



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**DISEÑO DE INVESTIGACIÓN DE PREDICCIÓN DE INVENTARIOS Y
COMPRAS MÍNIMAS PARA UNA INDUSTRIA DE IMPORTACIÓN Y
COMERCIALIZACIÓN**

Jorge Armando Rodríguez Carvajal

Asesorado por el MSc. Ing. Neftalí de Jesús Calderón Mendez

Guatemala, abril de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**MODELO DE PREDICCIÓN DE INVENTARIOS Y COMPRAS MÍNIMAS PARA
UNA INDUSTRIA DE IMPORTACIÓN Y COMERCIALIZACIÓN**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA

POR

JORGE ARMANDO RODRÍGUEZ CARVAJAL

ASESORADO POR MSC. ING. NEFTALÍ DE JESÚS CALDERÓN MENÉNDEZ

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, ABRIL DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Armando Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Julio I. González Podszueck
EXAMINADOR	Ing. Jorge Luís Álvarez Mejía
EXAMINADOR	Ing. Sergio Silva
EXAMINADOR	Ing. Edwin Sánchez
SECRETARIO	Ing. Francisco J. González López

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

MODELO DE PREDICCIÓN DE INVENTARIOS Y COMPRAS MÍNIMAS PARA UNA INDUSTRIA DE IMPORTACIÓN Y COMERCIALIZACIÓN

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Posgrado, con fecha 12 de enero de 2022.

Jorge Armando Rodríguez Carvajal



EEPFI-PP-0159-2022

Guatemala, 12 de enero de 2022

Director
null
Escuela De Ingenieria En Sistemas
Presente.

Estimado Ing. Alonzo

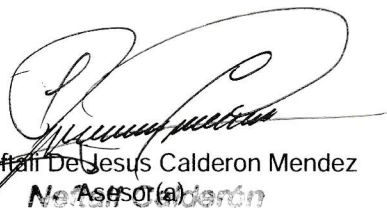
Reciba un cordial saludo de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería.

El propósito de la presente es para informarle que se ha revisado y aprobado el Diseño de Investigación titulado: **MODELO DE PREDICCIÓN DE INVENTARIOS Y COMPRAS MÍNIMAS PARA UNA INDUSTRIA DE IMPORTACIÓN Y COMERCIALIZACIÓN**, el cual se enmarca en la línea de investigación: **Análisis de datos - Análisis de datos**, presentado por el estudiante **Jorge Armando Rodríguez Carvajal** carné número **8511537**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la Maestría en ARTES en Ingeniería Para La Industria Con Especialidad En Ciencias De La Computación.

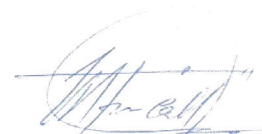
Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

Atentamente,

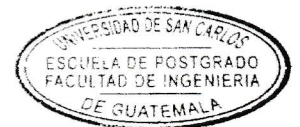
"Id y Enseñad a Todos"



Mtro. Neftalí De Jesús Calderón Mendez
Asesor(a)
Ingeniero en Ciencias y Sistemas
Colegiado No. 8051



Mtro. Mario Renato Escobedo Martínez
Coordinador(a) de Maestría



Mtro. Edgar Darío Álvarez Cotí
Director
Escuela de Estudios de Postgrado
Facultad de Ingeniería





EEP-EICS-0159-2022

El Director de la Escuela De Ingenieria En Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **MODELO DE PREDICCIÓN DE INVENTARIOS Y COMPRAS MÍNIMAS PARA UNA INDUSTRIA DE IMPORTACIÓN Y COMERCIALIZACIÓN**, presentado por el estudiante universitario **Jorge Armando Rodriguez Carvajal**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería en esta modalidad.

ID Y ENSEÑAD A TODOS

A handwritten signature in black ink is written over an official oval stamp. The stamp contains the text: 'UNIVERSIDAD DE SAN CARLOS DE GUATEMALA', 'DIRECCION DE INGENIERIA EN CIENCIAS Y SISTEMAS'.

Ing. Carlos Gustavo Alonzo
Director
Escuela De Ingenieria En Sistemas


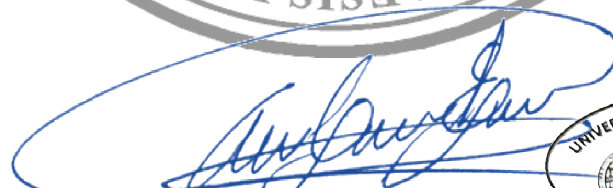
Guatemala, enero de 2022

Decanato
Facultad de Ingeniería
24189101- 24189102
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.314.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN DE PREDICCIÓN DE INVENTARIOS Y COMPRAS MÍNIMAS PARA UNA INDUSTRIA DE IMPORTACIÓN Y COMERCIALIZACIÓN**, presentado por: **Jorge Armando Rodríguez Carvajal**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:



Inga. Aurelia Anabela Cordova Estrada ★

Decana

Guatemala, abril de 2022

AACE/gaoc

ACTO QUE DEDICO A:

- Dios** Por darme su bendición, inspiración, sabiduría, salud y ser mi guía en este camino, para obtener uno de mis objetivos más deseados.
- Mis padres** Mariano Rodríguez y Lila Carvajal. Por su amor, esfuerzo y por la oportunidad de educación que me brindaron.
- Mi esposa** Patty Carrera, por su amor, cariño, apoyo, compañía y soporte, siempre impulsándome y alentándome a no darme por vencido.
- Mis hijos** Javier y Daniela Rodríguez, por ser fuente de mi inspiración, que me permite siempre seguir adelante.

AGRADECIMIENTOS

Universidad de San Carlos de Guatemala	Por ser el centro de estudios superiores que me permitió acrecentar mis conocimientos, la dedicación y la perseverancia, pilares fundamentales en mi vida profesional.
Facultad de Ingeniería	Por ser mi segunda casa y por haberme permitido pasar por sus aulas, para vivir momentos alegres y experiencias inolvidables como estudiante.
Mis catedráticos	Por brindarme sus conocimientos y experiencias, que definitivamente me sirvieron para enfrentar el mundo profesional, lleno de tantos retos y vicisitudes.
Mis compañeros	Por su apoyo y tantos momentos de alegría y esfuerzo en el transcurso de la carrera.
Mis abuelitos	Manuel Rodriguez e Hilda Leiva, por su apoyo, guía y cariño de toda la vida.
Mis familiares y amigos	Por siempre estar allí, en las buenas y las no tan buenas.

ÍNDICE GENERAL

INDICE DE ILUSTRACIONES.....	V
LISTA DE SÍMBOLOS.....	VII
GLOSARIO.....	IX
1. INTRODUCCIÓN	1
2. ANTECEDENTES	3
3. PLANTEAMIENTO DEL PROBLEMA	9
3.1. Contexto general	9
3.2. Descripción del problema	9
3.3. Formulación del problema	10
3.4. Delimitación del problema	10
4. JUSTIFICACIÓN	11
5. OBJETIVO.....	13
5.1 General.....	13
5.2 Específico	13
6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN	15
7. MARCO TEÓRICO.....	17
7.1. Modelos de inventario.....	17
7.1.1. Conceptos y términos	17

	7.1.1.1.	Definición de inventario.....	18
	7.1.1.2.	Modelos comerciales y de producción	18
	7.1.1.3.	Modelos determinísticos y probabilísticos	19
	7.1.1.4.	Tiempo de entrega (lead time)	19
7.2.		Cadena de suministro	20
	7.2.1.	Definición de cadena de suministro.....	20
	7.2.2.	Etapas de una cadena de suministro	20
	7.2.3.	Demanda y oferta.....	21
	7.2.4.	Aranceles	22
	7.2.5.	Efecto látigo.....	23
7.3.		Modelos de análisis predictivo	24
	7.3.1.	Modelos predictivos.....	24
	7.3.2.	Modelos descriptivos.....	24
	7.3.3.	Modelos de decisión.....	25
7.4.		Técnicas de análisis predictivo.....	25
	7.4.1.	Regresión lineal.....	26
	7.4.2.	<i>Machine learning</i>	28
	7.4.3.	Redes neuronales	29
	7.4.3.1.	Máquinas de vectores de soporte	30
	7.4.3.2.	K-vecinos más cercanos.....	31
7.5.		Minería de datos	33
	7.5.1.	Fases del proceso de minería de datos	33
	7.5.1.1.	Comprensión del negocio	34
	7.5.1.2.	Compresión de los datos	34
	7.5.1.3.	Preparación de los datos	34
	7.5.1.4.	Modelado.....	35
	7.5.1.5.	Evaluación	35
	7.5.1.6.	Implementación	35
	7.5.1.7.	Tipos de información	35
7.6.		Lenguajes aplicados a <i>Machine Learning</i>	36

7.6.1.	Lenguaje R	36
7.6.2.	Lenguaje Python.....	37
8.	PROPUESTA DE ÍNDICE DE CONTENIDOS	39
9.	METODOLOGÍA.....	43
9.1.	Características del estudio	43
9.2.	Unidad de análisis	43
9.3.	VARIABLES del estudio	44
9.4.	Fases del Estudio	45
9.4.1.	Revisión documental	45
9.4.2.	Casos de estudio	45
9.4.3.	Recolección de datos	46
9.4.4.	Revisión de datos	46
9.4.5.	Depuración de datos.....	47
9.4.6.	Presentación de resultados	47
9.4.7.	Redacción de informe final	47
9.5.	Técnicas de análisis de información	48
9.5.1.	Preparación de datos.....	48
9.5.2.	Tamaño del conjunto de datos.....	50
9.5.3.	Regresión lineal	51
9.5.4.	Árboles de decisión	52
9.5.5.	Redes neuronales.....	52
9.5.6.	Algoritmos genéticos	53
9.5.7.	Series temporales.....	53
9.5.8.	Análisis de la demanda.....	55
9.5.9.	Análisis del nivel de inventario.....	55
9.5.10.	<i>Machine learning</i> y lenguaje R.....	57

10.	CRONOGRAMA	59
11.	FACTIBILIDAD DEL ESTUDIO	61
11.1.	Factibilidad operativa	61
11.2.	Factibilidad técnica.....	62
11.3.	Factibilidad económica.....	62
12.	REFERENCIAS	65

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	Etapas de la cadena de suministro	21
2.	Efecto látigo.....	23
3.	Regresión lineal.....	26
4.	Red neuronal.....	29
5.	SVM solución lineal en el espacio	31
6.	Aplicación k-vecinos más cercanos.....	32
7.	Diagrama CRISP-DM para minería de datos	33
8.	RStudio – Ambiente de desarrollo integrado	37
9.	Anaconda Navigator with Spyder editor	38
10.	Niveles de inventario	57
11.	Cronograma de actividades.....	59

TABLAS

I.	Descripción de variables del estudio	44
II.	Venta histórica formateada para machine learning	48
III.	Conjunto de datos (dataset) históricos de venta	49
IV.	Conjunto de datos (dataset) para entrenamiento y prueba	50
V.	Actividades y costos de la investigación.....	63

LISTA DE SÍMBOLOS

Símbolo	Significado
i7	Procesador de computadora con arquitectura x86 de 64 bits.
GB	Un gigabyte es una cantidad de almacenamiento con un valor de 10e+9 bytes.
x64	Versión 64 bits del conjunto de instrucciones x86. Soporta una cantidad mucho mayor de memoria virtual y memoria física, permitiendo a los programas almacenar grandes cantidades de datos en la memoria.

GLOSARIO

Lenguaje Python	Es un lenguaje de programación de alto nivel, se utiliza para desarrollar aplicaciones de todo tipo. Se trata de un lenguaje interpretado, no es necesario compilarlo para ejecutar el código.
Lenguaje R	Entorno de software libre y lenguaje de programación interpretado. Ejecuta las instrucciones directamente, sin previa compilación. Comúnmente utilizado para la computación estadística y gráfica.
RAM	Memoria de acceso aleatorio que conforma la memoria principal de un equipo de cómputo.
RStudio	Entorno de desarrollo utilizando el lenguaje R.
SQL	Lenguaje estructurado de consulta, utilizado en las bases de datos relacionales.

1. INTRODUCCIÓN

Es un hecho indiscutible que todas las cadenas de suministro tienen el reto de la planificación. Cuanto mejor evaluemos el futuro, mejor podremos prepararnos. En este contexto, el reto de lidiar con la incertidumbre del comportamiento futuro y cómo protegerse contra esta incógnita siempre ha sido crucial para cada cadena de suministro. Desde negociar volúmenes de contratos con proveedores hasta establecer objetivos de stock de seguridad, todo se relaciona con poder predecir el futuro con un nivel de certeza aceptable.

