su importancia y cómo se lleva a cabo.

En primer lugar, el análisis de información preliminar se realiza con el objetivo de identificar y evaluar la calidad de los datos disponibles antes de emprender un estudio más profundo. Permite a los investigadores obtener una comprensión inicial de los datos, su alcance, su fiabilidad y sus posibles limitaciones. Al realizar un análisis preliminar, se pueden identificar lagunas en la información o áreas que requieren una investigación más exhaustiva.

El análisis de información preliminar implica varias etapas, como la recopilación de datos relevantes, la organización y la revisión inicial. Durante esta fase,

se pueden utilizar diversas herramientas y técnicas para examinar la información, como gráficos, tablas, estadísticas descriptivas y técnicas de visualización de datos. Esto nos permitirá identificar patrones, tendencias y relaciones iniciales entre las variables, lo que puede ayudar a definir áreas clave de interés y establecer hipótesis iniciales.

Es importante destacar que el análisis de información preliminar, no pretende llegar a conclusiones definitivas o tomar decisiones basadas únicamente en esta fase inicial. Su objetivo principal es, proporcionar una base sólida para el análisis posterior y la interpretación adecuada de los datos. El análisis preliminar ayuda a los investigadores a plantear preguntas relevantes, desarrollar hipótesis más sólidas y diseñar estrategias de investigación más efectivas.

Esta etapa es crucial. Podremos comprender inicialmente los datos disponibles y establecer las bases para algo más detallado. Además, identificaremos patrones, tendencias y lagunas en la información, lo que ayuda a definir áreas clave de interés y desarrollar estrategias más efectivas.

En el análisis preliminar de la metodología se realizan los siguientes pasos:

1. **Importación y limpieza de datos.** Estos se obtienen de la fuente, en este caso, un archivo de Excel generado por el sistema ERP de la empresa. Estos datos se importan en el lenguaje de programación R, donde se llevará a cabo el análisis. La importación y limpieza de los datos son necesarias para asegurar la calidad de los mismos y eliminar cualquier información innecesaria o inconsistente.
2. **Análisis de datos exploratorio.** Se realiza un análisis de datos exploratorio para, examinar las variables de la tabla de la base de datos que sean relevantes para el análisis. Aquellas variables que no aporten información relevante, se descartan. Este análisis permite obtener una comprensión inicial del comportamiento de los datos y calcular medidas de tendencia para la información

relevante.

3. **Selección de variables críticas.** Una vez realizado el análisis de datos exploratorio, se seleccionan las variables de interés. En primer lugar, se eligen aquellas variables que son relevantes para la clasificación ABC. Posteriormente, se seleccionan las variables necesarias para la técnica de clasificación propuesta en esta investigación, con el objetivo de realizar comparativas entre las técnicas tradicionales y las propuestas.
4. **Análisis de correlación.** Realiza un análisis de correlación entre las variables seleccionadas para identificar posibles relaciones y dependencias entre ellas. Esto te ayudará a comprender mejor cómo se relacionan los diferentes componentes y cómo pueden influir en la clasificación y agrupación de los inventarios.

### 3.3 TÉCNICAS DE ANÁLISIS Y CLASIFICACIÓN COMUNES EN INVENTARIOS

En esta parte, pasamos a lo que son el análisis de datos y gestión de inventario mediante las técnicas más utilizadas. Las siguientes dos que se describirán aquí, son valiosas para organizar y priorizar los elementos en función a su importancia. Y, es por eso que en esta sección se exploran las técnicas de clasificación ABC y XYZ, para obtener el resultado de optimizar y gestionar los inventarios y se lleguen a tomar buenas decisiones .

Tanto la clasificación ABC como la clasificación XYZ, son métodos de análisis efectivos para tomar decisiones informadas sobre la gestión de inventario, la asignación de recursos y la optimización de costos. Estas técnicas permiten a las empresas identificar y priorizar los elementos más críticos o aquellos que requieren una mayor atención y control, lo que puede resultar en una mejora significativa en la eficiencia operativa y la rentabilidad.

### 3.3.1 CLASIFICACIÓN ABC.

La clasificación ABC es una técnica ampliamente utilizada para clasificar los artículos en función de su importancia o valor, según lo describen Pandya y Thakkar (2016). Considerando la información y la metodología propuesta, se aplica el método ABC de la siguiente manera:

- Seleccionar las variables relevantes para el análisis ABC y construir una tabla que contenga esta información para cada artículo.
- Establecer factores y porcentajes de clasificación para el análisis ABC. Estos factores pueden estar relacionados con el valor de los artículos, la demanda o cualquier otra métrica relevante que se guste tomar en cuenta.
- Establecer los intervalos correspondientes en las categorías ABC, donde se identificará el valor mínimo y máximo de cada categoría. Estos intervalos se calcularán con base a los factores y porcentajes establecidos previamente.
- Asignar el puntaje de clasificación a cada artículo, utilizando los resultados de los factores de ponderación y teniendo en cuenta la categoría ABC a la que pertenece el artículo. Este puntaje puede ser utilizado posteriormente para la toma de decisiones relacionadas con la gestión de inventarios.

### 3.3.2 CLASIFICACIÓN XYZ

Así mismo Pandya y Thakkar (2016), mencionan que otra técnica para clasificar artículos en función de su frecuencia de demanda ó rotación es, la clasificación XYZ. Los pasos según este método serían:

- Analizar el histórico de la demanda de los artículos y calcular la frecuencia de rotación o demanda para cada uno.

- Establecer los criterios de clasificación XYZ basados en la frecuencia de demanda. Por ejemplo, se puede utilizar una clasificación basada en la distribución porcentual de la frecuencia de demanda (por ejemplo, los artículos de alta demanda pueden ser clasificados como "X", los de demanda media como "Y", y los de baja demanda como "Z").
- Asignar la clasificación XYZ correspondiente a cada artículo según los criterios establecidos.

### 3.4 MÉTODO DE CLASIFICACIÓN DE INVENTARIOS BASADO EN ANÁLISIS DE CLÚSTERS

El método de clasificación de inventarios propuesto, se basa en el uso de técnicas de *clustering* de *machine learning*. Este método busca agrupar los componentes de ensambles finales con base a características similares, como la demanda, el valor, la criticidad, entre otros factores relevantes. El procedimiento propuesto para el método de reclasificación de inventarios es el siguiente:

**Definir las variables de interés.** Tras completar el análisis de la información e identificar las variables de interés basadas en datos históricos y/o relevantes para el analista, se procede a definir la tabla inicial para evaluar la relevancia de dichas variables. Algunos ejemplos de variables consideradas pueden ser la demanda, el precio, el costo, la estabilidad, la criticidad, entre otros.

**Obtener la importancia de las variables.** Esto se genera al aplicar el algoritmo de *random forest* a los datos: su finalidad es obtener el nivel de importancia de las variables y con esto, se seleccionan las variables de interés, según el valor sean los más altos.

Un Bosque Aleatorio utiliza árboles  $h_j(X, \Theta_j)$  como aprendices base. Para los datos de entrenamiento  $D = (x, y), \dots, (x, y)$ , donde  $x = (x_1, \dots, x_p)^T$  denota los  $p$

predictores e  $y_i$  denota la respuesta, y una realización particular  $\theta_j$  de  $\Theta_j$ , el árbol ajustado se denota como  $\hat{h}_j(x, \theta_j, D)$ . Así lo cita Cutler *et al.* (2012) y también mencionan el siguiente algoritmo del *random forest*.

---

**Algorithm 1** Bosques Aleatorios
 

---

**Require:** Datos de entrenamiento  $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$ , donde  $x_i = (x_{i,1}, \dots, x_{i,p})^T$

- 1: **for**  $j = 1$  hasta  $J$  **do**
  - 2:     Tomar una muestra bootstrap  $D_j$  de tamaño  $N$  de  $D$ .
  - 3:     Utilizar la muestra bootstrap  $D_j$  como datos de entrenamiento y ajustar un árbol utilizando particionamiento binario recursivo:
    - 4:         a. Comenzar con todas las observaciones en un solo nodo.
    - 5:         b. Repetir los siguientes pasos de forma recursiva para cada nodo no dividido hasta que se cumpla el criterio de parada:
      - 6:             i. Seleccionar  $m$  predictores al azar de los  $p$  predictores disponibles.
      - 7:             ii. Encontrar la mejor división binaria entre todas las divisiones posibles en los  $m$  predictores del paso (i).
      - 8:             iii. Dividir el nodo en dos nodos descendientes utilizando la división del paso (ii).
  - 9:     Para hacer una predicción en un nuevo punto  $x$ :
  - 10:     Para regresión:
 
$$11: \quad f(x) = \frac{1}{J} \sum_{j=1}^J \hat{h}_j(x)$$
  - 12:     Para clasificación:
 
$$13: \quad f(x) = \arg \max_y \sum_{j=1}^J I(\hat{h}_j(x) = y)$$
  - 14:     donde  $\hat{h}_j(x)$  es la predicción de la variable de respuesta en  $x$  utilizando el  $j$ -ésimo árbol.
- 

Para la obtención de la importancia de las variables usando el coeficiente de Gini, se puede especificar que, en cada nodo de rama del proceso de construcción del árbol, se seleccionan aleatoriamente varios subconjuntos de todas las características para obtener el mejor método de segmentación del subconjunto de características

mediante el cálculo de la pureza, (por ejemplo, ganancia de información, tasa de ganancia de información y coeficiente de Gini). El "bosque" se refleja en el crecimiento completo de cada árbol sin poda, y el número de árboles afecta al valor de decisión final. Así lo cita Tang *et al.* (2020) en su investigación.

Por ende, el proceso de construcción de un árbol de decisión mediante la partición de variables continuas, utilizando el coeficiente de Gini, va a dividir los nodos y construir el árbol de manera recursiva, de esta forma se explica a continuación:

- El cálculo del punto de partición  $S_i$  para las variables continuas  $a_i$  y  $a_i + 1$  es la siguiente:

$$S_i = \frac{a_i + a_{i+1}}{2}$$

Esta fórmula toma el valor de  $a_i$  y el valor adyacente  $a_i + 1$ , y calcula el punto medio entre ellos para obtener el punto de partición  $S_i$ . Este proceso se repite para todas las muestras y se obtiene un total de  $N - 1$  puntos de partición para discretizar las características continuas.

- La fórmula representa el cálculo del coeficiente de Gini para una partición  $P$ , donde  $K$  es el número de categorías y  $P_i$  es la probabilidad de ocurrencia de la categoría  $i$ .

$$Gini = 1 - \sum_{i=1}^K P(i)^2 \quad (3.1)$$

$$Gini(P) = 1 - \sum_{i=1}^K P(i)^2$$

**Para mejorar la coherencia y reducir la dispersión** en ciertas variables que podrían afectar el desempeño de los algoritmos de *clustering*, se aplicará una técnica de suavización de datos mediante el uso de la función logaritmo, cuando sea

necesario. Esta suavización ayudará a manejar la dispersión excesiva en las variables, promoviendo así un mejor rendimiento de los algoritmos de *clustering*.

La **función logarítmica** tiene la propiedad de comprimir los valores, lo que significa que los valores más grandes se reducen más que los valores más pequeños. Esto puede ser útil para reducir la escala de los datos y estabilizar la varianza cuando se encuentran datos con una dispersión excesiva.

La aplicación de la función logarítmica a los datos se define matemáticamente como:

$$y = \log_b(x)$$

Donde:

- $x$  es el valor original de los datos.
- $y$  es el valor transformado después de aplicar el logaritmo.
- $b$  es la base del logaritmo.

El logaritmo puede tener diferentes bases, como base 10 (logaritmo decimal) o base  $e$  (logaritmo natural). Si no se especifica la base, se asumirá que es el logaritmo natural (base  $e$ ).

Es importante tener en cuenta que la aplicación de la función logarítmica puede alterar la interpretación de los datos y es necesario considerar el contexto y los objetivos del análisis antes de aplicar esta transformación.

**Aplicación de algoritmos de *clustering***, se evalúan paralelamente el algoritmo de *clustering* k-medias denominado en inglés, como *k-means* y el *clustering* jerárquico, el propósito de estos es agrupar los componentes en clústers o grupos con base a sus características similares.

*K-Means* es un método de aprendizaje no supervisado simple que resuelve el problema de agrupamiento. El procedimiento busca clasificar un conjunto de datos

dado en un número predefinido de clústers ( $k$  clústers). La idea principal es encontrar  $k$  centroides, uno para cada clúster, de tal manera que los datos se agrupen de manera óptima. Para ello, los centroides se colocan inicialmente de forma estratégica en el espacio de datos para maximizar la distancia entre ellos.

