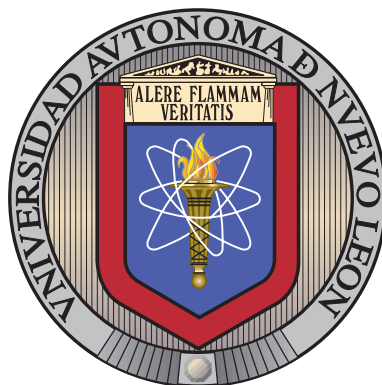


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO



Big data COMO ESTRATEGIA COMPETITIVA
APLICADO EN LA LOGÍSTICA Y EL TRANSPORTE
DE PYMES

POR

FÁTIMA SÁNCHEZ RUIZ

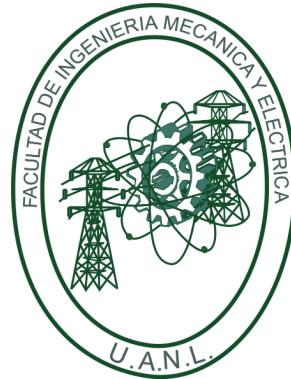
COMO REQUISITO PARCIAL PARA OBTENER EL GRADO DE
MAESTRÍA EN LOGÍSTICA Y CADENA DE SUMINISTRO.

AGOSTO 2019

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

SUBDIRECCIÓN DE ESTUDIOS DE POSGRADO



Big data COMO ESTRATEGIA COMPETITIVA
APLICADO EN LA LOGÍSTICA Y EL TRANSPORTE
DE PYMES

POR

FÁTIMA SÁNCHEZ RUIZ

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

AGOSTO 2019

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica
Subdirección de Estudios de Posgrado

Los miembros del Comité de Tesis recomendamos que la Tesis «*Big data* como estrategia competitiva aplicado en la logística y el transporte de pymes», realizada por el alumno Fátima Sánchez Ruiz, con número de matrícula 1937481, sea aceptada para su defensa como requisito parcial para obtener el grado de Maestría en Logística y Cadena de Suministro.

El Comité de Tesis

Dra. Jania Astrid Saucedo Martínez

Asesor

Dr. José Daniel Mosquera Artamonov

Revisor

Dr. Eduardo Valdés García

Revisor

Vo. Bo.

Dr. Simón Martínez Martínez

Subdirector de Estudios de Posgrado

San Nicolás de los Garza, Nuevo León, agosto 2019

*El tiempo es uno de los recursos más valiosos del hombre, cada día que pasa, no se recupera jamás, es imposible retroceder el tiempo, es por ello, que este trabajo de investigación, al igual que todas las horas que invertí durante el desarrollo del mismo, son para ti **Humberto**, mi pequeño hijo, cada día que dedique en la investigación, experimentación y documentación de esta tesis fue un día que no pude estar a tu lado, sin embargo, quiero que no te quede ni la menor duda de que **siempre estuviste presente en mi mente y en mi corazón.***

Sin ti mamá, tampoco hubiera podido culminar con este proyecto, siempre me diste todo tu apoyo desde el primer momento en que decidí dejar todo por alcanzar esta meta, con tu ejemplo me demostraste que con empeño y constancia es posible conseguir todo lo que uno se propone, eres el ejemplo perfecto de perseverancia, de la persona en que quiero convertirme algún día.

A mis hermanos, ya que siempre han sido mi apoyo y ejemplo.

Finalmente, a mi padre, al ser el mi mayor crítico, despertó en mi el deseo de superarme y demostrar que siempre se puede ser mejor en cada aspecto de la vida y no parar hasta concretar mis metas.

ÍNDICE GENERAL

Agradecimientos	XI
Resumen	XIII
1. Planteamiento y fundamentación	1
1.1. Introducción	1
1.2. Descripción del problema	4
1.3. Objetivo	5
1.4. Hipótesis	5
1.5. Justificación	6
1.6. Metodología	9
1.7. Estructura de la tesis	10
2. Antecedentes	12
2.1. Cadena de suministro y estrategias competitivas	13
2.1.1. Tecnologías de la información e Industria 4.0 en la Cadena de suministro	18

2.1.2.	Aplicaciones <i>Big Data</i> en la logística y transporte	24
2.1.3.	Técnicas de análisis <i>Big Data</i>	30
2.1.4.	Herramientas de análisis <i>Big Data</i>	38
2.2.	Análisis de series de tiempo en el pronóstico de la demanda	41
2.2.1.	Series de tiempo	42
2.2.2.	Series de tiempo intermitentes	46
2.2.3.	Métodos de pronóstico en series de tiempo	48
2.2.4.	Métodos de suavización	48
2.2.5.	Promedios móviles	49
2.2.6.	Métodos auto regresivos	50
2.2.7.	Métodos de pronostico en series de tiempo intermitentes	51
2.2.8.	Evaluación de los métodos de pronóstico	54
3.	Metodología	57
3.1.	Desarrollo de la herramienta	57
3.1.1.	Pseudocódigos de la herramienta	59
4.	Caso de Estudio	66
4.1.	Información de la empresa	66
4.2.	Descripción del proceso de atención	68
4.3.	Exploración de la problemática y áreas de oportunidad	69