Por otro lado, se sabe que las grandes corporaciones a nivel mundial utilizan software de cálculo predictivo desde muchos años atrás. Esto les permitió utilizar el pronóstico estadístico como la base fundamental de su proceso de planificación de ventas y operaciones. Estos modelos de pronóstico estadístico fueron planteados hace algún tiempo y no han cambiado mucho desde entonces. En la actualidad, existen nuevas opciones gracias al crecimiento y evolución constante de la tecnología, a la capacidad y el poder informático, al flujo y almacenamiento de grandes cantidades de datos, al surgimiento de la minería de datos y *machine learning*, lo que permite a las empresas marcar la diferencia.

El presente trabajo de investigación pretende aportar, a las empresas que realizan las actividades de importación y luego de comercialización de sus productos, un modelo que sea de utilidad en el proceso de reabastecimiento de inventario, basado en el aprendizaje de su propia información. Para ello se llevó a cabo un proceso de recolección, revisión y depuración de datos, se utilizaron modelos matemáticos y algoritmos de clasificación y agrupamiento, que

sustancialmente reducen el error en la precisión y predicción del inventario mínimo y cantidad de compra de productos.

En el primer capítulo se describen los antecedentes de trabajos que utilizaron minería de datos y *machine learning* en distintos contextos del ámbito empresarial y social, así como también estudios sobre modelos de sistemas de ventas e inventario y otros sobre herramientas de uso frecuente para la implementación estos conceptos.

En el segundo capítulo se hace el planteamiento del problema desde su contexto general hasta su delimitación. Luego en el tercer capítulo se hace la justificación y razones importantes por la cuales se elabora este protocolo. Posteriormente, en el cuarto capítulo, se presentan los objetivos generales y específicos bajo los cuales se enmarca este trabajo.

Las necesidades que se cubren y el esquema general de la solución se esbozan en el quinto capítulo. En el sexto capítulo se presenta el marco teórico que sustenta la investigación, en el cual se definen entre otros conceptos, modelos de inventario, lenguajes aplicados a aprendizaje automatizado, cadena de suministro, modelos de análisis predictivo, aprendizaje automatizado y minería de datos.

En el capítulo siete se sugiere el índice de contenidos que tendrá el trabajo final de graduación que tiene su base en este protocolo. Luego de esta propuesta de contenido, se describe la metodología a utilizar en la investigación en el capítulo ocho, para que después se presente en el capítulo nueve el cronograma de desarrollo del trabajo con los tiempos estimados de cada una de las fases y actividades. Finalmente, en el capítulo diez se pueden encontrar el estudio de factibilidad del proyecto.

2. ANTECEDENTES

La minería de datos combinada con el aprendizaje automatizado permite establecer modelos de predicción en muchas áreas de la vida cotidiana y de las empresas. En cuanto a estos conceptos Piscoya Ordoñez, L. (2016), elaboró un estudio sobre aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular de la región de Lambayeque y quien propone utilizar las técnicas la minería de datos como una herramienta que permita realizar predicciones sobre alumnos que se matriculen en los próximos años, sirviendo de apoyo a la institución para las decisiones futuras que puedan tomar.

En dicha investigación se establece la utilización de algoritmos ETS (suavizamiento exponencial), para la predicción de series temporales y redes neuronales, además se construye un modelo de minería de datos basado en series de tiempo, logrando establecer las predicciones de deserción escolar en la educación primaria y secundaria en la región de Lambayeque, Ugel Chiclayo, Perú.

Por otro lado, Salas-Navarro, K., Miguél-Mejía, H. y Acevedo-Chedid, J. (2016) quienes realizaron: metodología de gestión de inventarios para determinar los niveles de integración y colaboración en una cadena de suministro, afirman en su investigación que, para ser competitivas, las empresas deben implementar estrategias que relacionen estrechamente el control del inventario con la cadena de suministro. Además, presentan una metodología con varios pasos e indicadores, los cuales permiten analizar el rendimiento de la cadena de suministro.

La disponibilidad adecuada de inventario es la clave aquí, su gestión es horizontal a todos los eventos o procesos de la cadena de suministro, lo que permite estar prevenido para eventualidades que nunca faltan, en aras de lograr siempre un nivel de servicio adecuado.

En ese mismo orden de ideas, Jara-Cordero, S., Sanchez-Partida, D. y Martinez-Flores, J. (2017) realizaron: Análisis para la mejora en el manejo de inventarios de una comercializadora y describen el método para el cálculo de la cantidad económica a ordenar (EOQ por sus siglas en inglés) y el Punto de Reorden (ROP) con los artículos de una empresa comercializadora. Dentro de los objetivos del estudio se tienen la reducción de los pedidos pendientes de entrega (*backorder*) y la mejora del servicio al cliente, por ende, la mejora de los costos asociados al proceso logístico. Con el fin de lograr la realización y cumplimiento de los objetivos propuestos, se pusieron en práctica metodologías de administración e investigación de operaciones y cadena de suministro.

Se contaba con un estudio antiguo y los valores de EOQ y ROP estaban obsoletos, por lo que fue necesario actualizar el sistema de control de inventarios con los nuevos valores EOQ y ROP para todos los artículos de la empresa y así aplicar las metodologías escogidas.

También sobre este tópico Chamorro Corea, J., Díaz Camejo, J., Fuentes Espinoza, O. y Lovo Gutiérrez, H. (2018) presentaron: Política de inventarios máximos y mínimos en cadenas de suministro multinivel. Caso de estudio: una empresa de distribución farmacéutica, un sobre la política de inventarios máximos y mínimos de una empresa, en cadenas de suministros multinivel. Dicha política se evalúa por medio de los indicadores de nivel de servicio y rotación de inventarios. Entre los métodos de optimización utilizados, se incluyeron algoritmos metaheurísticos de búsqueda tabú y búsqueda dispersa,

cuyos resultados experimentales derivados de las simulaciones demuestran que la solución derivada de la teoría garantiza un alto nivel de servicio y posibles mejoras en la rotación de inventarios.

Como un complemento de estos estudios, Burger, S. (2018) publicó: *Introduction to Machine Learning with R (Rigorous Mathematical Modeling)*, en donde brinda una muestra de los modelos, algoritmos y entrenamiento de datos disponibles para aprendizaje automatizado utilizando el lenguaje R.

En la misma se profundiza sobre los modelos de regresión lineal más utilizados en los negocios y la ciencia, y brinda la forma de utilizar redes neuronales de una o varias capas que nos permita calcular los resultados y una muestra del funcionamiento de los modelos basados en árboles, incluidos los árboles de decisiones populares. Basado en estos conceptos minería de datos y aprendizaje automatizado, Gutierrez Alvarez, R. (2019) realizó: Implementación de un modelo predictivo basado en data mining para la mejora de la gestión de ventas en la Distribuidora Jimenez Iriarte, S.A. En este trabajo se establece un modelo que permite clasificar los clientes con la finalidad de ofrecerles promociones y descuentos personalizados, sugeridos de ventas conociendo su comportamiento de compra y planificar la preventa de la fuerza de ventas.

Además, realiza un análisis comparativo de las técnicas de data mining como: Holt, Holtwinters, *suavizamiento exponencial (ETS)* y movimiento promedio integrado autorregresivo (Arima), para determinar la que más aplica para el caso de estudio, acompañado de la metodología CRISP-DM la cual posee un método probado para orientar trabajos de minería de datos. Finalmente, se puede concluir, que la investigación establece que el uso de técnicas de data mining permiten a las empresas mejorar muchos elementos que interactúan en

los procesos del negocio, tales como su nivel de inventario y realizar compras basadas en información.

Adicionalmente, otro estudio elaborado hecho por Calatayud, A. y Katz, R (2019), titulado: Cadena de suministro 4,0 (Mejores prácticas internacionales y hoja de ruta para América Latina), nos presenta la situación las empresas en América Latina y la necesidad de impulsar una transformación digital a la cadena de suministro, lo que constituye un aspecto clave para conducir las empresas hacia la cadena de suministro 4,0. Dentro de los objetivos del estudio se encuentran la identificación de la experiencia, mejores prácticas en la transformación digital de la cadena de suministro en economías avanzadas, incluyendo las lecciones aprendidas y, también, presentar las recomendaciones para acelerar este proceso en la región.

Establece que la revolución tecnológica es la base de la cadena de suministro 4,0 y analiza cada una de las tecnologías disponibles como internet de las cosas (IoT), analítica de *big data*, inteligencia artificial y *machine learning*, robótica, impresión 3D y digitalización.

Otra investigación que nos muestra estos conceptos es la realizada por Dongo Pozo, A. y Silva Cama, X. (2020), Quiénes realizaron: Análisis de la minería de datos aplicada en empresas del sector retail. En este trabajo indican que las pequeñas empresas no tienen tanto conocimiento de lo que es minería de datos, en comparación con las grandes empresas que buscan estar siempre a la vanguardia. Muchas de las empresas grandes han logrado un buen posicionamiento en el mercado, porque conocen las necesidades de sus clientes y eso ayuda a que se tomen mejores decisiones orientadas en los mismos.

Adicionalmente, el trabajo establece como se identifica y muestra cómo aplican el proceso de minería de datos algunas empresas del sector retail en sus

operaciones, cuáles son sus fases, su arquitectura, sus modelos y las diferentes técnicas que existen, que les ha permitido tener un crecimiento.

De igual manera encontramos que Lara Cabrera, M. (2020) realizó: Desarrollo de un prototipo que permita realizar el análisis predictivo de delitos de investigación criminal en el departamento de Guatemala. En este trabajo se describe que es posible relacionar el análisis criminal y la predicción del delito, por medio de un enfoque de sistemas de información. Para ello se implementó un proceso de minería de datos, con algoritmos de agrupamiento jerárquico K-means y Difuso, permitiendo al analista conocer e identificar patrones y características de los delitos, agrupándolos en conglomerados homogéneos en el contexto de la incidencia criminal.

Finalmente, existe una publicación hecha por Vandeput, N. (2021): *Data Science for Supply Chain Forecasting*, quien nos presenta un análisis y enfoque práctico, sobre el uso adecuado de modelos tradicionales de pronóstico estadístico (entre otros: media móvil y suavizamiento exponencial) y, por otro lado, establece la aplicación de modelos o algoritmos de *machine learning* que tienen un rendimiento sobresaliente, en términos de precisión del pronóstico, para un conjunto de datos de demanda en la cadena de suministros. Cabe destacar que el aprendizaje es una característica primordial de estos modelos, estos encuentran relaciones ocultas que no es factible descubrir con los modelos estadísticos tradicionales y las aplican sobre los datos futuros.

3. PLANTEAMIENTO DEL PROBLEMA

3.1. Contexto general

Las empresas de importación y comercialización, sobre todo las pequeñas y medianas, requieren de una herramienta que les permita tener un inventario de productos controlado y al nivel óptimo, para que puedan brindar un buen nivel de servicio a sus clientes y que sus almacenes no se encuentren llenos de productos que tienen baja rotación.

Estas empresas no tienen control de la cadena de suministro, deben conocer entre otros: los tiempos de entrega del proveedor, los mínimos de inventario o stock de seguridad y el punto de reorden, lo que les permitiría comprar de una forma inteligente.

3.2. Descripción del problema

Las empresas de importación y comercialización no poseen una herramienta para el control del nivel óptimo de inventario y mucho menos para la gestión de la cadena de suministro, desde el pedido que se realiza al proveedor hasta el despacho que se envía al cliente, pasando por la importación, internación, almacenamiento y logística de distribución.

3.3. Formulación del problema

Las empresas muchas veces realizan compras desmedidas y sin control de sus productos y contraen compromisos con sus proveedores. Esto reduce rápidamente la capacidad de pago y las conduce, muchas veces, a realizar préstamos financieros. Esto se convierte en un círculo vicioso, lo que las puede llevar a la quiebra.

- Pregunta central
 - ¿Cómo desarrollar una mejora en la cadena de suministro, en una empresa de importación y comercialización, con el uso de *machine learning* integrado con su sistema de gestión actual?

- Preguntas auxiliares
 - ¿Cómo se determina la cantidad de compra de productos de inventario que este dentro del tiempo de entrega establecido?
 - ¿Cómo se determina la cantidad mínima de inventario de productos que satisfacen las necesidades de los clientes?
 - ¿Cómo crear un modelo de previsión de inventario utilizando *machine learning*, para optimizar la cadena de suministro?

3.4. Delimitación del problema

El presente trabajo de investigación se circunscribe al estudio de una muestra de empresas que pertenecen al segmento importación y comercialización, porque las variables y condiciones varían significativamente según el segmento al que pertenece la empresa.

4. JUSTIFICACIÓN

La realización del presente trabajo de investigación se justifica en la línea de investigación de minería de datos y aprendizaje automatizado, con un enfoque de aplicación a la industria de importación y comercialización en la Maestría en Ingeniería para la Industria con especialización en Ciencias de la Computación.