Luego, cada punto de datos se asigna al centroide más cercano y se crea un clúster provisional. Después, se recalculan los nuevos centroides para cada clúster basados en la media de los puntos que lo componen. Este proceso de asignación y actualización de centroides se repite hasta que los centroides convergen y los clústers ya no cambian.

El objetivo del algoritmo es minimizar una función de error, en este caso, la suma de las distancias al cuadrado entre los puntos y los centroides de sus clústers respectivos.

En resumen, *k-means* busca agrupar los datos en  $k$  clústers, de manera que los puntos dentro de cada clúster sean similares entre sí y que los clústers estén bien separados unos de otros. El proceso iterativo de asignación y actualización de centroides busca lograr una partición óptima de los datos.

La fórmula del algoritmo *k-means* se representa de la siguiente manera:

$$\operatorname{argmin}_{c_i \in C} \sum_{i=1}^k \sum_{x \in c_i} \|x - \mu_i\|^2$$

La fórmula describe el objetivo del algoritmo *k-means*, que es encontrar una partición de los datos en  $k$  clústers que minimice la suma de las distancias al cuadrado entre cada punto de datos y el centro de su clúster. La notación *argmin* indica que estamos buscando los clústers ( $c_i$ ) que minimizan esta suma.

Cada variable tiene el siguiente significado:

- $c_i$ : Es un clúster (grupo de puntos) dentro del conjunto de todos los clústers  $C$ .

- $k$ : Es el número total de clústers en los que se dividirán los datos.
- $x$ : Es un punto de datos que se está agrupando.
- $\mu_i$  : Es el centroide o punto medio del clúster  $ci$ .
- $\|x - \mu_i\|^2$  : Es la distancia al cuadrado entre el punto  $x$  y el centroide  $\mu_i$  del clúster al que pertenece.

El algoritmo está compuesto por los siguientes pasos:

1. Coloca  $k$  puntos en el espacio representado por los objetos que están siendo agrupados. Estos puntos representan los centroides iniciales de los grupos.
2. Asigna cada objeto al grupo que tiene el centroide más cercano.
3. Cuando todos los objetos hayan sido asignados, recalcula las posiciones de los  $k$  centroides.
4. Repite el paso 2 y 3 hasta que los centroides ya no se muevan.

---

**Algorithm 2** Algoritmo de k-medias

---

- 1: **Input:**Conjunto de datos  $X$ , número de clusters  $k$
  - 2: **Output:**Conjunto de centroides  $C$
  - 3: **Inicialización:** Colocar  $k$  centroides iniciales aleatorios en el espacio representado por los objetos en  $X$
  - 4: **repeat** Los centroides ya no se muevan
  - 5:     **Asignación:** Asignar cada objeto en  $X$  al grupo con el centroide más cercano
  - 6:     **Recálculo:** Recalcular la posición de cada centroide como el centroide de los objetos asignados a ese grupo =0
- 

**Clustering jerárquico** construye un árbol (generalmente binario) sobre los datos. Las hojas son elementos de datos individuales, mientras que la raíz es un solo

clúster que contiene todos los datos. Entre la raíz y las hojas se encuentran clústers intermedios, que contienen subconjuntos de los datos. La idea principal del *clustering* jerárquico es formar -clústers de clústers- ascendiendo para construir un árbol. Los dos clústers que están más cercanos, se agrupan. Este proceso se repite hasta que todos los puntos pertenecen a un clúster construido jerárquicamente.

La estructura final del clúster jerárquico se llama **dendrograma**, la cuál es simplemente un árbol que muestra qué clústers se aglomeraron en cada paso. Así lo cita Adams (2009) en su investigación.

---

### Algorithm 3

---

Clustering Jerárquico Aglomerativo

- 1: **Input:** Vectores de datos  $\{x_n\}_{n=1}^N$ , distancia entre grupos  $D(G, G')$
  - 2:  $A \leftarrow \emptyset$  ▷ El conjunto activo comienza vacío.
  - 3: **for**  $n \leftarrow 1$  to  $N$  **do** ▷ Bucle sobre los datos.
  - 4:    $A \leftarrow A \cup \{x_n\}$  ▷ Agregar cada dato como su propio clúster.
  - 5:  $T \leftarrow A$  ▷ Almacenar el árbol como una secuencia de fusiones. En la práctica, se utilizan punteros.
  - 6: **while**  $|A| > 1$  **do** ▷ Bucle hasta que el conjunto activo solo tenga un elemento.
  - 7:    $G'_1, G'_2 \leftarrow \arg \min_{G_1, G_2 \in A; G_1 \neq G_2} D(G_1, G_2)$  ▷ Elegir el par en A con la mejor distancia.
  - 8:    $A \leftarrow (A \setminus \{G'_1\}) \setminus \{G'_2\}$  ▷ Eliminar cada uno del conjunto activo.
  - 9:    $A \leftarrow A \cup \{G'_1 \cup G'_2\}$  ▷ Agregar la unión al conjunto activo.
  - 10:    $T \leftarrow T \cup \{G'_1 \cup G'_2\}$  ▷ Agregar la unión al árbol.
  - 11: **Return:** Árbol  $T$ . =0
- 

**Métricas de evaluación de los clústers, coeficiente de *Silhouette*.** Uno de los indicadores de evaluación es el **coeficiente de *Silhouette***, también denominado de **Silueta (SC)**, que evalúa el clúster en función de la medida de la distancia promedio, entre un punto de datos y otros puntos de datos dentro del mismo clúster (cohesión) y la distancia promedio entre diferentes clústers (separación). La ventaja

del coeficiente de *Silhouette* es que depende únicamente de la partición del conjunto de datos y no del algoritmo de agrupamiento en sí. Su fórmula se representa de la siguiente manera:

$$s(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}}$$

$s(x)$  : El valor del coeficiente de *Silhouette* para un punto de datos  $x$ . Mide qué tan bien está agrupado el punto  $x$  dentro de su clúster comparado con otros clústers.

$a(x)$  : la distancia promedio entre el punto  $x$  y todos los otros puntos en el mismo clúster. Representa la cohesión del clúster. Es decir, que tan cerca están los puntos dentro del clúster de  $x$ .

$b(x)$  : La distancia promedio entre el punto  $x$  y todos los puntos en el clúster más cercano al que  $x$  no pertenece. Representa la separación entre clústers. Es decir, que tan lejos está  $x$  del clúster más cercano que no es su propio clúster.

$\max\{a(x), b(x)\}$ : el máximo valor entre la distancia promedio dentro del clúster  $a(x)$  y la distancia promedio al clúster más cercano  $b(x)$ . Se usa para normalizar el valor del coeficiente de *Silhouette*.

Donde el valor de  $s(x)$  puede variar entre -1 y 1. –

- -1 = mal agrupamiento
- 0 = indiferente
- 1 = bueno

El coeficiente de *Silhouette* para todo el agrupamiento es :

$$SC = \frac{1}{N} \sum_{i=1}^N s(x_i)$$

- $SC$  : El coeficiente de *Silhouette* para todo el agrupamiento, es una medida global que evalúa la calidad del agrupamiento en función del promedio del coeficiente de *Silhouette* de todos los puntos en el conjunto de datos.

- $N$  : el número total de puntos de datos en el conjunto de datos, es el tamaño del conjunto de datos.
- $\sum_{i=1}^N s(x_i)$  : La suma de los coeficientes de *Silhouette* individuales  $s(Xi)$  para cada punto de datos  $xi$  en el conjunto de datos. Cada  $s(Xi)$  mide la calidad del agrupamiento del punto de datos  $xi$ .
- $s(Xi)$  : El coeficiente de *Silhouette* para el punto de datos  $Xi$  mide qué tan bien está agrupado el punto  $Xi$  dentro de su clúster comparado con otros clústers, utilizando la fórmula mencionada anteriormente.

Estos dos conceptos: la cohesión intra-cluster y la separación inter-cluster son fundamentales en la evaluación de algoritmos de *clustering*, ya que se utilizan para medir la calidad de la agrupación resultante. Estas métricas proporcionan información sobre cómo se agrupan los datos dentro de los clústers y cómo se separan entre sí.

La cohesión intra-clúster se refiere a la similitud o proximidad de los puntos de datos dentro de un mismo clúster. Mide qué tan bien los elementos dentro de un clúster se agrupan entre sí y qué tan similares son. Mientras que por otro lado, la separación inter-cluster se refiere a la distancia o la diferencia entre los clústers. Mide qué tan bien los clústers están separados unos de otros y qué tan distintos son entre sí. Una alta separación inter-clúster implica que los clústers son diferentes y están bien definidos, lo que significa que los puntos de datos de un clúster están alejados de los puntos de datos de otros clústers.

Por lo tanto, para calcular el coeficiente de *Silhouette*, se compara la distancia promedio entre un punto y los demás puntos en su propio clúster (cohesión intra-cluster) con la distancia promedio entre el punto y los puntos en el clúster más cercano (separación inter-cluster). El resultado de su valor, cuando es cercano a 1 indica que, el punto está bien clasificado en su clúster y está alejado de los puntos en otros clústers. Pero, si por otro lado, el valor es cercano a -1 indica que el punto

puede haber sido asignado incorrectamente a su clúster y estaría más cerca de los puntos en otro clúster.

**Asignar la clasificación de inventarios** a cada grupo formado según su importancia, criticidad o cualquier criterio definido.

### 3.4.1 TÉCNICA DE COMPARACIÓN DE LOS MÉTODOS

La técnica de comparación de los métodos consiste, en evaluar y comparar el desempeño de los métodos ABC, XYZ y el método propuesto en términos de sus resultados y su capacidad para cumplir con los objetivos establecidos. Algunos pasos que se pueden incluir en esta técnica son los siguientes:

- Definir los criterios de evaluación, como la eficiencia en la clasificación, la precisión en la identificación de componentes críticos, la capacidad para minimizar los costos totales, entre otros aspectos relevantes.
- Aplicar los métodos ABC, XYZ y el método propuesto a un conjunto de datos de prueba.
- Comparar los resultados obtenidos por cada método en términos de los criterios definidos, identificando fortalezas y debilidades de cada enfoque.
- Realizar un análisis estadístico para determinar la importancia de las diferencias entre los métodos en cuanto a su desempeño.
- Presentar las conclusiones y recomendaciones basadas en la comparación de los métodos, destacando las ventajas y desventajas de cada uno y su idoneidad para cumplir con los objetivos establecidos en la investigación.

Y de esta manera se da por concluido este capítulo de la metodología. A continuación, sigue la parte de los resultados de la metodología, donde se describen que hallazgos obtuvimos por medio de nuestra metodología propuesta.

## CAPÍTULO 4

# APLICACIÓN Y RESULTADOS DE LA METODOLOGÍA

---

En este capítulo, se exponen los resultados obtenidos a partir de la metodología ya antes mencionada. Con este enfoque, se busca abordar los desafíos asociados con la gestión de inventarios y la optimización de la cadena de suministro en un entorno altamente dinámico y competitivo.

En el capítulo anterior, discutimos la importancia de una adecuada clasificación de inventarios en la industria automotriz y cómo es que puede influir en la eficiencia operativa, la reducción de costos y la mejora del nivel de servicio al cliente. También, se reconoce que las empresas enfrentan retos significativos debido a diferentes factores externos que no se pronostican y que no se piensan que van a ocurrir, lo que destaca aún más, la necesidad de una gestión efectiva en los inventarios.

Nuestra metodología se basa en el uso de técnicas avanzadas de aprendizaje automático y análisis de datos, usando técnicas ya conocidas como la clasificación ABC, la XYZ y la de multicriterio, en donde se emplea una herramienta no supervisada para identificar patrones y agrupar los componentes de manera automática, permitiendo una clasificación eficiente y sin sesgos predefinidos. Además, complementamos este enfoque con el uso de árboles de decisiones, con el que nos permitió

identificar las variables críticas que influyen en la clasificación de los componentes.

En este capítulo, se comparten los resultados de la aplicación de esta metodología. Se presentan las categorías identificadas por la herramienta no supervisada y la relevancia de cada grupo en el contexto de la cadena de suministro. Así mismo, se explicarán los hallazgos derivados del análisis de variables críticas mediante árboles de decisiones, lo que nos brindó una comprensión más profunda de los factores que afectan la clasificación de inventarios.

Los resultados obtenidos en esta sección estarán divididos en tres puntos diferentes, los cuáles son:

1. Prueba de la metodología contra una base de datos con clasificación.
2. Aplicación de la metodología al caso práctico.
3. Comparativa de resultados de la metodología contra ABC y XYZ.

Los resultados obtenidos tienen como objetivo proporcionar una base sólida para la toma de decisiones en la gestión de inventarios y la planificación de la cadena de suministro. Los datos y conclusiones presentados en este capítulo serán fundamentales para el desarrollo de estrategias eficaces que permitan a la empresa optimizar sus recursos, reducir costos y garantizar una respuesta ágil y efectiva ante las demandas del mercado.

En resumen, este capítulo de resultados representa un paso significativo hacia una gestión más inteligente y eficiente de los inventarios en la industria automotriz. Los avances tecnológicos y las técnicas de análisis de datos aplicados, han brindado nuevas perspectivas y soluciones para enfrentar los desafíos que conlleva la clasificación de componentes, en un mercado altamente cambiante y competitivo. A través de estos resultados, esperamos contribuir al desarrollo sostenible y exitoso de la empresa, asegurando su posición en el sector y la satisfacción de las necesidades de sus clientes.

## 4.1 COMPROBACIÓN DE EFICIENCIA DEL MODELO CON BASES DE DATOS EXISTENTES

Para la ejecución de nuestro modelo, era necesario usar una base de datos cuya información fuera de uso público y no tan extenso, con el fin de comprobar que nuestro modelo funcionara y se obtuvieran los resultados esperados. Para eso se usó una base de datos que estaba incluida dentro del programa, con el que se trabajó para la ejecución de nuestra base de datos. El programa usado fue el -R project- el cuál es un software con un entorno y lenguaje de programación y el cuál tiene un enfoque al análisis estadístico.

La base de datos que este software incluye dentro de sus paquetes lleva el nombre de -iris-, este archivo muestra información de las especies que esta flor tiene y así como esta, existen otras bases de datos que son fáciles de usar para ejemplificar ciertos modelos y verificar con ellos que los resultados sean exitosos.

A continuación, se describe un poco más el proceso que se llevó a cabo para la realización de este modelo con la base de datos iris.

### 4.1.0.1 COMPARACIÓN CON BASE DE DATOS IRIS

Se utilizó una base de datos donde se observa información de las diferentes especies de iris que existen y en él se tenía las dimensiones de tamaño (largo y ancho) por sus sépalos y pétalos. Considerando que esta base de datos mostraba muy poca información a comparación de la base de datos a experimentar, se pudo observar que la metodología utilizada, brinda el mismo resultado de clasificar de mejor forma las especies de iris considerando diferentes variables al mismo tiempo.

Por lo tanto, aquí se presentan los resultados del experimento aplicado a la base de datos "iris". Este famoso conjunto de datos de iris (también conocido como

conjunto de datos de *Fisher o Anderson*) proporciona las medidas en centímetros de las variables longitud y ancho del sépalo, y longitud y ancho del pétalo, respectivamente, para 50 flores de cada una de las 3 especies de iris. Las especies son iris setosa, versicolor y virginica. Esta base de datos está incluida en el paquete base de R. La base de datos contiene cuatro variables independientes, y una variable dependiente que corresponde a la clasificación de las especies.

A continuación, se muestra en la siguiente tabla 4.1 el encabezado de la base de datos “iris” para visualizar sus primeras filas:

TABLA 4.1: Base de datos iris

Sepal.Largo	Sepal.Ancho	Petal.Largo	Petal.Ancho	Especie
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

#### 4.1.0.2 OBTENCIÓN DE LA IMPORTANCIA DE LAS VARIABLES CON EL COEFICIENTE DE GINI

Para tener un contexto de que es el coeficiente de Gini, según a como lo describe Daniya *et al.* (2020), son medidas que se pueden utilizar para medir condiciones de desigualdad, distribución uniforme, distribución de probabilidad discreta y se pueden aplicar a los árboles de clasificación y regresión para, clasificar los puntos de decisión para el conjunto de datos dado y regla de división aplicada que mejora el rendimiento

de los árboles; del cuál da un resultado de un árbol óptimo.

En otras palabras mientras tenga un valor más alto el coeficiente de Gini, indica una mayor desigualdad, mientras que un valor más bajo indica una distribución más equitativa.

Una vez que tenemos nuestra base de datos de iris, usaremos el coeficiente de Gini, el cuál se aplicará una vez que ya hayamos considerado todas las variables que creemos que son relevantes para nosotros. Por eso mismo, al momento de ser aplicado a las variables que nosotros consideramos de gran importancia, se hace referencia a las que, cuyo valor son mayor. Tomando en cuenta estas variables con mayor valor se ejecuta el modelo para, que brinde un resultado preciso con base en lo que se busca obtener.

Para determinar el orden jerárquico de las variables en la generación de los clústers, se aplicó el algoritmo de *random forest* y se calculó el coeficiente de Gini para obtener la importancia de cada variable como se muestra en la tabla 4.3.

TABLA 4.3: Niveles de importancia de las variables en términos de calidad de clúster

según el coeficiente de Gini

	Importancia
Petal.Largo	45.46807
Petal.Ancho	40.32552
Sepal.Largo	38.33385
Sepal.Ancho	24.32587

Para mostrar el efecto de la importancia de las variables y seleccionar las mejores para el clúster, se comparan gráficamente las dos variables con mayor importancia (figura 4.1) contra las dos variables con menor importancia (figura 4.2) según la tabla anterior (4.3). Esto permite observar la separación adecuada de las especies para

generar un mejor clúster.

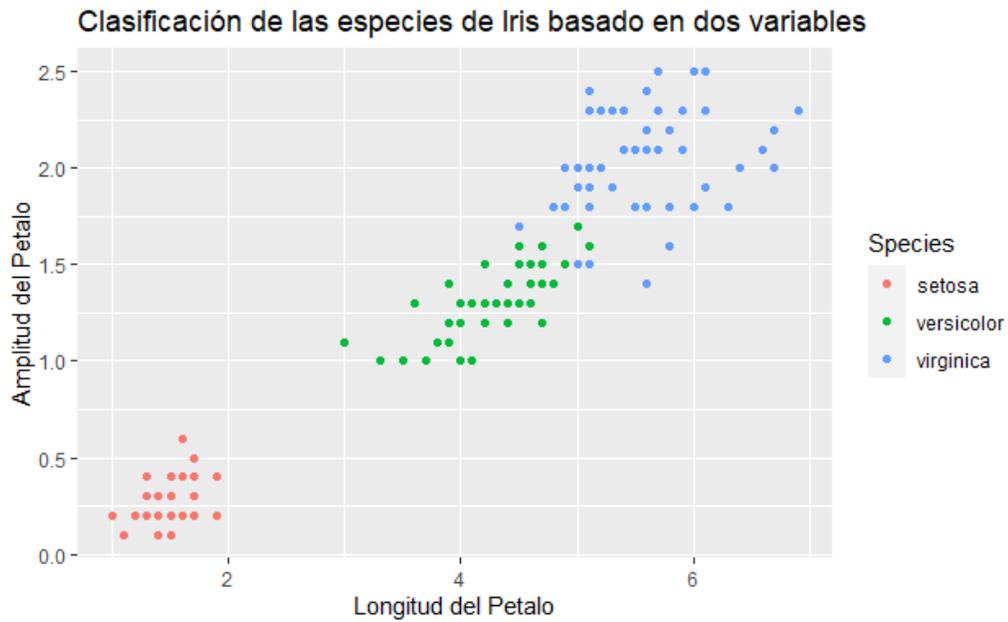


FIGURA 4.1: Clasificación de las especies de iris basado en las dos variables con mayor importancia

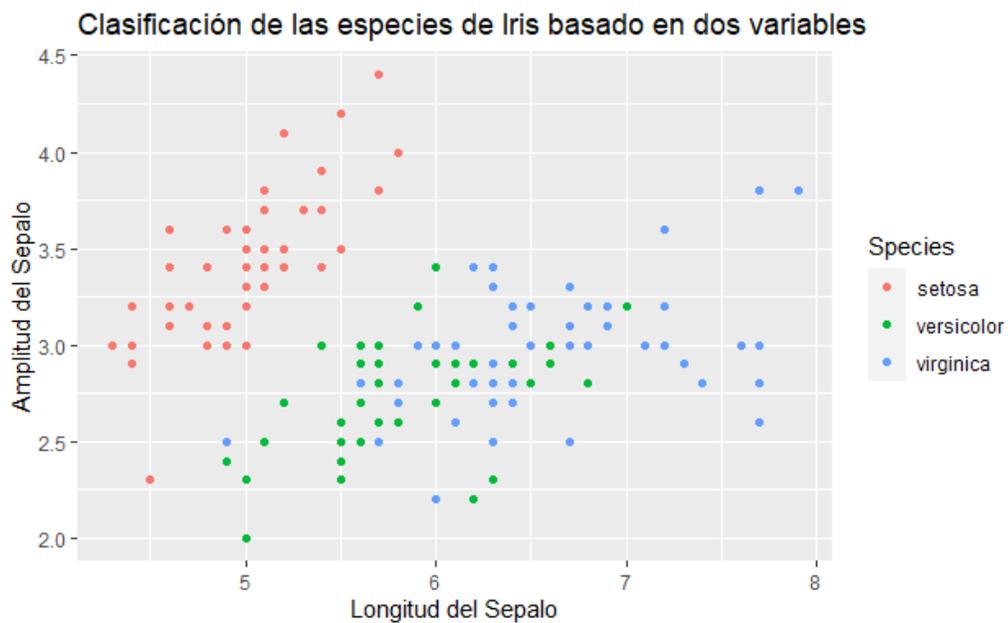


FIGURA 4.2: Clasificación de las especies de iris basado en las dos variables con menor importancia

Con esta visualización, se puede apreciar claramente cómo las dos variables más importantes muestran una mejor separación de las especies, lo que sugiere, que son más relevantes para la creación de clústers precisos en el conjunto de datos “iris” .

Al ser datos con escalas muy parecidas determinamos con la exploración, que no es necesario aplicar la suavización con la función logarítmica y realizar los clústers de manera directa.

Ahora aplicamos el algoritmo de *clustering* jerárquico usando las dos variables mas importantes que nos arrojó el coeficiente de Gini y obtenemos el siguiente valor de Gini.

```
## [1] 0.6573949
```

En la siguiente figura 4.3, se puede observar la calidad de la clasificación del clúster obtenido, tomando en cuenta las variables con mayor importancia basado en su largo y ancho del pétalo.

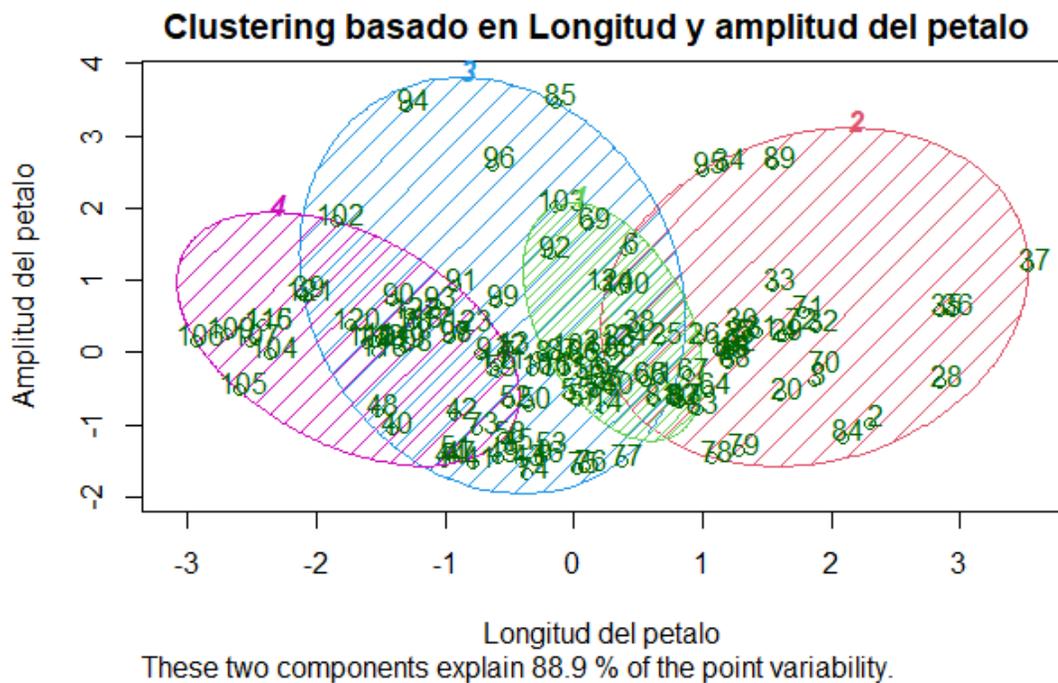


FIGURA 4.3: Clúster obtenido basado en longitud y amplitud del pétalo.