En la actualidad las pequeñas y medianas empresas de importación y comercialización necesitan cada vez más apoyarse en la tecnología, necesitan realizar análisis de la información que generan por medio de la inteligencia de negocios y la minería de datos, lo que les permite tomar decisiones más acordes a la realidad del negocio.

La investigación pretende establecer un modelo predictivo orientado a la mejora del proceso de reabastecimiento de inventario y a la optimización de la cadena de suministro. En este proceso intervienen varios elementos clave, entre los que se encuentran el mínimo de inventario o stock de seguridad, el punto de reorden y el tiempo de entrega y una vez conociendo su valor, permiten realizar una compra de productos de forma eficiente e inteligente.

El establecimiento de este modelo predictivo beneficiará a la empresa a disminuir sobrecostos para el cumplimiento de acuerdos comerciales, realizar sugeridos de compra de acuerdo con el comportamiento de la oferta y demanda; tener mayor alcance de las preferencias con los proveedores.

Los beneficios influyen positivamente en el nivel de servicio al cliente impulsando la lealtad de este, la cual será proporcional al valor económico que genere para la empresa. Además, puede servir de base para que otras áreas de

la empresa, a futuro, implementen el uso de la minería de datos y *machine learning*.

5. OBJETIVO

5.1 General

Diseñar un modelo de predicción, basado en minería de datos y *machine learning*, que permita mejorar el proceso de la cadena de suministro.

5.2 Específico

- Establecer un proceso que permita sugerir las cantidades de productos a comprar.
- Determinar los mínimos de inventario con base en la demanda de los clientes.
- Crear un modelo de previsión de inventario, para optimizar la cadena de suministro.

6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN

Existe una carencia de herramientas o modelos que permitan orientar a las empresas en el manejo adecuado de los valores de inventario y compras mínimas, que permitan mejorar positivamente la cadena de suministro. Las grandes empresas han logrado sobresalir en el mercado, porque conocen las necesidades de sus clientes y eso ayuda a que se tomen mejores decisiones orientadas en los mismos.

En tal sentido definitivamente existe la necesidad de establecer las cantidades y montos mínimos a comprar; tomando en cuenta el tiempo de entrega, para que los productos se encuentren disponibles y así satisfacer la demanda de los clientes. Esto nos lleva a establecer un modelo que apoye a la gestión del proceso de cadena de suministro y a conocer el nivel de servicio que se tiene con los clientes y aplicar las acciones de mejora, para que dicho nivel mantenga la continuidad de ventas a los clientes.

Se analizarán todos los productos que manejan la empresa importadora y comercializadora. Como resultado de dicho análisis, se determinarán las cantidades optimas a pedir por producto, tomando en cuenta stock de seguridad, lo que conlleva una reducción del costo de inventario.

Se tomará la información de varios años de historia de su sistema informático actual, lo que permitirá obtener el volumen anual de ventas y compras por producto. Con esta información se realizará por un lado un proceso de pronóstico estadístico para plantear un modelo de optimización que incluye cálculos como media móvil, suavizamiento exponencial simple y doble, según

sea el caso, que nos permitirá pronosticar tendencias lineales, estacionalidad, valores atípicos y errores de previsión. Por otro lado, se aplicará minería de datos y se utilizarán modelos de *machine learning*, aplicando regresión lineal, árboles de decisión, generadores de demanda, *clustering* o redes neurales, lo que permitirá descubrir tendencias y patrones del comportamiento de los productos.

7. MARCO TEÓRICO

7.1. Modelos de inventario

La estructuración de un modelo de inventario basado en sus características, impacto en el negocio, condiciones de venta, tiempo de entrega; permite a las empresas lograr un aprovisionamiento oportuno.

7.1.1. Conceptos y términos

El tema de análisis del inventario y la cadena de suministro es clave en la planeación estratégica y operativa de las empresas, dado que los inventarios son una parte integral de la mayoría de éstas, sobre todo en las que realizan la importación y comercialización. Con frecuencia, los inventarios son el concepto de mayor magnitud que aparece en el lado de los activos del balance general de una empresa y en muchas compañías representa un porcentaje muy alto del total de activos. Cuando se considera este hecho, resulta evidente que una reducción en los inventarios, aun en un porcentaje pequeño, puede representar ahorros sustanciales en términos de dinero.

Los fabricantes mantienen inventarios para satisfacer la demanda de los clientes, sin importar si el cliente es mayorista o detallista. Si se llevan inventarios a niveles elevados, entonces puede maximizarse el servicio a los clientes. Entonces, desde el punto de vista del cliente los inventarios elevados son deseables.

Es evidente que una empresa no puede reducir inventarios para disminuir la inversión en activos y, al mismo tiempo, mantener inventarios considerables para satisfacer la demanda de los clientes. Por lo tanto, se puede afirmar que la

empresa debe buscar alcanzar un equilibrio entre la satisfacción del cliente y las inversiones en activo a través de una buena gestión de inventarios apoyada con el control de la cadena de suministro.

7.1.1.1. Definición de inventario

El inventario es un recurso disponible que se encuentra almacenado en algún momento del tiempo. En el caso de una empresa de importación y comercialización, se puede decir que el inventario es un conjunto de artículos que están disponibles para la venta. Para una empresa productora o fabricante el inventario incluye materias primas, material de empaque, material en proceso y producto terminado.

Tradicionalmente se ha relacionado el inventario con empresas comerciales y empresas productoras, pero existen elementos como los equipos, los materiales y las personas que forman parte del inventario de otra clase de empresas como universidades, hospitales, escuelas y el mismo sector público que conforman el gobierno de un país. El concepto y los modelos de inventario pueden aplicarse también a esta clase de organizaciones de igual manera que a la empresa de comercialización o producción.

El estudio del inventario y el análisis de su comportamiento tienen una relación directa con el análisis de líneas de espera o teoría de colas, ya que no existe un solo modelo que aplique a todos los modelos de inventario. Es más, muchas veces aplican varios modelos según las características del problema.

7.1.1.2. Modelos comerciales y de producción

Un modelo comercial es aquel en el que el inventario de reabastecimiento es adquirido a un proveedor externo de la empresa y un modelo de producción

es aquel en el que el inventario de reabastecimiento se fabrica internamente, en la misma empresa.

La diferencia entre estos estriba en la forma en que se reabastecen los inventarios, los modelos de producción suponen que el reabastecimiento ocurre a lo largo del tiempo, de acuerdo con la capacidad instalada de la planta de producción, mientras que los modelos comerciales suponen que los inventarios se reabastecen de forma instantánea.

7.1.1.3. Modelos determinísticos y probabilísticos

Si se conoce la demanda, con certidumbre y se supone que es constante en el tiempo, entonces el modelo se denomina: modelo determinístico de demanda (Hillier & Lieberman, 2010). Ahora bien, el hecho de clasificar un modelo determinístico no implica que la demanda sea constante; existe la posibilidad que la demanda sea variable y aun así el modelo puede seguir siendo determinístico. Se considera un modelo probabilístico o estocástico cuando algunas de las relaciones funcionales, es decir, los parámetros del modelo, no se conocen con certidumbre. Un modelo probabilístico puede tener algunas relaciones funcionales que sean determinísticas y probabilísticas o todas pueden ser probabilísticas.

7.1.1.4. Tiempo de entrega (lead time)

El tiempo que transcurre entre el inicio del proceso de reabastecimiento (el pedido del cliente o la orden de producción) y la recepción o entrega del reabastecimiento de los inventarios se conoce como tiempo de entrega (*lead time*). En una empresa de importación y comercialización este tiempo, a veces muy difícil de conocer, suele tener una dependencia muy directa con el origen o procedencia del inventario.

7.2. Cadena de suministro

El análisis de la cadena de suministro comprende la planificación, ejecución y control de todas las actividades relacionadas con el flujo de productos de una empresa y debe ser considerado como una actividad preponderante en el quehacer diario.

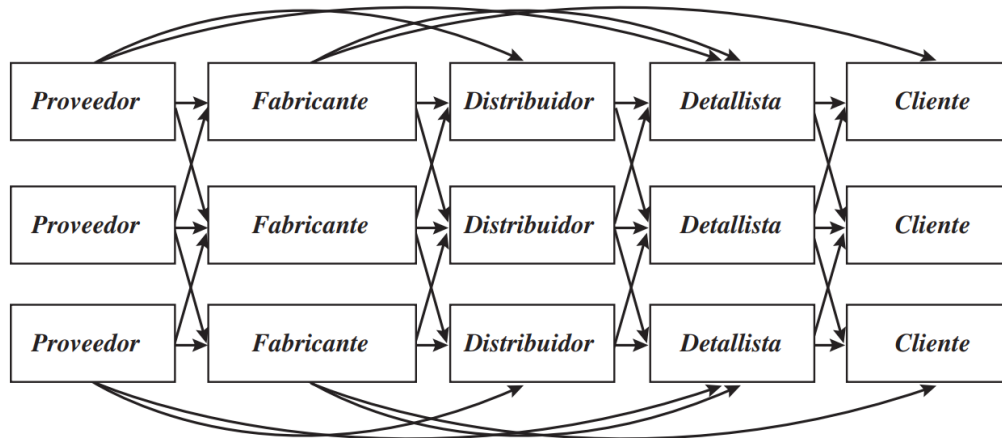
7.2.1. Definición de cadena de suministro

La cadena de suministro está formada por todos los elementos involucrados en el proceso, de forma directa o indirecta, que permiten satisfacer una necesidad establecida por el pedido de un cliente o la producción (Chopra & Meindl, 2008). Es importante mencionar que dentro de los elementos involucrados se encuentran: los transportistas, los encargados de almacenamiento, los agentes aduanales, los vendedores, los clientes y no solo el fabricante o proveedor.

7.2.2. Etapas de una cadena de suministro

Las cadenas de suministro pueden ser tan complejas o sencillas de acuerdo con el tipo de producto que se desea gestionar. En la figura 1 se puede observar que las distintas etapas o elementos de una cadena de suministro se asemejan a una red o malla de conexiones. Podemos usar entonces el término red de suministro, para definir sustancialmente la estructura de las cadenas de suministro. También, en la figura 1, se puede observarse que el flujo del proceso podría darse en distintas direcciones, tanto de forma horizontal como vertical, o incluso de izquierda a derecha y viceversa.

Figura 1. **Etapas de la cadena de suministro**



Fuente: Chopra, Sunil; Meindl, Peter (2008). *Administración de la Cadena de Suministro. Estrategia, planeación y operación.*

7.2.3. **Demanda y oferta**

El estudio de la demanda está basado en el mercado, que a su vez aglomera a los vendedores y compradores de un bien o servicio determinado. En tal sentido se puede afirmar que: la demanda de un bien o servicio determinado es la cantidad que de este bien o servicio están dispuestos a adquirir los compradores (Mankiw, 2012).

En términos del funcionamiento de los mercados, que poseen al precio del bien o servicio como uno de sus elementos más importantes, se puede afirmar, también, que ésta se rige por la ley de la demanda, la cual establece que la cantidad demandada de un bien o servicio disminuye, cuando el precio del bien o servicio aumenta.

Por el otro lado tenemos a la oferta, que es la cantidad de bienes o servicios que los vendedores quieren y pueden vender (Mankiw, 2012). Basado siempre

en el precio del mercado como un elemento vital e importante, se puede decir que la ley de la oferta establece que la cantidad ofrecida de un bien o servicio determinado aumenta, cuando el precio del bien o servicio aumenta.

7.2.4. Aranceles

El arancel es un impuesto a un bien o servicio que se vende dentro de un país proveniente de un lugar fuera del país, es decir, a la importación de bienes o servicios.

Esto hace que el monto del impuesto se agregue al precio de venta del bien o servicio dentro del país y obviamente dicho precio es mayor al precio del bien o servicio, del país de origen. Este impuesto varía según el origen del producto, el tipo de producto, el destino de uso del producto, los regímenes fiscales de las empresas que lo importan, los acuerdos internacionales y tratados de comercio regionales. En tal sentido, lo que para una empresa que importa el producto tiene un gravamen del 10 %, para otra tiene un gravamen del 0 % porque ésta puede estar exenta dicho impuesto.