Finalmente, comparamos la clasificación real contra la clasificación del método propuesto, mediante la matriz de confusión, así como se muestra en la siguiente tabla 4.4, que describe las especies de iris.

TABLA 4.4: Resultado de la matriz de confusión con especies de iris.

setosa	versicolor	virginica
50	0	0
0	45	1
0	5	49

Como podemos observar en la matriz de confusión, el método para esta base de datos fue tan efectivo que se acercó al 94% de precisión de clasificación con este método explorado, detectando las variables importantes y clasificando a ciegas con la base de datos.

En la especie setosa no hubo errores al clasificar esta especie. Todas las muestras de setosa fueron correctamente identificadas (50/50). Para la especie versicolor 1 muestra fue incorrectamente clasificada como virginica (1 error) y para la especie virginica 5 muestras fueron incorrectamente clasificadas como versicolor (5 errores). Estos errores pueden deberse a la similitud entre las características de estas especies, lo que sugiere que se podría explorar un ajuste fino del modelo o el uso de métodos adicionales para mejorar la discriminación entre estas clases.”

Este tipo de descripción proporciona un análisis equilibrado que destaca tanto los logros del modelo como las áreas donde podría mejorar.

## 4.2 IMPLEMENTACIÓN CON EL MÉTODO DE CLASIFICACIÓN DE INVENTARIOS CON CLUSTERING.

En este apartado, se detalla la implementación práctica del método de clasificación de inventarios mediante técnicas de *clustering*. Se describe el proceso paso a paso, desde la recopilación de datos y la preparación de la información, hasta la aplicación de algoritmos de *clustering* para la clasificación de inventarios.

El enfoque se apoya en la identificación de patrones, relaciones y similitudes entre los componentes del inventario, permitiendo una agrupación eficaz y sin sesgos predefinidos. Este método no solo propone una clasificación robusta, sino que también destaca la relevancia de variables clave en el proceso de toma de decisiones para la optimización de la cadena de suministro.

El presente apartado proporcionará una visión detallada de la implementación del método, abordando las etapas críticas del proceso y destacando la importancia de esta metodología, en la gestión estratégica de inventarios. Los resultados obtenidos muestran el potencial de esta técnica para transformar la gestión de inventarios, permitiendo una mejor asignación de recursos y una planificación más eficiente para satisfacer las demandas del mercado en constante evolución.

### 4.2.1 LIMPIEZA Y SELECCIÓN INICIAL DE LOS DATOS

En esta sección del documento se presentan los resultados del experimento aplicado a la base de datos “*clustering*” ver tabla 4.5. La cuál corresponde a los datos obtenidos del caso práctico de la empresa y proporciona información como el número de parte, su descripción, cantidad de inventario actual, precio, el tiempo de tránsito, el tamaño de lote, cantidad de piezas por uso de ensamble, unidad de medida, costo, la media, su coeficiente de variación, la estabilidad y la rotación, para

la cantidad de 156 números de componentes usados en varios ensambles finales de una empresa.

TABLA 4.5: Base de datos *clustering*

PARTNO	DSC	INVACT
PECIO	QTYASSE	MEDIA
LT	LOTSIZE	QTY BOX
UNIDAD MEDIDA	COSTO	ESTAB
COEFVAR	ROTACION	

Existe gran variabilidad de componentes que son utilizados y algunos de estos rondan en materiales de aluminio, plásticos, anillos, motores, rotores, terminales, etc.

Brevemente aquí definimos cada una de las variables utilizadas en esta base de datos:

- PRTNO: es la identificación del componente con un número de parte.
- DSC: es la descripción o nombre del componente.
- “INV ACT”: es la cantidad actual con la que se cuenta con inventario.
- PRECIO: es el valor por pieza de dicho componente.
- LT: es el “lead time” o también conocido como el tiempo en tránsito que tarda cada componente en recibirse en el lugar destino.
- LOTSIZE: es el tamaño del lote descrito por cada número de parte.
- QTYBOX: es la cantidad que tiene cada caja por componente.

- QTYASSY: es la cantidad que se ensambla en el producto terminado.
- UNIDAD DE MEDIDA: es la unidad de medida. si son pieza, kilogramos, gramos, etc.
- COSTO: es el total de precio por el inventario.
- MEDIA: es la suma de un conjunto de números divididos por la cantidad de números que forman el conjunto.
- COEFVAR: es la variación ambicionada de un conjunto de datos respecto de su media.
- ESTAB: la estabilidad respecto a un periodo de tiempo.
- ROTACION: es la cantidad de veces que este va a cambiar.

Cabe recalcar que para la limpieza de datos, de la información previamente descrita, no se utilizaron en su totalidad los componentes o bien números de parte, debido a que no todos proporcionaban información de uso, o su variable de unidad de medida no era en piezas. Por lo tanto, se optó por usar solo los números de parte que si mantenían información respecto a un periodo específico y cuya unidad de medida fueran todas piezas y no kilogramos o gramos.

A continuación se muestra un ejemplo de la base de datos “clustering” en la tabla 4.6, para observar el encabezado con su respectiva descripción. También se puede ver que en las filas, el resultado de ciertas variables es similar, esto se debe a que estos componentes se ensamblan en un mismo producto terminado y pertenecen a un mismo clúster.

TABLA 4.6: Base de datos "CLUSTERING"

DSC	Costo	Media	Lot.Size	CoefVar	LT.Coeff
RETAINER	108336.746	246368	4500	0.14841	0.151692
ORING	18142.02156	246368	18000	0.14841	0.151692
CAP	24504.3185	246368	45000	0.14841	0.151692
SPRING	28681.52625	246368	12500	0.14841	0.151692
PLATE	64952.622	246368	15000	0.14841	0.151692

#### 4.2.2 DETERMINACIÓN DE LAS VARIABLES CON MEJOR DESEMPEÑO PARA CLUSTERING.

Ya que tenemos la base de datos con todas las variables, se toma la decisión de usar las que a criterio propio sean las más relevantes. Sin embargo, para identificar que variables van a funcionar mejor con nuestra metodología, se opta por aplicar el algoritmo de *random forest* y de esta manera también se calcula el coeficiente de Gini para obtener la importancia de cada variable.

De esta manera entre mayor sea el valor de las variables seleccionadas, se considerarán más influyentes en el modelo de *random forest*, para hacer predicciones. Estas variables contribuyen más a la reducción del error o la impureza en las predicciones del modelo; para determinar el orden jerárquico de las variables en la generación de los clústers.

En la tabla 4.7 se puede observar la calidad de las variables más importantes según el coeficiente de GINI del *random forest*. Con base a la metodología propuesta

y nuestra base de datos llamada *clustering*, se usaron las tres variables de rotación, crítico y costoinv. Esto se debe a que estas tres variables son las que mayor valor tienen y son las adecuadas para usar a nuestra metodología, esto fue posible obtenerlo ya que se aplicó el coeficiente de Gini.

De esta manera con las tres variables de rotación, crítico y costoinv esperamos que los resultados que arroje nuestra base de datos, sean precisos y certeros, a que si otras variables hubieran sido utilizadas.

TABLA 4.7: Variables importantes aplicando el coeficiente de GINI de *random forest*.

Variable	Coeficiente Gini
rotación	23.83125
crítico	21.76606
CostoInv	21.59665
Media	18.75183
LotSize	18.73616
coef	18.4732

### 4.2.3 OBTENCIÓN DE CATEGORÍAS MEDIANTE CLUSTERING

La siguiente sección pretende abordar la importancia de los agrupamientos de la base de datos de donde se realizó este caso práctico. Al igual que esta empresa y en términos generales, para cualquier empresa del rubro que sea, es sencillo la toma de decisiones y el análisis de ciertos datos, cuando están agrupados en diferentes categorías.

En este apartado se podrá observar los resultados obtenidos de las categorías mediante los agrupamientos de los componentes de la base de datos, tomando en cuenta diferentes variables.

Tomemos en cuenta que nuestra base de datos tenía aproximadamente trece variables que eran de nuestro interés, así como previamente se mencionaron y de las cuáles solo se usarán 3 con base a su mayor valor de acuerdo al coeficiente de Gini, a comparación del total de las 6 variables cuyo interés predominaba. También haremos un pequeño análisis de las diferencias que se obtuvieron y la manera en como esto pudiera impactar a la empresa.

Y, por último, hablaremos de la función logarítmica, su finalidad, su función y el beneficio que esta atrae al usarla en esta experimentación.

#### 4.2.3.1 ESTABILIZACIÓN DE LA VARIACIÓN DE LAS VARIABLES

Recapitulando el trabajo que hasta el momento se ha realizado, recordemos que inicialmente se comenzó por definir las variables de interés de la base de datos; una vez ya definidas, se usó la técnica de *random forest* para ver la calidad de las variables, con la cuál se le aplicó el coeficiente de Gini para identificar cuáles eran las más importantes.

Después se continuó con correr nuestro programa y se crea visualmente la relación de la lista de nuestros componentes en un histograma, donde podemos observar el total de los números de parte haciendo una comparación del acumulado y su costo de inventario, con la cuál podemos observar que existe una variabilidad muy grande y podemos ver que esta figura 4.4 está asimétrica.

En la siguiente figura 4.4 podemos observar nuestro histograma sin función logaritmo, y podemos ver que muchos componentes que se define como acumulado cuentan con un costo de inventario bajo y que muy pocos cuentan con un costo alto.

Esto quiere decir, que de acuerdo a las variables importantes y a nuestra programación, existen más de 75 números de parte cuyo costo de inventario es muy bajo y muy pocos con un costo muy alto. Ya, si esta información la enfocamos más en

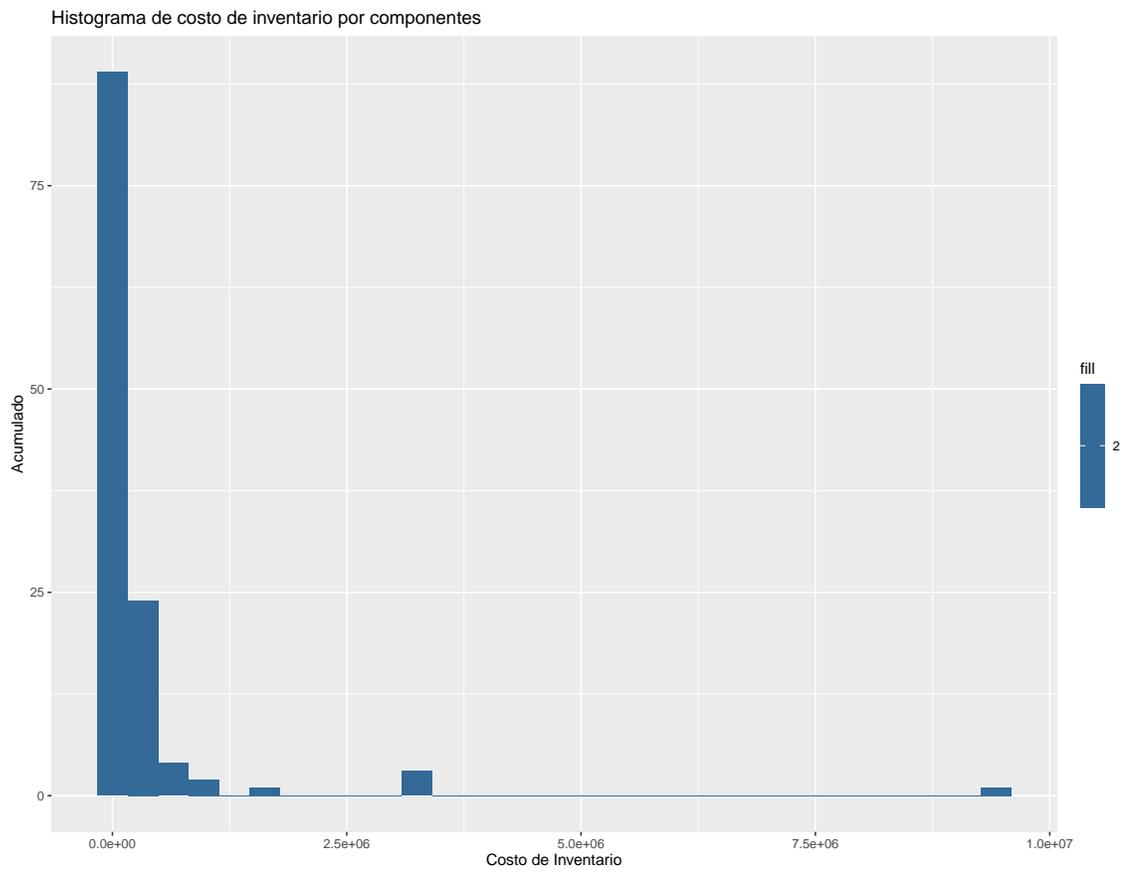


FIGURA 4.4: Histograma de costo de inventario por componentes sin función logarítmica.

términos para la empresa. Si la empresa desea identificar o agrupar sus inventarios con base a los costos de cada componente, tendrá más de 75 números de parte que deberá considerar como un solo grupo, ya que, su costo es muy bajo y si agrupa en otra clasificación los componentes más caros, solo 3 de estos, serán considerados críticos debido a su alto valor en costo.

Para esto, la empresa estará enfrentando problemas con respecto a su forma de planear los componentes, si considera estas agrupaciones. Ya que, si lo vemos en relación a costo de inventario y definen una política alta para los componentes con bajo costo, a la larga podrán tener problemas de sobre inventario si llegasen a tener fluctuaciones de clientes.

Si esto sucede y la empresa recibe reducciones, todos estos componentes agrupados generarán la acumulación de componentes de bajo costo y por consecuencia, podrán llegar a tener pérdidas monetarias debido a la falta de movimiento del material y sin ventas a la empresa, le generará mermas, sobre inventario y falta de espacio en el almacén, que puedan terminar en materiales dañados.

Por lo tanto, para evitar este tipo de problemas es que se recomienda aplicar a nuestra técnica, la función logarítmica, cuya finalidad es mejorar la coherencia y reducir la dispersión de nuestra evaluación, así como se puede observar en la figura 4.5.

De esta manera es que pasamos a la parte de estabilizar la variabilidad de las variables. La función logaritmo como ya lo expliqué, tiene el propósito de normalizar los datos, reducir el sesgo de las variabilidades, transformar los resultados a una forma más simétrica e identifica los patrones de una forma más clara. A continuación te describo lo que significa cada uno de estas finalidades.

Para la parte de normalizar los datos significa que, cuando existen diferentes escalas o distribuciones sesgadas, esta función ayuda a garantizar que todas las variables, contribuyan de manera más equitativa al proceso del *clustering*.

También la reducción del sesgo en valores grandes, quiere decir que cuando obtenemos o presenciamos valores extremadamente grandes que puedan dominar en el análisis del *clustering*, el uso de logaritmos puede reducir su impacto. En otros términos, es que estos valores grandes los escala hacia abajo, de forma que queden más simétricos.

Para la parte de la simetría, cuando algoritmos de *clustering* asumen una distribución normal o simétrica en los datos, podemos concluir que el resultado o funcionamiento será óptimo.

Y por último, la identificación de patrones, esto quiere decir que en algunos casos al tomar logaritmos de ciertas variables, puede revelar relaciones más simples o

lineales entre las variables y esto facilita la identificación de patrones o agrupamientos más claros mediante técnicas de *clustering*.

Por lo tanto para evitar este tipo de problemas es que se recomienda aplicar a nuestra técnica, la función logarítmica; esta ayudará a mejorar la coherencia y reducir la dispersión de nuestra evaluación, así como se puede observar en la figura 4.5

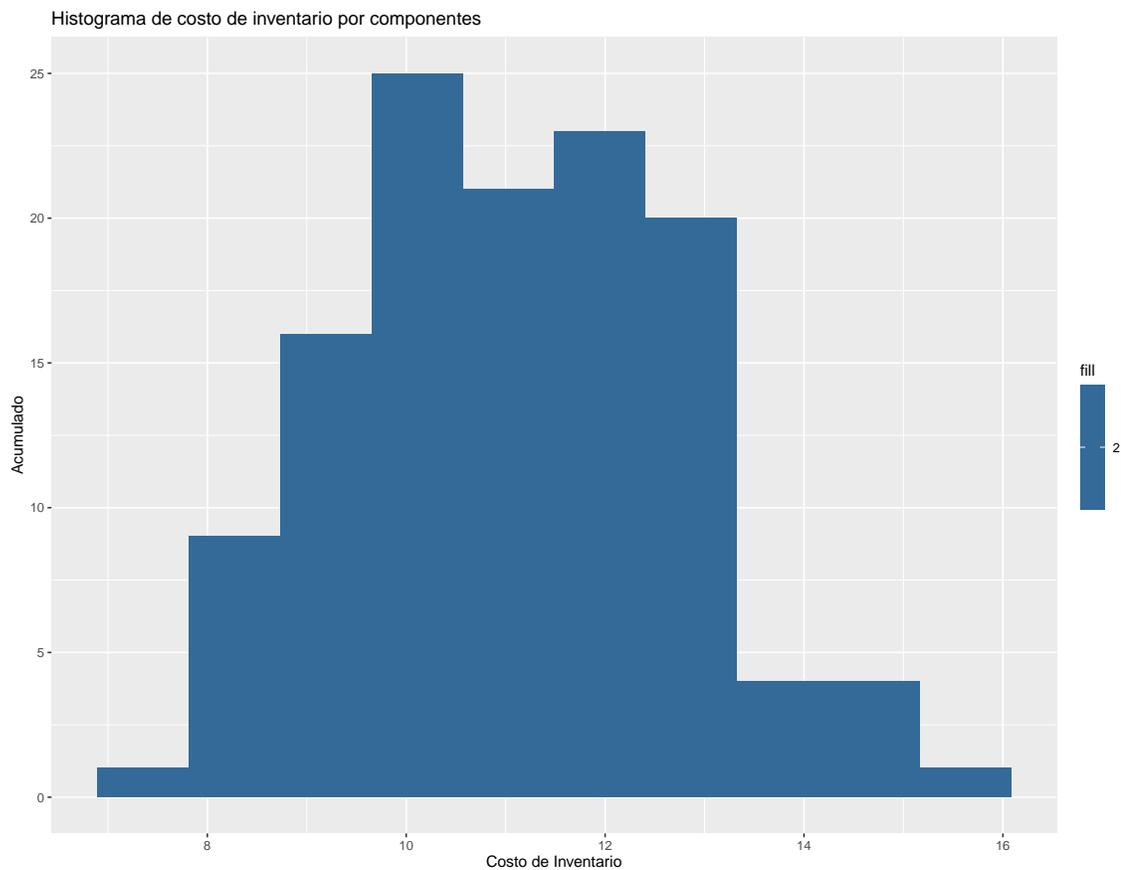


FIGURA 4.5: Histograma de costo de inventario por componentes con función logarítmica

En esta figura, se puede observar como al aplicar la función logaritmo a nuestros componentes, el costo de inventario se puede observar que es más variado, ya que existe cierta simetría en los acumulados de los costos por componentes y de esta forma, ya redujo el sesgo de los valores y se normalizan un poco más, ya que es más equitativo el proceso del *clustering*.

Al momento de observar una información más simétrica y la empresa toma en cuenta estos valores, existe mayor probabilidad que la toma de decisiones cambie, a diferencia de como se explicó al principio, cuando no se había aplicado la función logaritmo.

Se reconoce que con la función logaritmo los costos de inventarios de todos los componentes se estabilizan más y es más adecuado la agrupación de los componentes dependiendo su costo. Lo cuál si la empresa usa esta nueva información y considera definir una política de inventario de sus componentes con base a su costo, debe tomar en cuenta esta información, para que pueda agrupar correctamente los componentes.

Si existiesen fluctuaciones y agrupan políticas de inventario con base a los grupos de componentes como lo vimos anteriormente, será mucho más equitativo realizar los ajustes de órdenes de compra hacia los proveedores, esto con la finalidad de reducir la política de inventario de la empresa (siendo este el caso cuando existen reducciones del cliente).

Ya que, como se verá un incremento en costo, las barras incrementarán, pero si reducimos/eliminamos las órdenes de compra de nuestro proveedor, poco a poco el inventario también irá reduciendo, debido a que se consumirá lo que ya se tiene en almacén físico. De esta manera, podemos controlar la cantidad de inventario, la cantidad de dinero y las fluctuaciones que podamos tener por parte de nuestros clientes.

#### 4.2.3.2 APLICACIÓN DEL MÉTODO DE CLUSTERING JERÁRQUICO

Para comenzar con este apartado, damos una pequeña introducción de lo que es el *clustering* jerárquico. La cuál es una técnica de agrupamiento que se utiliza en análisis de datos y el aprendizaje automático.

Este método de agrupamientos tiene como función, agrupar elementos similares

en clústers o grupos con base en la proximidad o similitud entre ellos. Al identificar patrones y estructuras inherentes en los datos, se facilita la comprensión de la información que contienen los datos y adicional a esto, nos sirve para reducir la complejidad de gran cantidad de información de datos, al agrupar elementos similares y lo representan de una manera visual.

Esta herramienta visual que representa el método jerárquico se le denomina dendrograma y muestra la estructura jerárquica de agrupamientos, donde los elementos individuales se encuentran en la parte inferior y los clústers más grandes y similares se van formando a medida que ascendemos en el gráfico. Las ramas del dendrograma representan los enlaces entre los elementos o clústers, y la altura de estas ramas indica la distancia o similitud en la cual se unen.

Aquí en la siguiente figura (4.6), podemos observar el dendrograma obtenido con respecto a nuestro método aplicado con la base de datos en experimentación y podemos observar que la cantidad de agrupamientos que genera ronda entre los 3 grupos de componentes cuyas similitudes existen entre ellos.

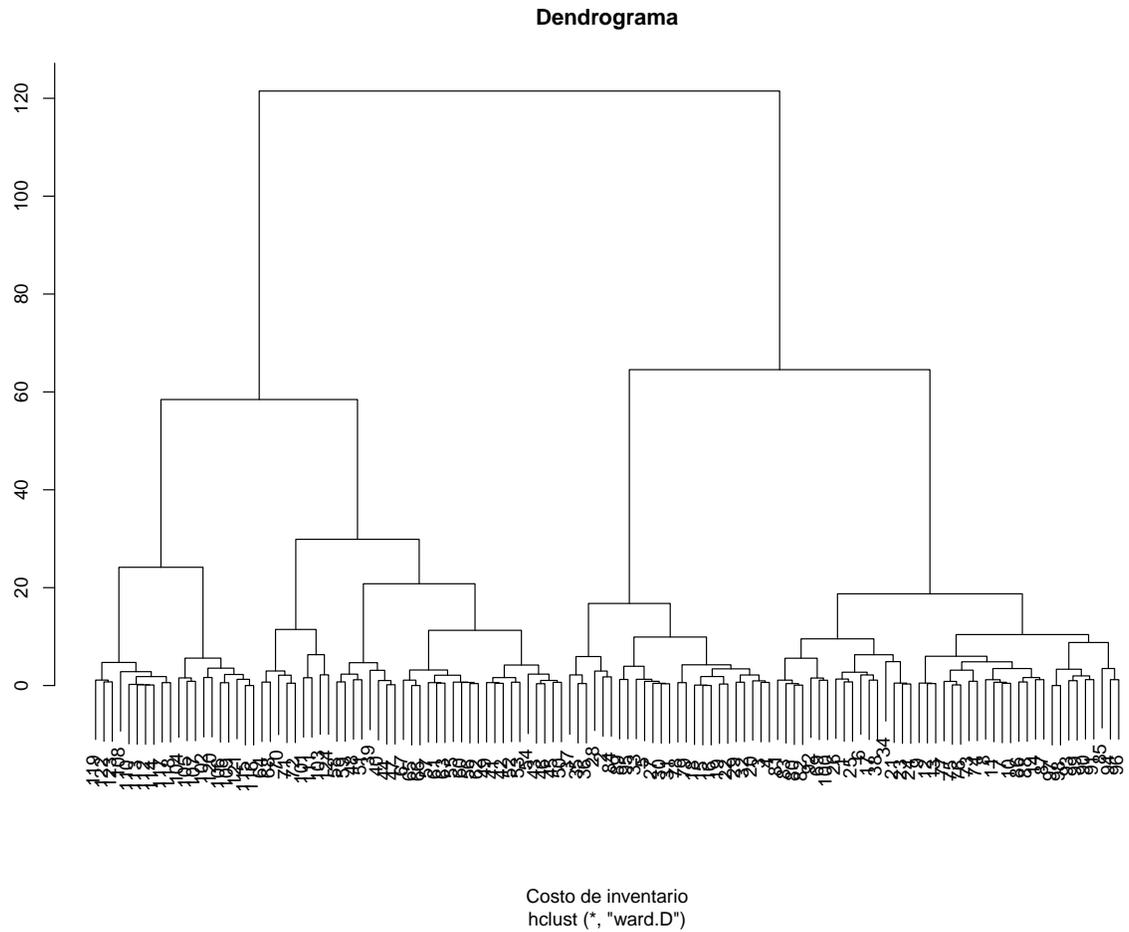


FIGURA 4.6: Dendrograma final después del tratamiento de los datos

Ya que pudimos observar el árbol del dendrograma, en la siguiente figura (4.7) , podemos observar los 4 clústers visualmente detallados y los que nos sugiere nuestra base de datos con respecto al método aplicado. Si la empresa toma en cuenta lo que nuestra técnica arroja, es recomendable que de todos los componentes que maneja, los divida en 4 grupos para realizar su correcta planificación y administración con base en el costo del inventario.