Existen varios tipos de aranceles, los más utilizados son: arancel de valor agregado, arancel específico y arancel mixto, que es una combinación de los anteriores. Por otro lado, los países imponen aranceles a las importaciones por varias razones, y las principales son:

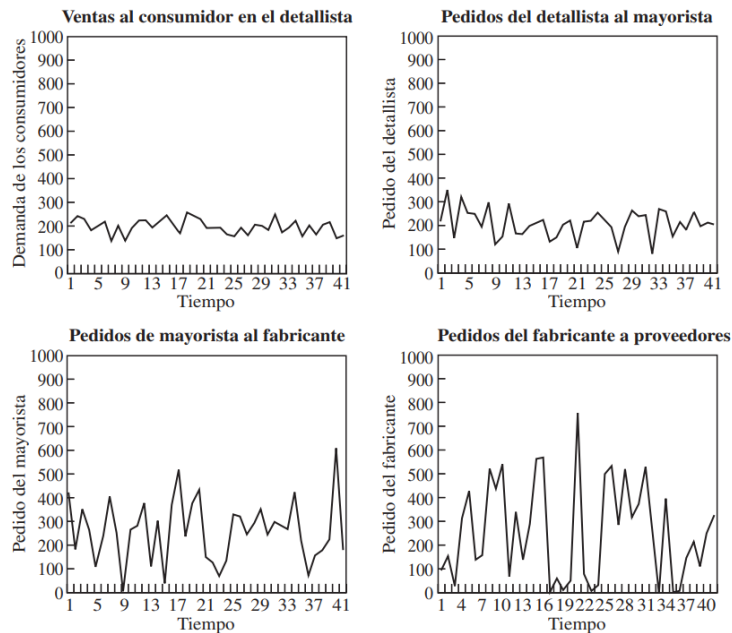
- Para obtener ingresos
 - Para la protección de la producción local sobre la competencia de la producción extranjera

7.2.5. Efecto látigo

Es una tendencia de cambios grandes en inventario en respuesta a cambios sustanciales y sorprendidos en la demanda del cliente. Este fenómeno se observa en canales de distribución y el reto fundamental radica en lograr la coordinación en la cadena de suministro a pesar de los diferentes y múltiples actores y la variedad de productos.

Este concepto fue nombrado así por el profesor Jay W. Forrester del MIT en su libro titulado Dinámica Industrial (1961), por lo que también se reconoce como el efecto Forrester. En las empresas se puede observar el efecto látigo, cuando las fluctuaciones en los pedidos aumentan a medida que avanzan en la cadena de suministro, de los detallistas a mayoristas, a los distribuidores, a los fabricantes y a su vez a los proveedores como se puede observar en la figura 2.

Figura 2. Efecto látigo



Fuente: Chopra, Sunil; Meindl, Peter (2008). *Administración de la Cadena de Suministro. Estrategia, planeación y operación.*

El efecto látigo tiende a reducir la rentabilidad de la cadena de suministro porque encarece el costo de ofrecer un nivel determinado de disponibilidad del producto. Por otro lado, este efecto tiende a aumentar el costo de fabricación y el costo de transporte entre otros.

7.3. Modelos de análisis predictivo

Los modelos de análisis predictivo son modelos matemáticos que predicen el comportamiento de una variable en función de un conjunto de otras variables. Una vez desarrollado el algoritmo de análisis predictivo, se requiere de datos históricos para la construcción del modelo.

7.3.1. Modelos predictivos

Un modelo predictivo permite representar la relación entre una variable de muestra y uno o más atributos que se conocen de dicha variable. Esta clasificación contiene a los modelos que podemos encontrar en áreas como mercadeo, inventario y ventas, en las cuales se requiere encontrar comportamientos o patrones ocultos para responder cuestionamientos sobre el desempeño de campañas, productos, ventas o clientes, según sea el análisis.

El análisis predictivo permite elaborar un modelo estadístico que utiliza los datos existentes para pronosticar datos futuros que se desconocen.

7.3.2. Modelos descriptivos

Un modelo descriptivo permite representar la cuantificación de las relaciones existentes entre los datos, a menudo es utilizado para clasificar proveedores, clientes o contactos en grupos. Estos modelos se enfocan en pronosticar el comportamiento de un cliente, producto o proveedor en particular, lo cual marca la diferencia con los modelos predictivos, y descubren e identifican

diferentes relaciones, por ejemplo: entre los productos y los clientes o entre los productos y los proveedores; los modelos descriptivos pueden ser utilizados para descubrir nuevos modelos adicionales, que podrían imitar una gran cantidad de casos individuales y hacer pronósticos. Dentro de estos modelos se pueden citar modelos de simulación, técnicas de previsión y teoría de colas.

7.3.3. Modelos de decisión

Un modelo de decisión permite representar la relación que existe entre los elementos de una decisión, esto incluye el conjunto de datos conocidos, la decisión como tal y el pronóstico resultante de la decisión tomada. Estos modelos permiten tomar en cuenta una cantidad considerable de variables, con el objeto de predecir los resultados de una decisión tomada. Dentro de estos modelos se pueden citar aquellos utilizados para optimizar ciertos resultados, cuando al mismo tiempo se reducen otros. Este es un escenario muy común en los negocios, cuando se desea conocer el efecto que podrían tener las nuevas políticas implementadas, sobre el resultado en un periodo de tiempo específico.

Cuando se toma una decisión y se crea el modelo de decisión correspondiente, este puede utilizarse una y otra vez, siempre y cuando este fue creado con un enfoque repetitivo. Dentro de estos modelos se pueden citar el análisis de Pareto, los árboles de decisión, el análisis de matriz de decisiones y el análisis de FODA (Fortalezas, Oportunidades, Debilidades y Amenazas).

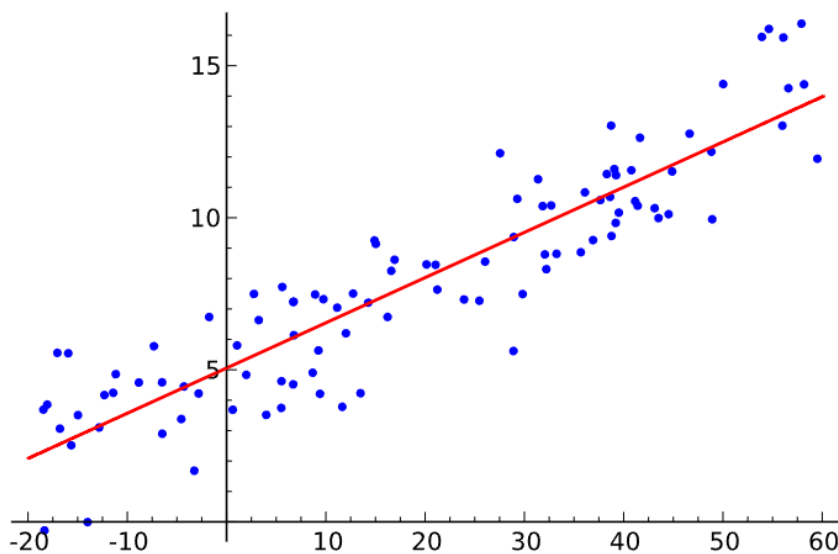
7.4. Técnicas de análisis predictivo

El análisis predictivo cuenta con varias técnicas y enfoques, que se dividen en dos grandes ramas: la primera rama se orienta a las técnicas de regresión y la segunda rama se enfoca en las técnicas de machine learning.

7.4.1. Regresión lineal

El modelo de regresión lineal analiza la relación existente entre la variable dependiente o de respuesta y un conjunto de variables independientes o predictoras. Esta relación se expresa como una ecuación que predice la variable de respuesta como una función lineal de los parámetros. Estos parámetros se ajustan para que la medida de ajuste sea óptima. El objetivo de la regresión es seleccionar los parámetros del modelo que minimizan la suma de los errores al cuadrado. Gran parte del esfuerzo en la adaptación del modelo se centra en minimizar el error, así como en asegurarse que está distribuido de forma aleatoria respecto a las predicciones del modelo. Esto se conoce como estimación de mínimos cuadrados ordinarios y los resultados en las mejores estimaciones lineales no sesgadas de los parámetros si y sólo si se satisfacen las suposiciones del teorema de Gauss-Markov.

Figura 3. Regresión lineal



Fuente: Wikipedia, the free encyclopedia (2021). Consultado el 20 de noviembre de 2021.

Recuperado de https://en.wikipedia.org/wiki/Linear_regression.

En la figura 3 se muestra un ejemplo regresión lineal simple. Este caso dispone de una variable dependiente y otra independiente, que proporcionan una función lineal que predice los valores de la variable dependiente según la función de la variable independiente. Se puede ver claramente como ciertos valores de la variable dependiente quedan por encima de la función lineal generada y otros por debajo, permitiendo al modelo predecir los valores de la variable dependiente.

Este modelo matemático se puede expresar así:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon$$

Donde:

Y = variable dependiente o variable de respuesta

$X_1 + X_2 \dots + X_n$ = variables independientes o explicativas

$\beta_0 + \beta_1 + \dots + \beta_n$ = parámetros del modelo, miden el efecto que las variables independientes tienen sobre la regresión

ε = variable aleatoria que recoge los factores de la realidad no controlados, errores o perturbación aleatoria.

La regresión lineal tiene varios modelos o análisis que se derivan de ella, tales como: la regresión lineal simple, en donde se realiza el análisis para dos variables estadísticas relacionadas por la ecuación. También existe la regresión lineal múltiple, en donde se realiza el análisis para dos o más variables estadísticas relacionadas a través de las ecuaciones.

Otro modelo derivado es el análisis de tiempo hasta el evento (análisis de supervivencia), en el cual suponemos que si los datos no fueron censurados serán representativos de la población de interés. Entonces, bajo esta situación podemos asumir que las observaciones censuradas surgen cuando la variable dependiente de interés representa el tiempo hasta un evento terminal, y la

duración del estudio es limitada al tiempo. Acá debemos darle suma importancia a la tasa de riesgo, definida como la probabilidad de que el evento ocurra en el tiempo t condicional a sobrevivir hasta el tiempo t y relacionado con la tasa de riesgo es la función de supervivencia que puede definirse como la probabilidad de sobrevivir al tiempo t .

Adicionalmente encontramos el modelo de curvas de regresión adaptativa multivariable, llamado MARS por sus siglas en inglés, la cual es una técnica no paramétrica que construye modelos flexibles al ajustar las regresiones por unidades. En este modelo aparece un concepto asociado con curvas de regresión que es el de un nudo. Podemos decir que un nudo es donde un modelo de regresión lineal da paso a otro modelo y por lo tanto es el punto de intersección entre dos curvas.

7.4.2. *Machine learning*

El concepto *machine learning* (aprendizaje automatizado por su traducción del idioma inglés) se utilizó originalmente para desarrollar técnicas que permitieran a las computadoras aprender. Este concepto fue evolucionando, de hecho, tiene su base en la matemática y la estadística (Burger, 2018) e incluye una serie de métodos estadísticos avanzados para la regresión y la clasificación. Tiene aplicación en una amplia variedad de campos, entre otros: diagnósticos médicos, reconocimiento facial, de imágenes y de voz, análisis del mercado de valores y detección de fraudes de tarjetas de crédito.

En ciertos casos es suficiente predecir directamente la variable dependiente sin tener en cuenta las relaciones subyacentes entre variables. En otros casos, las relaciones subyacentes pueden ser muy complejas y la forma matemática de las dependencias desconocida. Para estos casos, las técnicas de *machine learning* emulan la capacidad cognoscitiva del humano y aprenden de los ejemplos de entrenamiento para predecir eventos futuros.

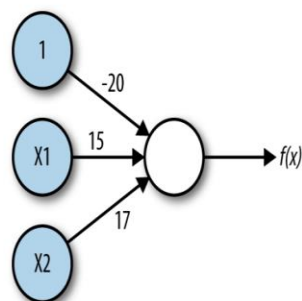
Existen dos grandes divisiones de *machine learning* como lo son los métodos supervisados y los no supervisados. Los métodos supervisados a su vez se subdividen en métodos de regresión, clasificación y mixtos. Dentro de los métodos supervisados, para el análisis predictivo, se encuentran las Redes neuronales, máquinas de vectores de soporte, clasificador Bayesiano y dentro de los métodos no supervisados se encuentra el análisis de K-vecinos más cercanos.

7.4.3. Redes neuronales

Las redes neuronales son técnicas de modelado no lineal sofisticadas que son capaces de modelar funciones complejas y toman su forma computacional de la forma en que funcionan las neuronas en un sistema biológico (Burger, 2018). Pueden aplicarse a problemas de predicción, clasificación o control en un amplio espectro de campos como las finanzas, la psicología cognitiva y neurociencia, la medicina, la ingeniería y la física.

Las redes neuronales se utilizan cuando no se conoce la naturaleza exacta de la relación entre los valores de entrada y de salida. Una característica clave de las redes neuronales es que aprenden la relación entre los valores de entrada y salida a través del entrenamiento.

Figura 4. Red neuronal



Fuente: Burger, Scott (2018). *Introduction to Machine Learning with R (Rigorous Mathematical Modeling)*.