La división de los 4 grupos, los divide de acuerdo a similitudes que los componentes comparten y el cuál nuestro método considera los más importantes, para que se tomen decisiones correctas con base a lo que la empresa buscaría cuidar con el costo, tamaño de lote y administración en su planificación.

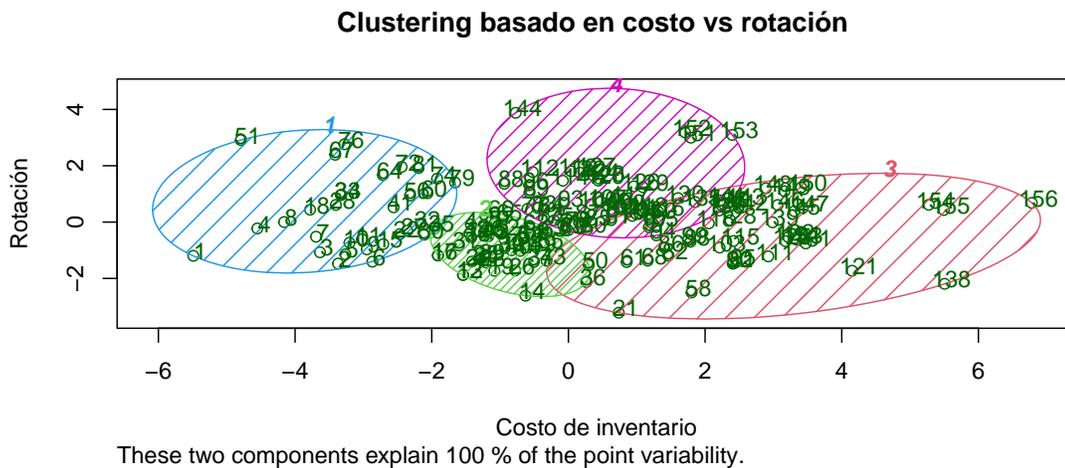


FIGURA 4.7: Clústers adquiridos considerando costo vs rotación

#### 4.2.3.3 APLICACIÓN DEL MÉTODO DE CLUSTERING JERÁRQUICO

Por otro lado se encuentra el método de las  $k$ -medias o también llamado en inglés *k-means* el cuál es un método de aprendizaje no supervisado simple que resuelve el problema de agrupamiento. El procedimiento busca clasificar un conjunto de datos dado en un número predefinido de clústers ( $k$  clústers). La idea principal es encontrar  $k$  centroides, uno para cada clúster, de tal manera que los datos se agrupen de manera óptima. Para ello, los centroides se colocan inicialmente de forma estratégica en el espacio de datos para maximizar la distancia entre ellos.

En otras palabras, el algoritmo de las  $k$  medias intenta minimizar la varianza dentro de cada clúster y maximizar la distancia entre clústers.

Para determinar el tamaño óptimo del clúster, la figura (4.8), muestra la relación entre el número de clústers y el indicador de la calidad de los clúster WCSS. Con base a la referencia de la figura, podemos observar en donde predomina más el codo que es de 2 clústers donde obtendríamos mayor calidad en la separación, sin embargo igualaremos a 4 clústers así como se mostraba con el clustering jerárquico con 4 clústers.

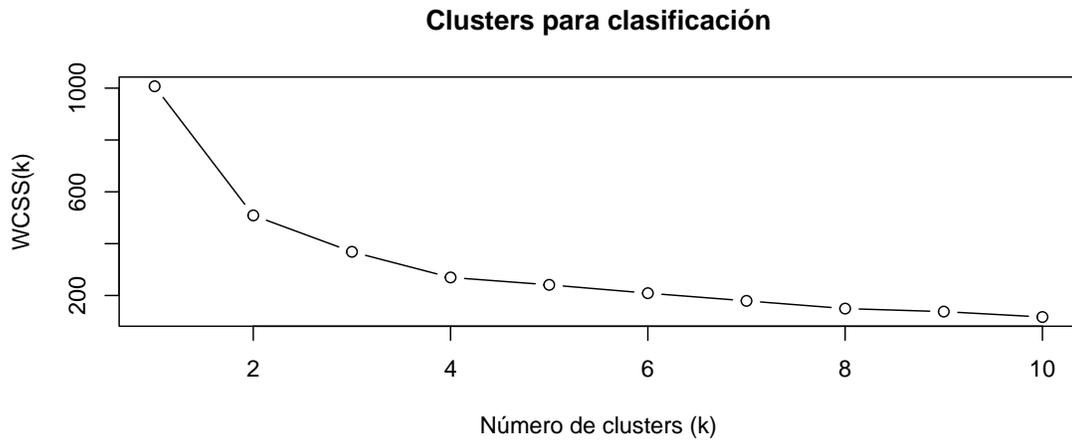


FIGURA 4.8: Clúster de k-medias

En la figura 4.9 se puede observar los clústers, así como anteriormente lo veíamos con el dendrograma, pero ahora es con la información del método K-medias. Al igual que el dendrograma, el método de las k-medias es posible apreciar 4 grupos con similitudes entre sí y cuya finalidad es obtener la cantidad apropiada para dividir y separar los componentes, para generar políticas de inventario apropiadas y una mejor planificación; para cumplir con ciertos niveles y KPIs de la empresa.

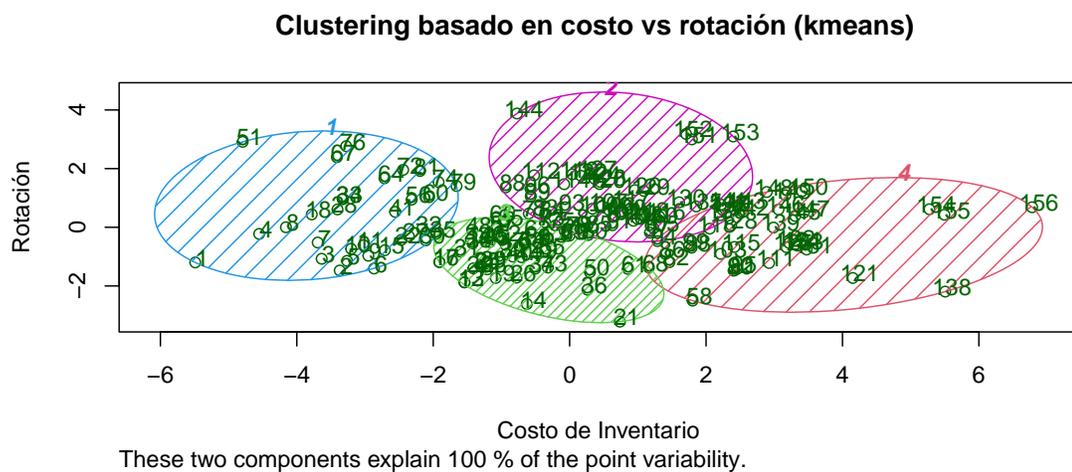


FIGURA 4.9: Clúster de k-medias basado en costo vs rotación

#### 4.2.4 COMPARATIVA ENTRE MÉTODOS

Para finalizar esta sección se comparan los métodos tradicionales de clasificación de inventarios, el método ABC y el XYZ, contra los grupos generados con la metodología propuesta.

Para mostrar la diferencia entre la generación de grupos del método propuesto, demostramos una clasificación donde se seleccionan 3 variables de interés diferentes para el gerente de almacén las cuales son: precio, rotación y tamaño de lote mínimo por pedido.

La tabla 4.8 muestra la clasificación con los distintos métodos usados, puede observarse mucha precisión del método propuesto con la clasificación de los métodos ABC y XYZ, sin embargo, la diferencia radica en que los métodos de *clustering* jerárquico y k-medias consideran variables importantes adicionales a diferencia de los métodos ABC y XYZ, los cuales son de mucho valor reclasificar para el gerente del almacén que no es posible considerar con ABC o XYZ. Entonces no se pierde calidad, si no que se gana precisión en la consideración de las variables de interés.

TABLA 4.8: Tabla comparativa de métodos clustering

Consecutivo	Numero de parte	ABC	XYZ	K-medias	Jerárquico
138	MX084663-0120	1	1	1	1
156	MX082159-0030	1	1	1	1
155	MX082158-0060	1	1	1	1
154	MX082152-0070	1	1	1	1
131	MX082130-0020	1	1	1	1
133	MX084663-0130	1	1	1	1
134	MX084632-5080	1	1	1	1
132	6655071-S	1	1	1	1

TABLA 4.8: (Continúa)

Consecutivo	Numero de parte	ABC	XYZ	K-medias	Jerárquico
147	MX082154-0150	1	1	1	1
139	MX082154-0130	1	1	1	1
145	MX084632-5100	1	1	1	1
146	MX082151-0110	1	1	1	1
150	MX192365-6060	1	2	1	1
128	MX082115-0030	1	2	1	1
149	8891002	1	2	1	1
142	MX082158-0080	1	2	1	1
148	MX084641-5090	1	2	1	1
143	MX084638-5110	1	2	1	1
136	MX082174-0020	1	2	1	1
137	MX192365-1700	1	2	1	1
140	MX082152-0100	1	2	1	1
141	MX082143-0050	1	2	1	1
135	TN136250-1344	1	2	3	1
130	MX084669-0410	1	2	3	1
129	8890976	1	3	3	3
153	MX082153-0060	1	3	3	3
151	3X060139	1	3	3	3
152	MX082159-0040	1	3	3	3
127	8891049	1	3	3	3
144	8890999	1	3	3	3
121	TN084622-5270	2	2	1	1
111	MX084613-5190	2	2	1	1
125	TN082115-0020	2	2	1	1
115	MX082154-0140	2	2	1	1

TABLA 4.8: (Continúa)

Consecutivo	Numero de parte	ABC	XYZ	K-medias	Jerárquico
118	MX014380-1140	2	4	1	1
116	MX082152-0090	2	4	3	3
110	MX084641-5070	2	4	3	3

Para mostrar un poco la distribución entre las clasificaciones tradicionales y las propuestas, se puede observar para el método ABC la figura 4.10, se muestra desproporcionado con respecto a los grupos generados.

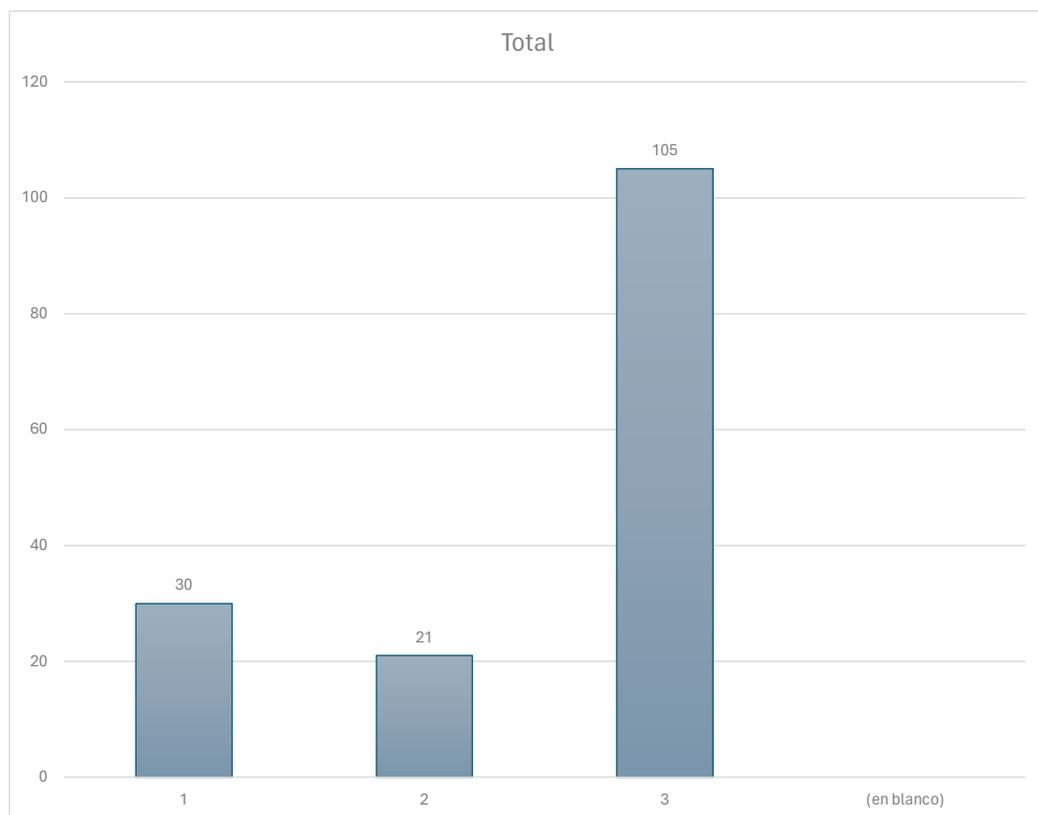


FIGURA 4.10: Distribución de los grupos generados con el método ABC

El método XYZ muestra también una clasificación muy desproporcionada en

la figura 4.11.

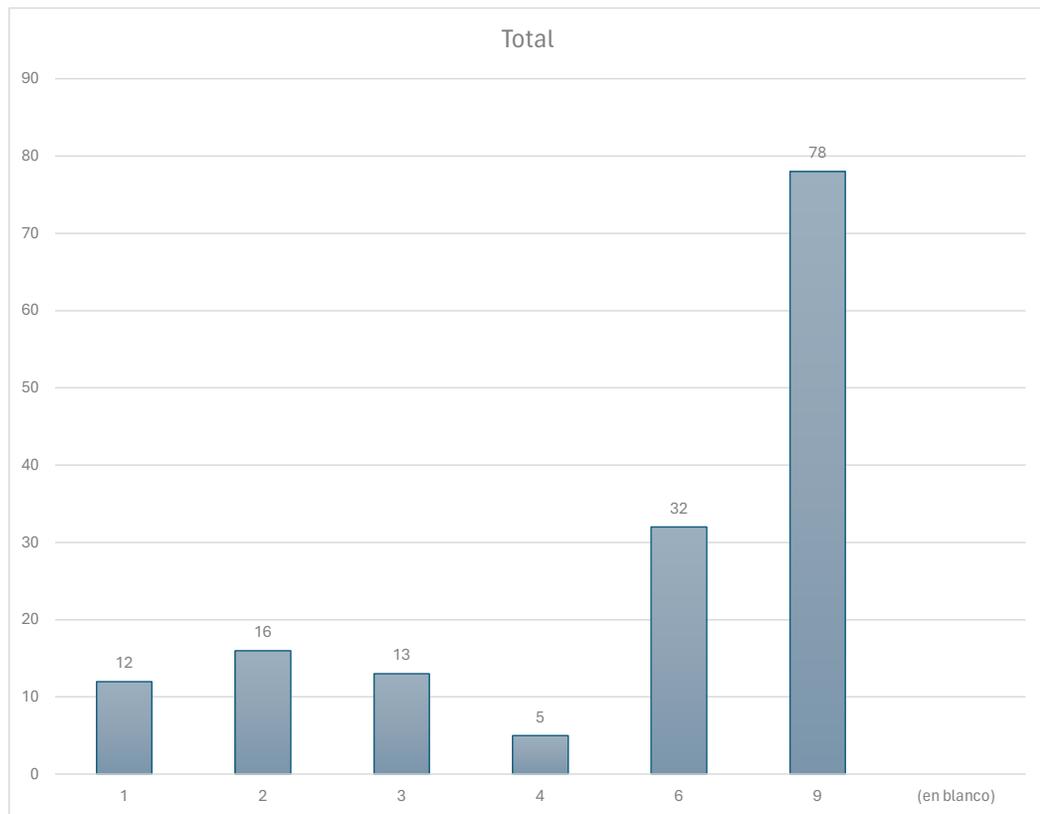


FIGURA 4.11: Distribución de los grupos generados con el método XYZ

Cuando observamos los clústers con el tratamiento de datos sugerido tanto el método jerárquico (figura 4.12) como el método de k-medias (figura 4.13) son un poco más uniformes en la distribución. Evitando grandes grupos o muy pequeños que no se puedan analizar o aglomerar correctamente los datos.

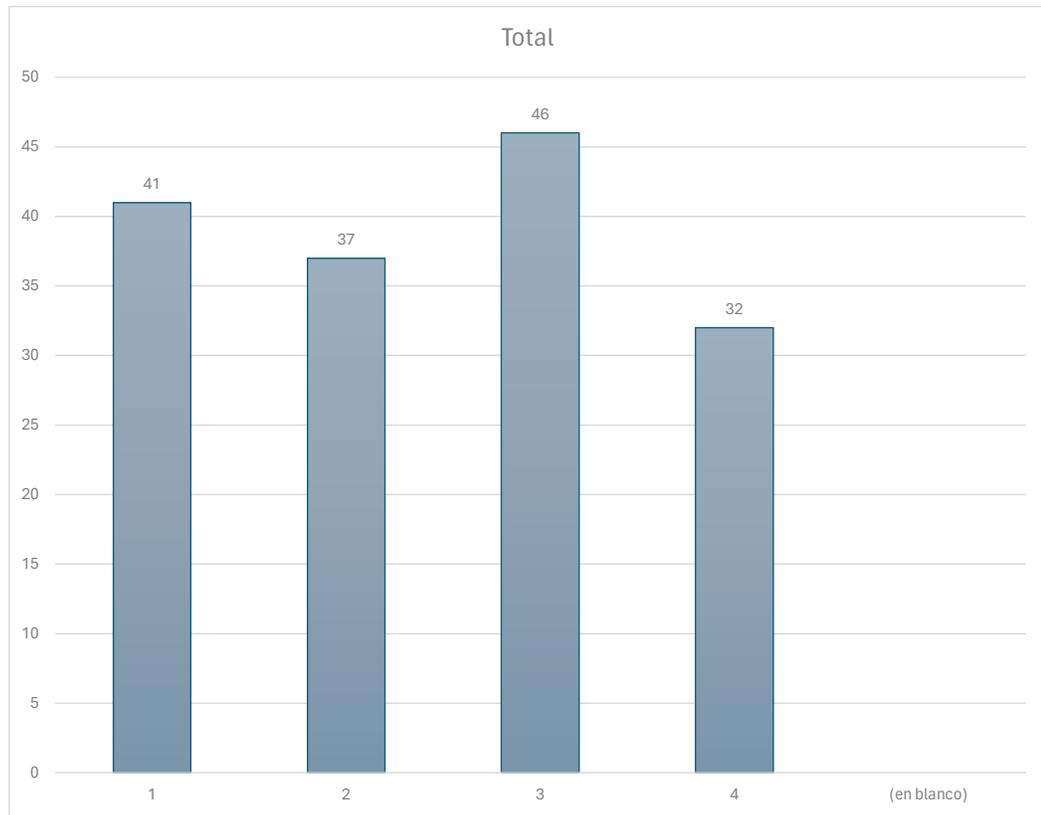


FIGURA 4.12: Distribución de los grupos generados con el método de clustering jerárquico

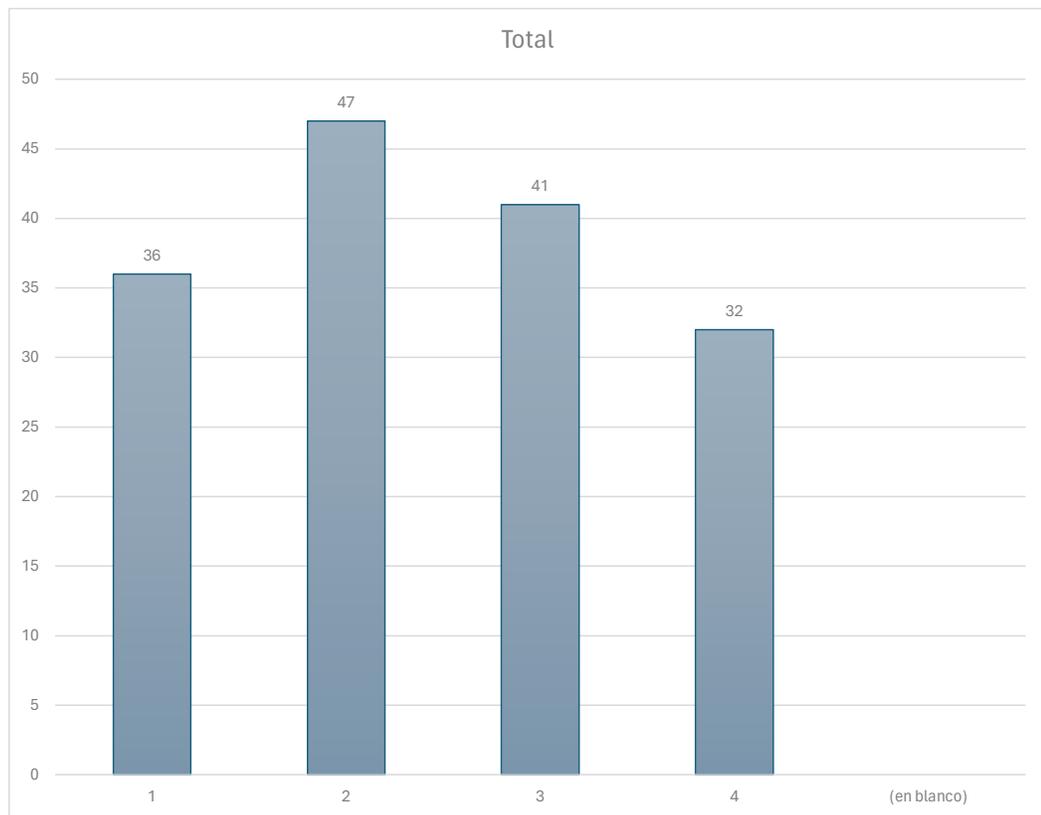


FIGURA 4.13: Distribución de los grupos generados con el método clustering de k-medias

Para finalizar, la coincidencia en la calidad de los clústers del método del *clustering* jerárquico contra el método de clúster con k-medias arroja una similitud del 90 % entre ellos, mientras que la similitud con el método de ABC y XYZ es del 50 %, esto se debe a que los últimos, no consideran las variables de interés adicionales e importantes para la persona que toma las decisiones.

## CAPÍTULO 5

# CONCLUSIONES

---

El análisis ABC clasifica los inventarios según su importancia y valor, permitiendo una mejor gestión de los artículos más relevantes. Por otro lado, el análisis XYZ considera las fluctuaciones en la demanda, clasificando los artículos en función de su variabilidad.

Estos modelos se han aplicado en diversos contextos, sea el giro industrial que sea, y son utilizados para reducir costos de almacenamiento sin comprometer al cliente. En la industria automotriz, se han utilizado para clasificar y gestionar los inventarios de componentes y ensambles finales que almacenan estas empresas.

Los resultados obtenidos revelaron que tanto el *clustering* jerárquico como el de k-medias, ofrecen una clasificación más precisa y detallada al considerar variables adicionales importantes para la toma de decisiones, en comparación con los métodos ABC y el XYZ.

El método de k-medias demostró ser efectivo al agrupar los datos en clústers óptimos, al minimizar la varianza dentro de cada clúster y maximizar la distancia entre ellos. Se observó que la elección del número adecuado de clústers es crucial para obtener una separación óptima de los datos.