En la figura 4 se muestra diagrama ejemplo del funcionamiento de una red neuronal emulando el comportamiento del operador de algebra booleana AND. Para calcular el resultado final de la puerta $f(x)$ (a la derecha de la figura) se necesita tomar los valores de entrada de las variables X_1 y X_2 , se tiene que definir también que si X_1 o X_2 es Verdadero entonces el valor es uno (1), en caso contrario el valor es cero (0). Luego construimos la ecuación utilizando esos pesos, para determinar el resultado. Después, se debe pasar el resultado por la función sigmoidea (el circulo vacío) y obtenemos el resultado de $f(x)$. En este contexto la función quedaría así:

$$h(x) = f(-20 + 15 * X_1 + 17 * X_2)$$

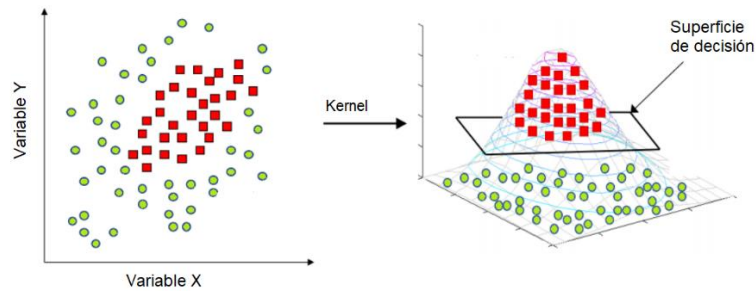
Como ejemplo para una combinación de valores, donde $X_1 = 1$ y $X_2 = 1$, tenemos lo siguiente:

$$h(x) = f(-20 + 15 + 17) \Rightarrow h(x) = f(12) \approx 1$$

7.4.3.1. Máquinas de vectores de soporte

Las máquinas de vectores de soporte (llamado SVM por sus siglas en inglés) se usan para detectar y explotar patrones complejos de datos agrupando, ordenando y clasificando los datos. Son máquinas de aprendizaje y algoritmos que se utilizan para realizar estimaciones de regresión y clasificaciones binarias (Burger, 2018). Usualmente usan métodos basados en kernel, que consiste en aplicar técnicas de clasificación lineal a problemas de clasificación no lineal. Hay una serie de tipos de SVM tales como lineal, polinomial y sigmoide.

Figura 5. **SVM solución lineal en el espacio**



Fuente: elaboración propia, realizado con Microsoft Power Point 365.

En la figura 5 se muestra la aplicación de un algoritmo kernel donde habitualmente no es necesario conocer explícitamente el mapeo, basta con conocer la función o kernel asociado, para realizar la transformación.

La SVM es el método kernel estándar de la clasificación y resuelve un problema de clasificación lineal en el espacio de características aplicando el principio de minimización del riesgo estructural (llamado SRM por sus siglas en inglés).

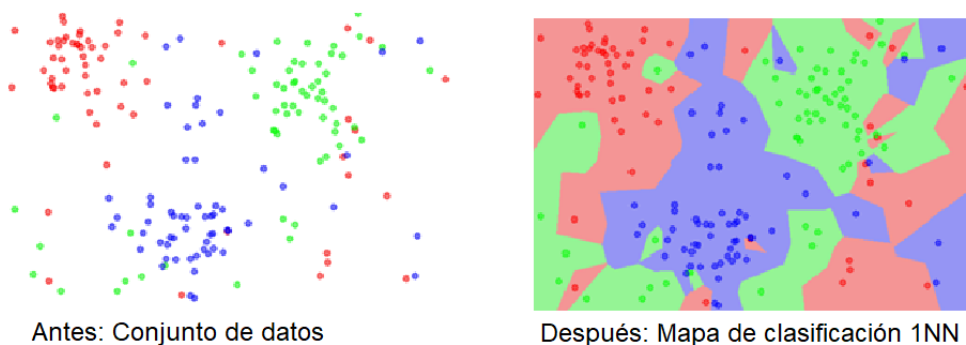
7.4.3.2. **K-vecinos más cercanos**

El algoritmo vecino más próximo k-NN (*k Nearest Neighbor*) pertenece a la clase de métodos estadísticos de reconocimiento de patrones. Se trata de un conjunto de entrenamiento con valores positivos y negativos. Una nueva muestra se clasifica calculando la distancia al vecino más cercano del conjunto de entrenamiento. El signo de ese punto determinará la clasificación de la muestra. En el clasificador k-vecino más cercano, se consideran los k puntos más cercanos y se utiliza el signo de la mayoría para clasificar la muestra. El rendimiento del algoritmo k-NN está influenciado por tres factores principales:

- La medida de distancia utilizada para localizar a los vecinos más cercanos.

- La regla de decisión usada para derivar una clasificación de los k-vecinos más cercanos.
- El número de vecinos utilizados para clasificar la nueva muestra.

Figura 6. **Aplicación k-vecinos más cercanos**



Fuente: Wikipedia, the free encyclopedia (2021). Consultado el 20 de noviembre de 2021.

Recuperado de https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm.

En resumen, se puede decir que a medida que el tamaño del conjunto de entrenamiento aumenta, si las observaciones son independientes e idénticamente distribuidas, independientemente de la distribución a partir de la cual se dibuja la muestra, ejemplificado por la figura 6, la clase predicha convergerá a la asignación de clase que minimiza el error de clasificación errónea.

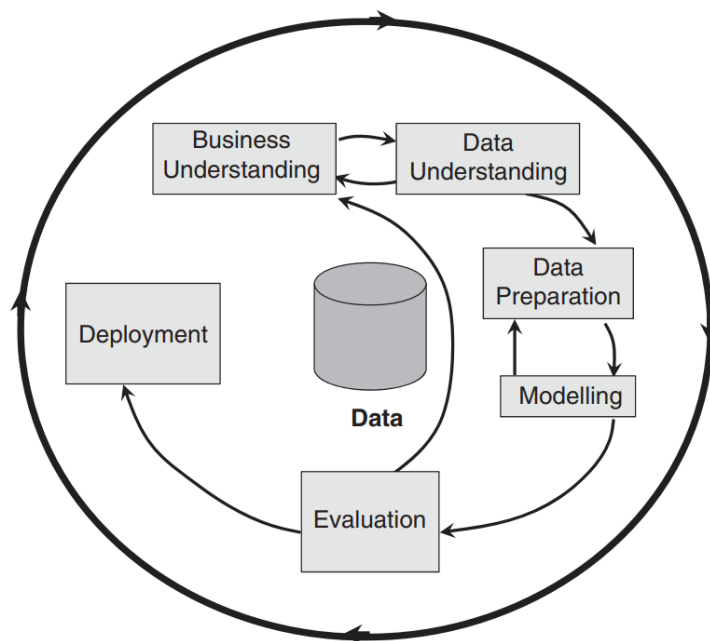
7.5. Minería de datos

Es el proceso de extracción de la información, de la cual se tiene interés, partiendo de los datos crudos, que permite descubrir nuevas e importantes relaciones, tendencias y patrones de comportamiento ocultos.

7.5.1. Fases del proceso de minería de datos

El proceso de minería de datos es un proceso de rango superior que pretende el descubrimiento del conocimiento. En este contexto existen varios modelos, pero uno de los más utilizados es el proceso estándar de la industria para la minería de datos (CRISP por sus siglas en inglés).

Figura 7. Diagrama CRISP-DM para minería de datos



Fuente: Leventhal, Barry (2010). *An introduction to data mining and other techniques for advanced analytics.*

En la figura 7 se muestra las fases establecidas por (Leventhal, 2010) como un proceso estándar de la industria, para la implementación de minería de datos y son los siguientes:

7.5.1.1. Comprensión del negocio

Fase en la cual se requiere tener el foco en la comprensión de los objetivos y requisitos de la industria, empresa o proyecto, para luego convertir este conocimiento en una definición de problema de minería de datos y un plan preliminar diseñado para lograr los objetivos.

7.5.1.2. Compresión de los datos

Esta fase inicia con la recopilación inicial de los datos para poder familiarizarse con los mismos e identificar posibles problemas de calidad de datos.

Esto permite obtener los primeros conocimientos sobre los datos y detectar posibles divisiones en subconjuntos interesantes. Todo esto sirve para después formular predicciones de información oculta.

7.5.1.3. Preparación de los datos

La fase de preparación de datos toma como base los datos iniciales que aún no han sido procesados, para crear un conjunto de datos finales. Existe la posibilidad que las tareas de preparación de datos se realicen varias veces, sin ningún orden en particular.

En su publicación (Leventhal, 2010) indica que dentro de estas tareas podemos incluir la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para herramientas de modelado.

7.5.1.4. Modelado

Con base en las técnicas de modelado conocidas o que se tienen a disposición, en esta fase se elige y aplica una para posteriormente ajustar o calibrar sus parámetros, lo que nos permite obtener los valores ideales. Cabe mencionar que como existe una diversidad de técnicas, las cuales tienen requisitos distintos entre sí, la escogencia de la técnica es vital para que no se tenga que regresar a la fase anterior las veces que el proceso lo requiera.

7.5.1.5. Evaluación

En esta fase del proceso, se puede contar con uno o más modelos que pueden reunir la calidad de datos deseada y antes de continuar con la implementación del modelo final, se recomienda evaluar exhaustivamente el modelo y repasar los pasos ejecutados, que permita asegurarse de lograr los objetivos de la industria, empresa o proyecto.

7.5.1.6. Implementación

Tomando en cuenta los objetivos o requisitos existentes, la fase de implementación puede variar desde generar un reporte simple hasta implementar un proceso de minería de datos repetible. Por otro lado, muchas veces esta fase ya no es implementada por el analista de datos, sino que muchas veces es el usuario que se sirve de los datos y realizada su análisis, para la toma de decisiones.

7.5.1.7. Tipos de información

El tipo de información que se obtiene de la minería de datos puede clasificarse de la forma siguiente: asociaciones, secuencias, clasificaciones, agrupamientos y previsiones.

El uso del tipo de información depende del problema que se pretenda resolver, como parte del presente estudio interesa particularmente los agrupamientos que es la existencia de similitud en los datos o posibles tendencias y las previsiones que nos brindan las predicciones estimadas con base a los patrones de comportamiento extraídos de los datos.

7.6. Lenguajes aplicados a *Machine Learning*

Existen varios lenguajes de programación orientados a *machine learning* entre los que se encuentran: Matlab, Julia, R y Python que serán abordados en esta investigación.

7.6.1. Lenguaje R

El lenguaje R es un lenguaje gratuito, de código abierto que tiene sus raíces en el mundo de la estadística y nació como una reimplementación del lenguaje S y su sucesor S+. Utilizado originalmente para la investigación científica, ahora es utilizado también en otras áreas como: el *machine learning*, minería de datos, investigaciones biomédicas, bioinformática y matemáticas financieras. Cuenta con una serie de bibliotecas que se pueden cargar con funcionalidades de cálculo, diferentes y útiles tipos de gráfica.

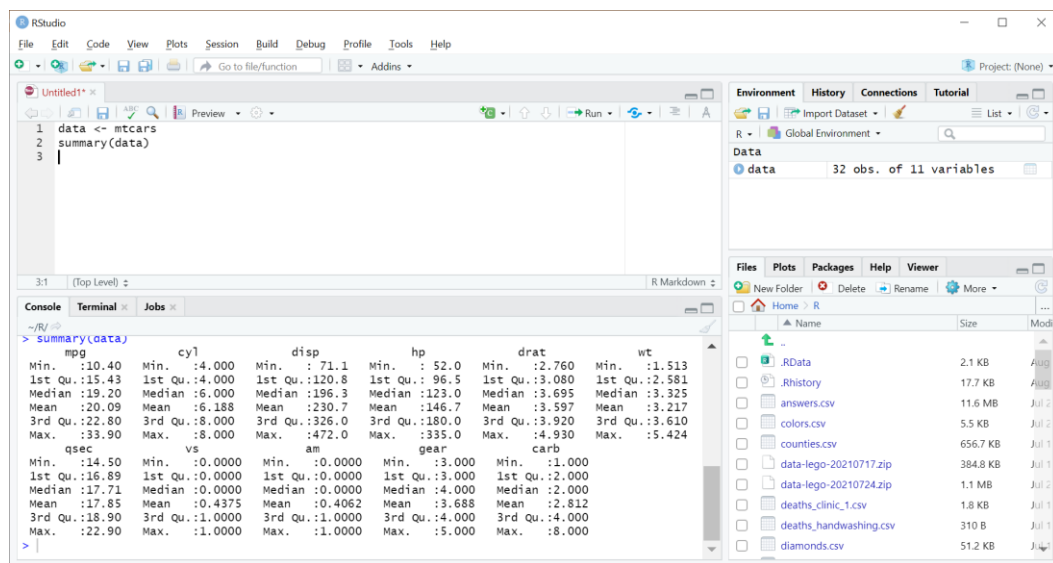
Existen varios ambientes soportados de desarrollo integrados (IDEs) para el lenguaje R, pero el más popular es R Studio (Burger, 2018).