Por otro lado, el *clustering* jerárquico también mostró resultados prometedores

al proporcionar una clasificación uniforme y significativa de los datos. La similitud del 90 % entre los clústers generados por el *clustering* jerárquico y k-medias refuerza la validez y consistencia de ambos métodos.

En contraste, los métodos tradicionales ABC y XYZ mostraron una clasificación desproporcionada y menos precisa, ya que no consideran variables importantes adicionales para la clasificación de inventarios.

En resumen, los resultados sugieren que el uso de métodos de agrupamiento como el *clustering* jerárquico y k-medias puede mejorar significativamente la precisión y relevancia de la clasificación de inventarios en comparación con los métodos tradicionales. Estos enfoques proporcionan una base sólida para la toma de decisiones informada y la optimización de políticas de inventario en entornos empresariales.

## 5.1 CONTRIBUCIONES

Esta investigación a pesar de que los datos fueron obtenidos de una empresa dedicada al sector automotriz, puede también ser utilizada para cualquier giro industrial diferente. Se puede decir, que con los resultados obtenidos, se podrá llegar a contribuir de manera que podamos identificar las variables más críticas, aparte de los costos, rotación y demanda; ya que informa una nueva manera de considerar más variables que puedan llegar a ser importantes para las empresas y esto ayudará a que puedan planear mejor sus componentes, materias primas, inventarios en general y controlen más preciso contemplando estas variables.

Por lo tanto, es que la contribución que esta investigación favorece es para los profesionistas, investigadores y empresas ó almacenes a que puedan darles un panorama más amplio de contemplar variables diferentes a las del precio y demanda, para que puedan tener una mejor planificación y control de inventarios, para evitar mermas y gastos innecesarios de dinero que puedan afectar en la cadena de suministro y a la rentabilidad de las empresas.

## 5.2 TRABAJO FUTURO

El trabajo a futuro sobre esta investigación es que abre una nueva forma de investigar y usar las técnicas de clasificación de inventarios por medio de herramientas no supervisadas, las cuáles pueden llegar a combinarse con la optimización de resultados y con otro tipos de variables como lo puede ser aspectos cualitativos o las formas de como son los materiales o el producto, que llegue a considerarse para su evaluación.

Esto también puede llegar a resultar en que no se evalúe un solo proceso, si no, más procesos al mismo tiempo y puedan evaluarse las formas ya sean regular, irregular o incómodas y que puedan también llegar a considerarse en productos mucho más robustos o tipos de industrias diferentes.

Adicional a esto, también permitirá mayores investigaciones con herramientas del machine learning para que sea mucho más fácil y más rápido los resultados, por medio de automatizaciones y programaciones.

## APÉNDICE A

# APÉNDICE

---

## A.1 LIBRERÍAS UTILIZADAS EN PROGRAMACIÓN R.

### A.1.1 LIBRERÍAS UTILIZADAS

En el contexto de programación en R, las "librerías", o "paquetes" son conjuntos de funciones y herramientas que extienden las capacidades del lenguaje R. Estas librerías están desarrolladas por la comunidad de usuarios y programadores de R y se comparten de manera gratuita para que otros usuarios puedan utilizarlas y beneficiarse de sus funcionalidades.

Las librerías en R son fundamentales para realizar análisis de datos, modelado estadístico, visualizaciones gráficas y muchas otras tareas. Al cargar una librería en R, se hacen disponibles todas las funciones y métodos que contiene, lo que permite al usuario acceder a potentes herramientas de manera rápida y sencilla. Aquí se describen las librerías que fueron utilizadas para la ejecución del modelo propuesto:

**library(dplyr)**: Esta librería proporciona un conjunto de funciones para manipulación y transformación eficiente de datos. Con dplyr, es fácil realizar tareas como filtrar filas, seleccionar columnas, agrupar datos y realizar operaciones de resumen.

**library(ggplot2)**: Esta librería es ampliamente conocida por su capacidad para crear gráficos y visualizaciones de datos altamente personalizadas y estéticamente atractivas. Se basa en el concepto de "gramática de gráficos" que permite al usuario crear gráficos complejos utilizando capas, ejes, etiquetas, colores y temas.

**library(readr)**: Esta librería proporciona funciones para leer y escribir datos en diferentes formatos de archivos, como CSV, TSV, Excel, entre otros. Es útil para la importación y exportación de datos en R de una manera sencilla y eficiente.

**library(randomForest)**: Esta librería implementa el algoritmo de Bosques Aleatorios (*random forest*), que es un método de aprendizaje automático utilizado tanto para tareas de clasificación como de regresión. Los Bosques Aleatorios son conocidos por su precisión y capacidad para manejar grandes conjuntos de datos con alta dimensionalidad.

**library(randomForestExplainer)**: Esta librería complementa a *random forest* y se utiliza para interpretar y explicar los resultados de los modelos creados con Bosques Aleatorios. Proporciona herramientas para analizar la importancia de las variables predictoras y entender cómo contribuyen al rendimiento del modelo.

**library(cluster)**: Esta librería se enfoca en el análisis y agrupamiento de datos. Proporciona funciones para realizar diferentes técnicas de *clustering* y segmentación, como el algoritmo *k-means*, agrupamiento jerárquico y otras técnicas de agrupamiento basadas en distancia.

**library(DT)**: Esta librería facilita la creación de tablas interactivas y dinámicas en R mediante la utilización de la biblioteca *DataTables* de JavaScript. Permite crear tablas con capacidad de búsqueda, filtrado y paginación, lo que resulta útil para presentar y explorar grandes conjuntos de datos.

**library(knitr)**: Esta librería es ampliamente utilizada para la generación de informes y presentaciones en R. Permite combinar código R, resultados de código, y texto en un solo documento, lo que facilita la comunicación de resultados y análisis

de datos.

Cada una de estas librerías tiene un propósito específico y proporciona funcionalidades adicionales que extienden las capacidades de R en diferentes áreas, como manipulación de datos, modelado estadístico, aprendizaje automático, visualización y generación de informes. Al utilizar estas librerías, los usuarios de R pueden aprovechar un conjunto diverso de herramientas para abordar una amplia variedad de tareas en el análisis de datos y ciencia de datos.

# BIBLIOGRAFÍA

---

- ADAMS, R. P. (2009), «Hierarchical agglomerative clustering», en *Proc. Ninth siam data mining conf.(SDM'09)*, Σελ, págs. 510–516.
- ALTAY GUVENIR, H. y E. EREL (1998), «Multicriteria inventory classification using a genetic algorithm», *European Journal of Operational Research*.
- BABAI, M., T. LADHARI y I. LAJILI (2014), «On the inventory performance of multi-criteria classification methods: empirical investigation», *International Journal of Production Research*, **0**(na), pág. na.
- BALAJI, K. y V. SENTHIL KUMAR (2014), «Multicriteria inventory ABC classification in an automobile rubber components manufacturing industry», *Procedia CIRP*.
- CHOPRA, S. y P. MEINDL (2008), *Contrucción de un marco estratégico para analizar la cadena de suministro*, tercera edición, Pearson, Prentice Hall, entender la cadena de suministro.
- CUTLER, A., R. CUTLER y J. R. STEVENS (2012), «Random forests», *Ensemble machine learning: methods and applications*, págs. 157–175.
- DANIYA, T., M. GEETHA y S. KUMAR (2020), «Classification and regression trees with gini index», *Advances in Mathematics: Scientific Journal 9*.
- DOUGLAS, T. (2000), *Control de inventarios*, segunda edición, MPrentice Hall.

- EHRHARDT y BRIGHAM (2007), «Administración del inventario: elemento clave para la optimización de las utilidades en las empresas», *Disponible en: <https://www.redalyc.org/articulo.oa?id=465545892008>*.
- EWING, J. y C. P (2021), «Cómo la escasez de carros pone en peligro la economía mundial», Excerpted from magazine The New York Times, URL <https://www.nytimes.com/es/2021/11/03/espanol/escasez-industria-automotriz.html>.
- FLORES, B., D. OLSON y V. DORAI (1992), «Management of multicriteria inventory classification», *Math Comput Modeling*, **16**(12), págs. 71–82.
- FLORES, B. y D. C. WHYBARK (1987), «Implementing multiple criteria ABC analysis», *Journal of Operations Management*, **7**(1-2), págs. 79–85.
- GÓMEZ HERNÁNDEZ, F. (2022), *Diseño de un sistema de gestión de inventarios para una compañía comercializadora de repuestos del sector automotriz.*, Tesis de Maestría, Universidad de La Sabana.
- LIMÓN DELOSANTOS, J. D. (2015), *Efectos de implementación de la manufactura esbelta en una línea de ensamble de la industria automotriz.*, Tesis de Maestría, Universidad Autónoma de San Luis Potosí.
- LOLLIA, F., A. ISHIZAKA, R. GAMBERINIA, E. BALUGANIA y B. RIMINIA (2017), «Decision trees for supervised multi-criteria inventory classification», *Procedia Manufacturing*.
- LÓPEZ SÁNCHEZ, V. M. (2007), *Un modelo de simulación dinámica de políticas de inventario bajo una estrategia de manufactura mixta en el sector automotriz mexicano.*, Tesis de Maestría, Instituto Tecnológico y de Estudios Superiores de Monterrey.
- M., G., A. A, J. M, L. T, S. R y G. M (2020), «COVID-19 and economy.», *Dermatologic therapy*, **33**(4).

- MILLSTEIN, M. A., L. YANG y H. LI (2014), «Optimizing ABC inventory grouping decisions», *Int. J. Production Economics*.
- MORA GARCÍA, L. A. (2000), *Gestión logística en centros de distribución, bodegas y almacenes*, segunda edición, Ecoe Ediciones.
- NAHMIA, S. (2007), *Production and Operation analysis*, quinta edición, McGraw Hill, control de inventarios sujeto al conocimiento de la demanda.
- PANDYA, B. y H. THAKKAR (2016), «A review on inventory management control techniques: ABC-XYZ analysis», *Journal on Emerging trends in Modelling and Manufacturing*.
- PAREDES RODRIGUEZ, A., V. CHUD PANTOJA y J. C. OSORIO (2019), «Sistema de control de inventarios multicriterio difuso para repuestos», *Scientia et Technica Año XXIV*, **24**(02), págs. 592–602.
- PRASAD, S. (1994), «Classification of inventory models and systems», *Elsevier*, **34**(34), págs. 209–222.
- RUÍZ, H. (2018), *Exceso de inventarios y la problemática de gestión*, Team, <https://www.teamnet.com.mx/blog/2018/04/exceso-de-inventarios>, ¿Cuáles son los principales problemas que causan el exceso de inventario?
- SINGH, P. (2014), «A study on inventory management with reference to leading automobile industry», : *International Journal of Management, Information Technology and Engineering (BEST: IJMITE)*, **2**(5), págs. 15–28.
- TANG, Z., Z. MEI, W. LIU y Y. XIA (2020), «Identification of the key factors affecting Chinese carbon intensity and their historical trends using random forest algorithm», *Journal of Geographical Sciences*, **30**, págs. 743–756.
- VILÁ BAÑOS, R., M. J. RUBIO HURTADO, V. BERLANGA SILVENTE y M. TORRADO FONSECA (2014), «Cómo aplicar un cluster jerárquico en SPSS», [*En línea*] *REIRE, Revista d'Innovació i Recerca en Educació*, **7**(1), págs. 113, 127, <http://www.ub.edu/ice/reire.htm>.

---

ZOWID, F., M. BABAI, M. R. DOUISSA y Y. DUCQ (2019), «Multi-criteria inventory ABC classification using Gaussian Mixture Model», *IFAC PapersOnLine*.

# RESUMEN AUTOBIOGRÁFICO

---

Fátima Abygail Dragustinovis López

Candidato para obtener el grado de  
Maestría en Logística y Cadena de Suministro

Universidad Autónoma de Nuevo León  
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:

CLASIFICACIÓN DE COMPONENTES CRÍTICOS BASADOS EN  
TÉCNICAS DE CLUSTERIZACIÓN

Yo, Fátima Abygail Dragustinovis López nací un 14 de abril del año 1995, en la Ciudad de Monterrey, N.L. mis padres son Jorge Dragustinovis y Olga López. Soy egresada de la Facultad de Contaduría Pública y Administración de la UANL, como Licenciada en Negocios Internacionales. Cuento con 8 años de experiencia en el área de control de producción en el sector automotriz manufacturero, mis conocimientos están enfocados en la cadena de suministro como planeadora de materiales y producto terminado, así mismo en el área de logística e importaciones y exportaciones.