Dentro de las principales características del lenguaje R se encuentran:

- Manejo y almacenamiento efectivo de los datos
- Un conjunto de operadores para la realización de cálculos con matrices
- Una gran colección de herramientas para el análisis de datos
- Utilidades gráficas para la visualización de datos

- Un lenguaje de programación bien desarrollado que incluye lógica condicional, ciclos de proceso, funciones recursivas, utilidades para la entrada y salida de datos.

Figura 8. **RStudio – Ambiente de desarrollo integrado**



Fuente: elaboración propia, realizado con captador de pantalla de Microsoft Windows 11.

Consultado el 20 de noviembre de 2021. Recuperado de <https://www.rstudio.com/products/rstudio/download/>.

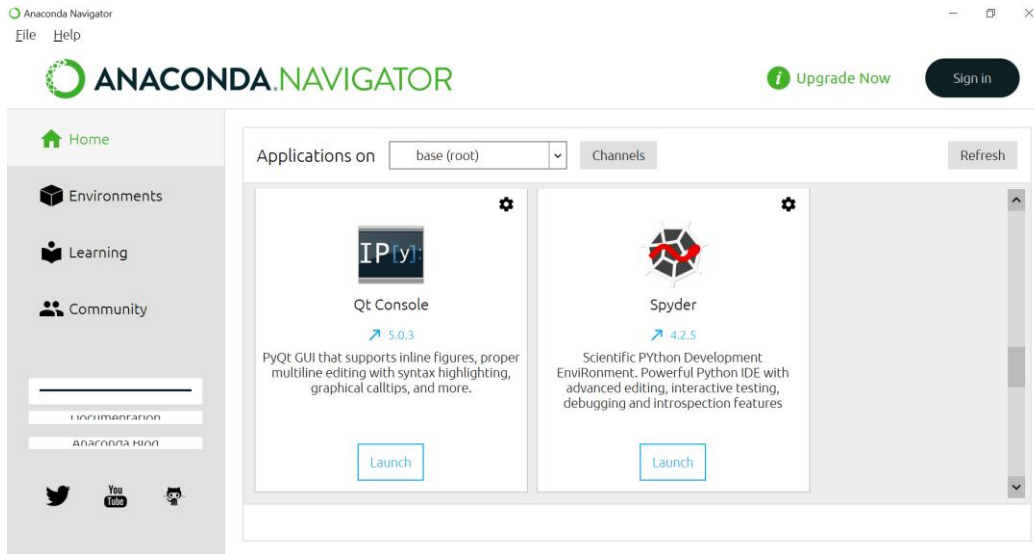
7.6.2. Lenguaje Python

El lenguaje Python es administrado por *Python Software Foundation*, que posee una licencia de código abierto. Es un lenguaje de programación interpretado, dinámico y multiplataforma, que soporta parcialmente la orientación a objetos, programación imperativa y, también, programación funcional.

La forma más sencilla para acceder a Python es utilizar la distribución Anaconda, la cual es una plataforma muy conocida y utilizada por científicos de datos de todo el mundo (Vandeput, 2021) en los campos de inteligencia artificial

y *machine learning*. Esta distribución, también instala los editores Spyder y Jupiter Notebook, que se pueden utilizar para ejecutar los scripts del lenguaje.

Figura 9. **Anaconda Navigator with Spyder editor**



Fuente: elaboración propia, realizado con capturador de pantalla de Microsoft Windows 11.

Consultado el 20 de noviembre de 2021. Recuperado de <https://www.anaconda.com/products/distribution#windows>.

8. PROPUESTA DE ÍNDICE DE CONTENIDOS

INDICE GENERAL

INDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTEAMIENTO DEL PROBLEMA

OBJETIVO

INTRODUCCIÓN

1. ANTECEDENTES

2. JUSTIFICACIÓN

3. ALCANCES

4. MARCO TEÓRICO

4.1. Modelos de inventario

4.1.1. Conceptos y términos

4.1.1.1. Definición de inventario

4.1.1.2. Modelos comerciales y de producción

4.1.1.3. Modelos determinísticos y probabilísticos

4.1.1.4. Tiempo de entrega (*lead time*)

4.2. Cadena de suministro

4.2.1. Definición de cadena de suministro

4.2.1.1. Etapas de una cadena de suministro

4.2.1.2. Demanda y oferta

- 4.2.1.3. Aranceles
 - 4.2.1.4. Efecto látigo
 - 4.3. Modelos de análisis predictivo
 - 4.3.1. Modelos predictivos
 - 4.3.2. Modelos descriptivos
 - 4.3.3. Modelos de decisión
 - 4.4. Técnicas de análisis predictivo
 - 4.4.1. Regresión lineal
 - 4.4.2. *Machine learning*
 - 4.4.3. Redes neuronales
 - 4.4.4. Máquinas de vectores de soporte
 - 4.4.5. K-vecinos más cercanos
 - 4.5. Minería de datos
 - 4.5.1. Fases del proceso de minería de datos
 - 4.5.1.1. Comprensión del negocio
 - 4.5.1.2. Comprensión de los datos
 - 4.5.1.3. Depuración de los datos
 - 4.5.1.4. Modelado
 - 4.5.1.5. Evaluación
 - 4.5.1.6. Implementación
 - 4.5.1.7. Tipos de información
 - 4.6. Lenguajes aplicados a *Machine Learning*
 - 4.6.1. Lenguaje R
 - 4.6.2. Lenguaje Python
- 5. CASO DE ESTUDIO
 - 5.1. Recolección de datos
 - 5.2. Revisión de datos
 - 5.3. Depuración de datos

6. PRESENTACIÓN DE RESULTADOS

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS

ANEXOS

9. METODOLOGÍA

En esta investigación se analizarán todos los productos que gestionan una empresa importadora y comercializadora. Como resultado de dicho análisis, se determinarán las cantidades mínimas de inventario y óptimas a comprar de cada producto.

9.1. Características del estudio

El enfoque del estudio propuesto es cuantitativo, porque por medio de la recolección de datos, tomando como base la medición numérica y el análisis estadístico, se puede descubrir patrones de comportamiento que nos permita alcanzar los objetivos trazados.

El alcance del estudio propuesto es descriptivo dado que pretende explorar y analizar el conjunto de datos, encontrar la correlación, tendencias o comportamientos ocultos, para interpretar y explicar los resultados obtenidos alineados al cumplimiento de los objetivos.

El diseño adoptado será cuantitativo y no experimental, pues la información sobre el mínimo de inventario y mínimo de compra se analizará en su estado original sin ninguna manipulación; además será transversal pues se estudiarán los eventos de compra y venta en un periodo de tiempo.

9.2. Unidad de análisis

La población en estudio será una muestra de empresas que pertenecen al segmento importación y comercialización, la cual se encuentra dividida en subpoblaciones como distribuidores mayoristas y ventas al detalle, cuyo

comportamiento, variables y condiciones varían significativamente entre sí. Por tal razón, se extraerá información del sistema informático actual de una empresa de distribución mayorista, tanto de las compras como las ventas por producto, en un período de tiempo específico que puede ser mayor o igual 1 año.

La información del año 2020 será excluida, porque representa información atípica y sesgada a causa de la pandemia Covid19.

9.3. Variables del estudio

Las variables establecidas para el estudio son el mínimo de inventario y el mínimo de compra, las cuales se describen a continuación:

Tabla I. Descripción de variables del estudio

Variable	Definición teórica	Definición operativa
Mínimo de inventario	Es la cantidad básica de unidades por producto, que por seguridad se debe disponer en almacén para satisfacer la demanda de los clientes.	Cantidad de inventario mínima por producto, disponible en cada ciclo de reabastecimiento.
Mínimo de compra	Es la cantidad mínima de unidades por producto que el proveedor está dispuesto a vender.	Cantidad de compra mínima por producto, en cada ciclo de reabastecimiento.

Fuente: elaboración propia, realizado con Microsoft Word 365.

9.4. Fases del Estudio

El estudio será dividido en fases y a su vez estas en actividades puntuales, que permitirán entender la situación actual y proponer opciones de solución con base en los hallazgos. Las fases del estudio se detallan a continuación:

9.4.1. Revisión documental

En esta fase se debe obtener el conocimiento teórico de los diferentes componentes que contempla la propuesta de solución. Será necesario investigar en diferentes fuentes de información como lo son: artículos científicos, libros, tesis de trabajos de postgrado y pregrado, revistas.

Este trabajo de investigación se basa en el desarrollo de un modelo de predicción de inventario utilizando para ello procesos de *machine learning* y minería de datos, por lo que es necesario establecer un conocimiento mediante revisión documental sobre los temas siguientes:

- Teoría de inventarios y cadena de suministro
- Procesos y algoritmos de optimización de inventarios
- Minería de datos orientada a la mejora de cadena de suministro
- *Machine learning* orientada a la mejora de cadena de suministro
- Herramienta y metodología de extracción de información

9.4.2. Casos de estudio

En esta fase se evaluarán distintas opciones de fuentes de información para realizar el estudio, que cumplan con las características mínimas de cantidad de productos, historia de la información, importación de productos y venta al mayoreo, categorización de productos.

9.4.3. Recolección de datos

En esta fase se elaborarán las sentencias en lenguaje SQL que se ejecutarán en la base de datos del sistema de información de los casos de estudio seleccionados, para obtener los datos que nos permitan desarrollar el modelo de predicción.

Las sentencias SQL son de elaboración propia y pretenden obtener los datos relevantes en las distintas fases en los que se requieran. Para ello será necesario realizar las actividades siguientes:

- Identificar período de análisis
- Diseñar método de extracción
- Elaborar herramienta de extracción
- Realizar proceso extracción

9.4.4. Revisión de datos

En esta fase se revisan y analizan los datos obtenidos, para identificar cuál de los métodos es factible utilizar que nos permita diseñar y elaborar el modelo de predicción.

Para ello será necesario realizar las actividades siguientes:

- Analizar conjunto de datos obtenidos
- Identificar el modelo a utilizar
- Diseñar modelo de predicción
- Elaborar modelo de predicción

9.4.5. Depuración de datos

En esta fase se analizan y depuran los datos como parte del proceso de minería de datos, en tal sentido será necesario realizar las actividades siguientes:

- Identificar si existen datos fuera de rango
- Eliminar datos fuera de rango que no afecten el estudio

9.4.6. Presentación de resultados

En esta etapa de resultados, se elaborarán las siguientes tablas para dar soporte a la presentación de resultados:

- Cuadro de datos obtenidos
- Cuadro de datos una vez se aplique el modelo de predicción

9.4.7. Redacción de informe final

En esta etapa se elabora el informe final, con los resultados de la investigación, a saber:

- Marco teórico
- Diseño de modelo de predicción
- Desarrollo de modelo de predicción
- Conclusiones
- Recomendaciones
- Entregar primera versión de trabajo de investigación

9.5. Técnicas de análisis de información

En el análisis de la información se pretende utilizar alguno de los métodos más utilizados en *machine learning* y minería de datos, dentro de los cuales están: regresión lineal para poder proyectar la demanda o la oferta según sea el caso, árboles de decisión para crear un mapa de opciones que permitan hacer predicciones, mejor llamados árboles de regresión porque nos permiten obtener valores. Con los resultados obtenidos se evaluará la utilización del algoritmo de validación cruzada de K iteraciones o aleatoria.

El ejemplo de la tabla II muestra la forma en que un algoritmo de *machine learning* aprende de la relación de la demanda entre los trimestres del Año1, para poder predecir el trimestre del Año2. Este proceso se hace con toda la información extraída por producto, para obtener mejores y certeros pronósticos.

Tabla II. **Venta histórica formateada para *machine learning***

	Entrada				Salida
	Año1				Año2
	Trimestre1	Trimestre2	Trimestre3	Trimestre4	Trimestre1
Producto A	4	1	5	3	2
Producto B	18	25	32	47	56
Producto C	7	2	3	1	1

Fuente: elaboración propia, realizado con Microsoft Word 365.

9.5.1. Preparación de datos

La primera actividad que se tiene que realizar en un proceso de *machine learning*, es formatear, limpiar y depurar los datos (Vandeput, 2021). En este estudio tendremos que formatear las venta y compra histórica por producto, para obtener una información como se muestra en la tabla II.

Por otro lado, durante el proceso de limpieza y formateo, se debe establecer que se utilice la notación estándar de la ciencia de datos, por ejemplo, la entrada se llamará: X y la salida se llamará: Y, puntualmente se llamará X_<dataset> al conjunto de datos que contendrá la información de entrada (histórica) y Y_<dataset> al conjunto de datos que contendrá la información de salida (pronostico).

Tabla III. **Conjunto de datos (*dataset*) históricos de venta**

Producto	Entrada								Salida			
	Año1				Año2				Año3			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
A	4	1	5	3	2	5	3	1	4	3	2	5
B	18	25	32	47	56	70	64	68	72	67	65	58
C	7	2	3	1	1	0	0	1	3	2	4	5
D	5	15	10	7	6	13	11	5	4	11	9	4

Fuente: elaboración propia, realizado con Microsoft Word 365.

En la tabla III se puede observar un ejemplo de un conjunto de datos típico de ventas históricas que sirven de base, para iniciar el estudio del pronóstico utilizando *machine learning*. Este conjunto de datos debe ser formateado de forma similar a la tabla I y se quiere predecir la venta de un producto (A, B, C o D) durante un trimestre en función de las observaciones de la venta de este producto durante los cuatro trimestres anteriores. Para ello se deben completar los conjuntos de datos (*dataset*) X_prueba y Y_prueba analizando los diferentes productos y cada vez se debe de crear una muestra de datos con cuatro trimestres consecutivos como X_prueba y el trimestre siguiente como Y_prueba. Es así como el algoritmo de machine learning aprenderá sobre la relación que existe entre un trimestre de la venta y los cuatro trimestres anteriores.

Tabla IV. **Conjunto de datos (*dataset*) para entrenamiento y prueba**

Iteración	Producto	Año1				Año2				Año3			
		T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
1	A	4	1	5	3	2							
1	B	18	25	32	47	56							
1	C	7	2	3	1	1							
2	A		1	5	3	2	5						
2	B		25	32	47	56	70						
2	C		2	3	1	1	0						
3	A			5	3	2	5	3					
3	B			32	47	56	70	64					
3	C			3	1	1	0	0					
...													
8	A								1	4	3	2	5
8	B								68	72	67	65	58
8	C								1	3	2	4	5

Fuente: elaboración propia, realizado con Microsoft Word 365.

En la tabla IV se puede observar un ejemplo del funcionamiento de las primeras iteraciones, en la iteración #1 se utilizan los datos desde la columna Año1 T1 hasta la columna Año1 T4 para predecir el Año2 T1, así sucesivamente para las siguientes iteraciones. El algoritmo de *machine learning* aprenderá las relaciones en X_{prueba} para predecir Y_{prueba} , lo cual tiene la siguiente notación: $X_{prueba} \rightarrow Y_{prueba}$.

9.5.2. Tamaño del conjunto de datos

Según (Vandeput, 2021) para la implementación de un algoritmo de *machine learning*: mientras más, mejor. Lo que significa que entre más datos podamos obtener de historia, para analizarla, mejor será el resultado y el pronóstico. Ahora bien, también se debe de tomar en cuenta la cantidad de períodos de historia, ya que entre más períodos utilicemos para hacer una predicción, menos podremos recorrer el conjunto de datos o sea se requieren más iteraciones.

En resumen, debemos encontrar el balance justo entre la cantidad de datos y el número de periodos del estudio. Una buena práctica es contar con un conjunto de datos mensuales para dos (2) años y así con ello tengamos dos (2) ciclos estacionales, para aprender de las relaciones y temporadas completas para probar.

9.5.3. Regresión lineal

La regresión lineal es una de las técnicas más utilizadas para realizar pronósticos en *machine learning*, que van desde una empresa que intenta pronosticar sus ganancias hasta las fronteras de la ciencia que intenta presentar nuevos descubrimientos que rigen las leyes del universo. Podemos utilizar regresión en cualquier escenario en el que necesite una predicción contra el tiempo. La elección entre un modelo de regresión logística, regresión lineal o regresión multivariante depende del problema y de los datos que se tengan.

Al ejecutar este algoritmo además de obtener la gráfica con los puntos dispersos y la línea ajustada resultante de la regresión, podemos obtener la fórmula utilizada para graficar, los residuos que representan la distancia vertical entre cada punto y la línea ajustada del modelo y los coeficientes que son los valores estimados para los coeficientes de la ecuación lineal.

Según (Burger, 2018) la regresión se presenta en dos formas: regresión lineal estándar, que pudimos conocer en algún curso de estadística básica o matemáticas, y regresión logística, que es totalmente diferente. La regresión logística es un método de clasificación que encuentra un límite que separa los datos en clases discretas. Para ello, pasa los datos a través de una función sigmoidea que asigna el valor real de los datos a un caso binario de 0 o 1.

9.5.4. Árboles de decisión

Los árboles de decisión son una clase de algoritmos de *machine learning* que crean un mapa, realmente crean un árbol de preguntas para hacer una predicción. Para hacer una predicción, el árbol comenzará en su base con una pregunta de sí o no y, dependiendo de la respuesta, continuará haciendo nuevas preguntas de sí o no hasta llegar a una predicción final.

Es importante tener en cuenta que se debe delimitar el crecimiento del árbol con la ayuda de algún criterio. Sin un criterio para detener el crecimiento del árbol, éste crecerá hasta que cada observación o muestra de datos tenga su propia hoja. (Vandeput, 2021) establece que dentro de estos criterios o parámetros debemos tomar en cuenta los siguientes: *max_depth* que es la cantidad máxima de preguntas consecutivas (nodos) que puede hacer el árbol, *min_samples_split* que es la cantidad mínima de observaciones que se requieren en un nodo para activar una nueva división del árbol y *min_samples_leaf* que es la cantidad mínima de observaciones que deben estar en una hoja. También indica (Vandeput, 2021) que este parámetro, cuanto más cerca de cero (0) mayor será el riesgo de sobreajuste, el árbol crecerá realmente hasta que haga suficientes preguntas para tratar cada observación por aparte.

9.5.5. Redes neuronales

Una red neuronal es un conjunto de ecuaciones que usamos para calcular un resultado. Dependiendo de la cantidad de características que tengamos en nuestros datos, la red neuronal casi se convierte en una caja negra. En principio, podemos mostrar las ecuaciones que componen una red neuronal, pero a medida que aumentan los niveles, la cantidad de información se vuelve demasiado engorrosa para intuir fácilmente. Las redes neuronales se utilizan ampliamente en la industria, en gran parte debido a su precisión, y vienen en muchos sabores

diferentes, pero las más populares provienen de redes neuronales de una o varias capas. A veces, es necesario negociar entre tener un modelo de alta precisión, pero velocidades de cálculo lentas. En tal sentido, es mejor probar varios modelos y usar redes neuronales solo si funcionan para su conjunto de datos en particular.

9.5.6. Algoritmos genéticos

Los algoritmos genéticos son un modelo matemático inspirado en la famosa idea de Charles Darwin sobre la selección natural. La selección natural preserva sólo a los individuos más aptos a lo largo de las diferentes generaciones. En *machine learning*, uno de los usos de los algoritmos genéticos es recoger el número correcto de variables para crear un modelo predictivo. Recoger el subconjunto correcto de variables es un problema de combinatoria y optimización. La ventaja de esta técnica, sobre otras, es que permite que la mejor solución surja de la mejor de las soluciones anteriores. Un algoritmo evolutivo que mejora la selección a lo largo del tiempo. La idea principal consiste en combinar las diferentes soluciones generación tras generación, para extraer los mejores genes (variables) de cada una de ellas. De esa manera crea nuevos y más adecuados individuos.

9.5.7. Series temporales

Una serie temporal es una sucesión de observaciones de una variable realizadas a intervalos regulares de tiempo. Según realicemos la medida de la variable considerada, podemos distinguir varios tipos de series temporales: discretas o continuas, de flujo o stock, de unidades de medida, de periodicidad de datos.

Para el análisis de series temporales es necesario tomar en cuenta los supuestos siguientes:

- Se considera que existe una cierta estabilidad en la estructura del fenómeno estudiado, para lograr esto se tiene que estudiar los períodos históricos más homogéneos posibles.
- Se debe mantener la definición y la medición de la magnitud objeto de estudio, dicho en otras palabras, los datos deben ser bastante homogéneos en el tiempo. Este supuesto no se da en muchas de las series económicas, ya que es frecuente que las estadísticas se perfeccionen con el paso del tiempo, produciéndose saltos en la serie debidos a un cambio en la medición de la magnitud estudiada.

Existen dos grandes grupos de clasificación de las técnicas de predicción basadas en series temporales:

- Métodos cualitativos, en donde el pasado no muestra información suficiente y directa sobre el estudio realizado.
- Métodos cuantitativos, en donde se extrae toda la información posible contenida en los datos, y con base al patrón de comportamiento en el pasado, es posible realizar predicciones sobre el futuro. A su vez los métodos de predicción cuantitativos se pueden segmentar en dos tipos de enfoque:
 - El análisis univariante, el cual se enfoca en realizar previsiones de valores futuros de una variable, utilizando como base la información existente en los valores pasados de la propia serie temporal.
 - El análisis multivariante, también llamado de tipo casual, ya que, en la explicación de la variable o variables del estudio, intervienen otras adicionales a ella o ellas mismas.

9.5.8. Análisis de la demanda

La información histórica de las empresas no es el único factor que activa las ventas o las compras, existen otros factores internos y externos que también impulsan la demanda. Los factores internos dependen de las decisiones de las empresas, entre ellos tenemos: promociones, cambios de precio, presupuesto de mercadeo, nuevos productos. Los factores externos no los puede controlar la empresa, son exógenos, entre ellos tenemos: el crecimiento de PIB, la tasa de desempleo, el clima, el tipo de cambio frente al dólar y pueden ser llamados también indicadores adelantados.

Cuando se utiliza un indicador adelantado, podemos decir que, en comparación con los modelos de regresión lineal tradicionales, la fortaleza de un modelo de *machine learning* es que no solo aprende de la relación entre un período y un indicador adelantado, sino que también, comprende automáticamente si existe una relación entre la demanda durante un período específico y cualquier versión rezagada de ese indicador. Un ejemplo de ello es que se podría observar un impacto en las ventas de productos en el cuarto trimestre del año, si el crecimiento del tipo de cambio fuera alto en el tercer trimestre de ese mismo año.

9.5.9. Análisis del nivel de inventario

En el proceso de gestión del inventario de una empresa, se puede contar con varios niveles de inventario que se deben conocer, para establecer un proceso integrado en toda la cadena de suministro. Entre estos se encuentran los siguientes:

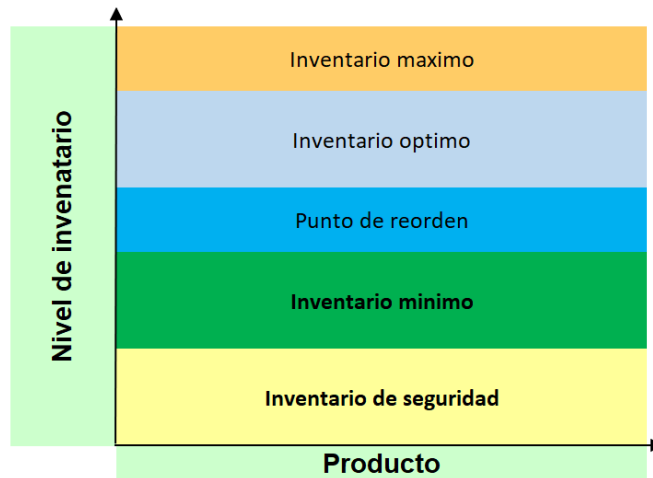
- Inventario mínimo: es la cantidad de producto que necesita el proceso para satisfacer la demanda, mientras se espera el arribo de los productos. El

nivel mínimo es el valor que permite seguir atendiendo a la demanda y supone el límite inferior de existencia dentro del cual no se debería bajar. Este valor debe ser establecido por cada producto que participa en el proceso.

- **Inventario máximo:** es la cantidad de producto máxima de existencia, regularmente se llega a este nivel cuando ingresa un nuevo pedido de compra al almacén. Este nivel está relacionado con el lote o cantidad de compra y al igual que el inventario mínimo, el inventario máximo se debe de establecer por cada producto que participa en el proceso.
- **Inventario de seguridad:** es la cantidad de producto que se establece para cubrir los incrementos no regulares o inesperados de la demanda y los retrasos en el envío de los proveedores. Este nivel de inventario depende de: tiempo de entrega (*lead time*), la efectividad del pronóstico de la demanda y del nivel de servicio esperado.
- **Punto de reorden:** es la cantidad de producto que exige la generación de un nuevo pedido a fin de satisfacer las necesidades de consumo o demanda.

Adicionalmente combinado con la determinación de los niveles de inventario por cada producto, es necesario conocer y determinar los ciclos de abastecimiento y el tiempo de entrega del proveedor, el cual podría ser normal o con demora.

Figura 10. **Niveles de inventario**



Fuente: elaboración propia, realizado con Microsoft Excel 365.

9.5.10. **Machine learning y lenguaje R**

En la actualidad, muchos sistemas de pronósticos del mundo real surgen del análisis de conjuntos de datos masivos (*big data*). La especialización en el análisis predictivo, así como los campos relacionados con *machine learning*, la ciencia de datos y el análisis de datos en general, están muy preocupados por la importancia del manejo de conjuntos de datos masivos, como ejemplos de esta realidad tenemos el motor de recomendación de productos de Amazon.com y el motor de recomendación de películas de Netflix.com.

La distribución base de R (lenguaje R) está diseñada para operar con datos que se ajustan a la memoria de la computadora. A menudo, los datos que queremos analizar son tan grandes que no es posible procesarlos todos en la

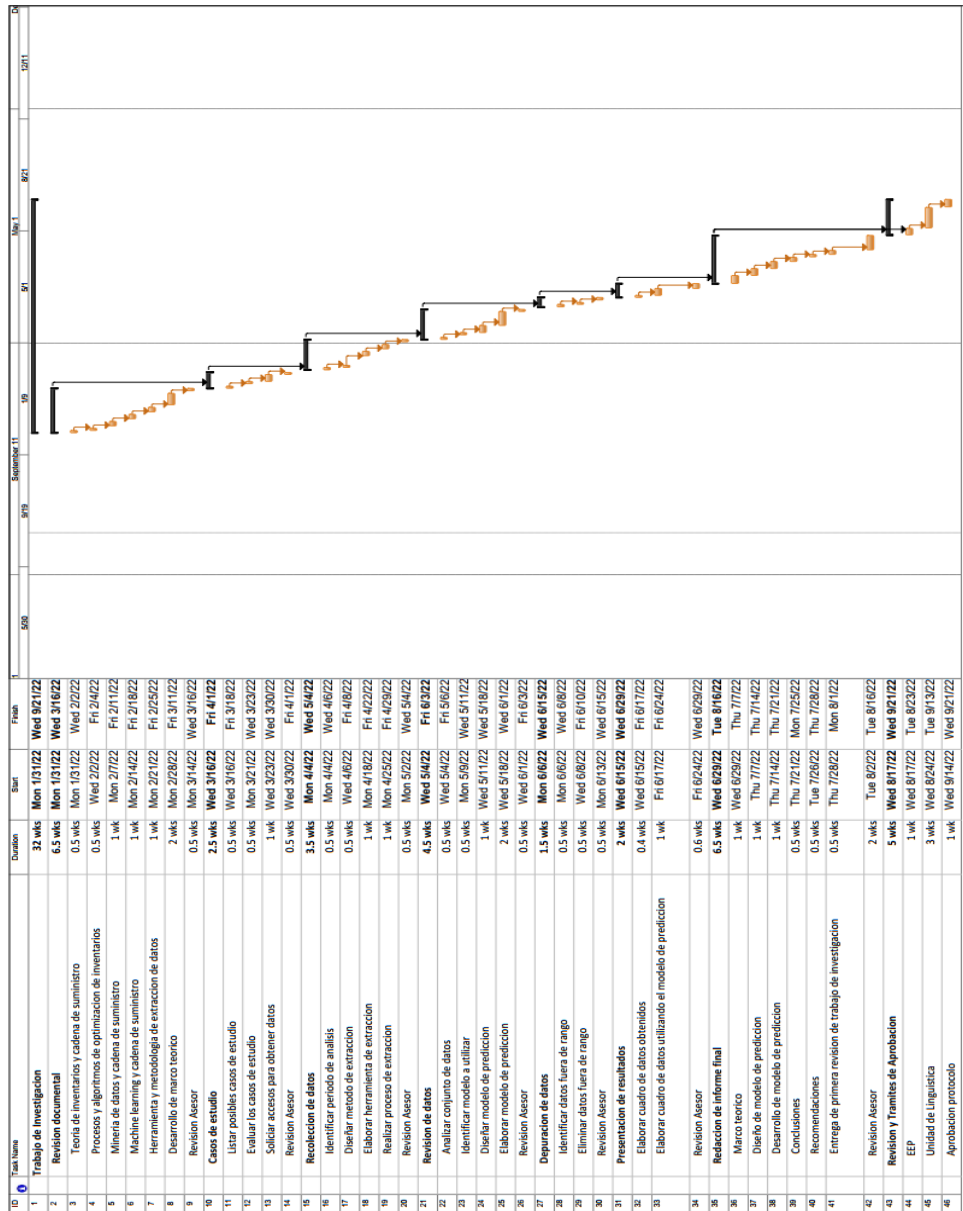
memoria de una sola computadora. Si este fuera el caso, es posible aprovechar los recursos informáticos bajo demanda en la nube y tener acceso a máquinas con características y capacidades mayores que las computadoras personales.

En un proyecto de programación en R, se puede contar con varios paquetes o librerías, para *machine learning* y computación de alto rendimiento que interactúan con bibliotecas de procesamiento paralelo. El catálogo de paquetes es muy amplio, se puede destacar, por ejemplo: RODBC, RpostgreSQL, SQLdf y RSQLite, que permiten cargar datos directamente de una base de datos, o ggplot2 y rgi que proporcionan la capacidad de generar gráficos en 2D y 3D. También está Shiny, que es capaz de generar gráficos interactivos que puede ser visualizados en la web mientras caret permite analizar la calidad de los datos, seleccionar características y construir modelos predictivos.

Además, es posible cargar conjuntos de datos en el entorno de trabajo desde múltiples formatos (archivos con extensiones como .csv, .xlsx, .txt, .json, .dbf o .xml) y realizar sobre ellos los pasos previos necesarios para la generación de modelos predictivos.

10. CRONOGRAMA

Figura 11. Cronograma de actividades



Fuente: elaboración propia, realizado con Microsoft Project

11. FACTIBILIDAD DEL ESTUDIO

Para el desarrollo y elaboración del presente trabajo de investigación, se cuenta con la posibilidad de extraer la información de un sistema de gestión empresarial (ERP) de una empresa de importación y comercialización, para poder realizar el análisis y la creación del modelo objeto de esta investigación.

11.1. Factibilidad operativa

A continuación, se describirán los elementos más importantes y necesarios para llevar a cabo la investigación:

- Recurso humano necesario
 - 1 persona del departamento de informática para las pruebas técnicas, quien debe contar con los accesos al sistema de gestión empresarial, para la ejecución de la herramienta de extracción de la información.
- Tiempo de pruebas de observación
 - El proceso de compra se realiza todos los meses con base en los acuerdos con los proveedores, las necesidades de los clientes y la disponibilidad de inventario. En tal sentido se realizará la extracción de la información de los últimos dos (2) periodos anuales de compras y ventas de los productos más importantes de la empresa, que se encuentran almacenados en el sistema de gestión empresarial.
 - El tiempo aproximado para realizar dichas pruebas es de 5 días hábiles.

11.2. Factibilidad técnica

Los recursos que son necesarios para efectuar las actividades y procesos que permitan llevar a cabo la investigación, se detallan a continuación:

- Aspectos técnicos de software y hardware
 - MS SQL Server Management Studio
 - RStudio con Lenguaje R
 - Estación de trabajo (PC o laptop)
- Procesador Intel Core i7 x64
- 16 GB RAM
- 512 GB libres en disco de estado solido
- Herramienta de extracción de la información
 - Script con sentencias SQL, queries, para obtener la información de la base de datos que utiliza el sistema de gestión empresarial.
 - Se utilizarán 2 horas en horario no laboral, durante 2 semanas, para la elaboración de los queries de extracción.

Basado en los aspectos descritos para la factibilidad técnica, se puede concluir que se cuenta con el software, hardware, tiempo y conocimiento necesario para la elaboración del proyecto.

11.3. Factibilidad económica

A continuación, se describen las actividades consideradas en la elaboración del presente trabajo de investigación con el costo y observación asociada.

Tabla V. **Actividades y costos de la investigación**

Actividad	Costo Unitario	Observación
Costo de infraestructura de trabajo (hardware y software)	Q 0,00	El equipo por utilizar es propiedad del investigador, no es necesario adquirir componentes adicionales.
Costo del software a utilizar en la infraestructura de trabajo	Q 0,00	El software que se utilizará es software libre que puede descargarse y distribuirse de forma gratuita, por lo que no se requiere ninguna inversión en licencias de uso.
Costo por personal en pruebas de observación y pruebas de tecnología	Q 0,00	El investigador aportara su tiempo al proyecto, para la realización de la extracción de datos del sistema de gestión empresarial. Únicamente se requiere acceso al servidor en donde se encuentra la base de datos.
Costo asociado por el desarrollo de herramienta de extracción	Q 0,00	El investigador aportara su tiempo al proyecto, para desarrollar la herramienta de extracción de datos de la base de datos del sistema de gestión empresarial.
Total Proyecto (+)	Q 0,00	
Costos asumidos investigador (-)	Q 0,00	
Costo de inversión	Q 0,00	

Fuente: elaboración propia, realizado con Microsoft Word 365

Finalmente, es posible afirmar que en la fase de factibilidad económica que no se incurre en ningún gasto ya que el costo asociado a la extracción de los datos, el desarrollo de la herramienta de extracción y creación del modelo de predicción es asumido por mi persona como parte interesada de llevar a cabo esta investigación.

Se concluye como factible el desarrollo e implementación de esta investigación debido al cumplimiento operativo, técnico y económico.

12. REFERENCIAS

1. Burger, S. (2018). *“Introduction to Machine Learning with R”* (Rigorous Mathematical Modeling), OREILLY.
2. Calatayud, A. y Katz, R. (2019). “Cadena de Suministro 4.0 (Mejores Prácticas Internacionales y Hoja de Ruta para América Latina)”, BID – Banco Interamericano de Desarrollo, recuperado de <https://publications.iadb.org/es/cadena-de-suministro-40-mejores-practicas-internacionales-y-hoja-de-ruta-para-america-latina>
3. Chamorro Corea, J., Díaz Camejo, J., Fuentes Espinoza, O. y Lovo Gutiérrez, H. (2018). “Política de inventarios máximos y mínimos en cadenas de suministro multinivel. Caso de estudio: Una empresa de distribución farmacéutica” (Nexo, revista científica). Universidad Nacional de Ingeniería. Facultad de Tecnología de la Industria. Managua, Nicaragua.
4. Chopra, S. y Meindl, Peter. (2008). “Administración de la Cadena de Suministro. Estrategia, planeación y operación” (3ª. Edición) Pearson Educación / Prentice Hall.
5. Dongo Pozo, A. y Silva Cama, X. (2020). “Análisis de la minería de datos aplicada en empresas del sector retail” (Tesis de pregrado). Universidad Católica San Pablo, Facultad de Ingeniería y Computación, Escuela Profesional de Ingeniería Industrial, Perú.

6. Gutierrez Alvarez, R. (2019). "Implementación de un modelo predictivo basado en Data Mining para la mejora de la gestión de ventas en la Distribuidora Jimenez Iriarte, S.A." (Tesis de pregrado). Universidad Nacional de Lima Sur, Facultad de Ingeniería y Gestion, Escuela Profesional de Ingeniería de Sistemas, Perú.
7. Hillier, F. y Lieberman, G. (2010). "Introducción a la Investigación de Operaciones" (9ª. Edición) McGraw-Hill
8. Jara-Cordero, S., Sanchez-Partida, D. y Martinez-Flores, J. (2017). "Análisis para la mejora en el manejo de inventarios de una comercializadora" (Revista de Ingeniería Industrial, septiembre 2017 Vol.1 No.1 1-18), Universidad Popular Autónoma del Estado de Puebla, México.
9. Lara Cabrera, M. (2020). "Desarrollo de un prototipo que permita realizar el análisis predictivo de delitos de investigación criminal en el departamento de Guatemala" (Tesis de maestría), Universidad de San Carlos de Guatemala, Facultad de Ingeniería, Escuela de Estudios de Postgrado, Guatemala.
10. Leventhal, Barry (2010). An introduction to data mining and other techniques for advanced analytics. Journal of Direct, Data and Digital Marketing Practice 12, 137-153
11. Mankiw, N. (2012). "Principios de Economía" (6ª. Edición) Cengage Learning

12. Piscoya Ordoñez, L (2016). “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular de la región de Lambayeque” (Tesis pregrado), Universidad Señor de Sipán, Facultad de Ingeniería, Arquitectura y Urbanismo, Escuela Académico Profesional de Ingeniería de Sistemas, Perú

13. Salas-Navarro, K., Miguél-Mejía, H. y Acevedo-Chedid, J. (2017). “Metodología de Gestión de Inventarios para determinar los niveles de integración y colaboración en una cadena de suministro” (Artículo). Universidad de la Costa en Barranquilla y de la Universidad Tecnológica de Bolívar, Colombia. *Ingeniare*, revista chilena de ingeniería, Chile.

14. Vandeput, N. (2021). “Data Science for Supply Chain Forecasting” (2nd Edition), DE GRUYTER, Foreword by Prof. Spyros Makridakis, Founder, Makridakis Open Forecasting Center.