4.4. Pautas para el desarrollo de la metodología y la aplicación de la herramienta en la empresa	71
4.5. Formulación de los criterios de decisión	72
4.6. Análisis de resultados	74
5. Conclusiones	80
5.1. Conclusiones generales	80
5.2. Recomendaciones	82
5.3. Contribuciones	82
6. Anexos	84
6.0.1. Pseudocódigo de limpieza:	87
6.0.2. Pseudocódigo para realizar la extracción de datos	88
6.0.3. Pseudocódigo para descomponer las fechas:	89
6.0.4. Pseudocódigo para extraer la fecha:	90
6.0.5. Pseudocódigo para aplicar método de pronóstico suavización exponencial simple:	91
6.0.6. Pseudocódigo para aplicar método de pronóstico Croston: . . .	93

ÍNDICE DE FIGURAS

1.1. Metodología de la investigación	11
2.1. Cadena de suministro tradicional	14
2.2. Cadena de suministro en entorno 4.0	23
2.3. Fases <i>Big Data</i>	25
2.4. Áreas de aplicación <i>Big Data</i>	26
2.5. Ventajas análisis Big Data	29
2.6. Técnicas de análisis <i>Data Mining</i>	35
2.7. Funciones y algoritmos de técnicas de análisis de datos <i>Data Mining</i> .	37
2.8. Componentes de las series de tiempo	44
2.9. Serie de tiempo intermitente	46
2.10. Clasificación de demanda intermitente	47
2.11. Métodos de solución y análisis en las series de tiempo	48
2.12. Diagrama de flujo método Croston	53
3.1. Proceso KDD	58

4.1. Ubicación geográfica de la empresa	67
4.2. Diagrama de Flujo	70
4.3. Base de datos ejemplo	73
4.4. Gráfico de la demanda del producto 1	75
4.5. Clasificación de intermitencia	78
6.1. Gráfico de la demanda del producto 2.	85
6.2. Gráfico de la demanda del producto 3	86
6.3. Gráfico de la demanda del producto 4	87

ÍNDICE DE TABLAS

2.1. Herramientas de analisis de datos de código abierto	39
4.1. Datos del producto 1.	75
6.1. Datos del producto 2.	84
6.2. Datos del producto 3.	85
6.3. Datos del producto 4.	86

AGRADECIMIENTOS

A Dios, por permitirme cumplir esta meta y poner a personas maravillosas en mi camino, las cuales me apoyaron y estuvieron siempre a mi lado en los momentos mas difíciles.

A mi tutora la Dra. Jania Astrid Saucedo Martinez, por “adoptarme” cuando me quedé sin tutor, por todo el tiempo que dedicó para ayudarme a desarrollar mi proyecto y el apoyo durante estos dos años en el ámbito académico.

A mis revisores el Dr. José Daniel Mosquera Artamonov y el Dr. Eduardo Valdéz García, por todos los consejos, por su ayuda, por abrirme las puertas de “yalma” e introducirme al mundo del *Data Science*, sobre todo al Dr. Daniel por su enorme paciencia, por estar siempre detrás de mí para ayudarme pero sobretodo para enseñarme y no dejarme dar por vencida en los momentos de estrés, por sus consejos y toda la retroalimentacion.

A mis amigos, Allan Cubillo y Yang-Li, más que amigos se convirtieron en mi familia, esos juevesitos salvaron mi cordura, ¡pura vida maes!

A la Universidad Autónoma de Nuevo León y a Facultad de Ingeniería Mecánica y Eléctrica por apoyarme con las becas de rectoráa, inscripción y pago de materias respectivamente.

A CONACyT por el apoyo económico para poder desarrollarme como estudiante de tiempo completo.

A la empresa caso de estudio ya que en ella aprendí demasiado, fue mi puerta al mundo laboral ya que me dieron la oportunidad de llevar a cabo grandes proyectos, les debo mucho de lo que soy actualmente como profesionista.

A toda mi familia, por animarme y estar siempre pendiente de mí y de mi hijo.

Finalmente pero no menos importante, Cristian gracias por estar siempre a mi lado, darme ánimo siempre que lo necesite, pero sobre todo por enseñarme que el amor mas importante, es el amor propio.

RESUMEN

Fátima Sánchez Ruiz.

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

Universidad Autónoma de Nuevo León.

Facultad de Ingeniería Mecánica y Eléctrica.

Título del estudio: *Big data* COMO ESTRATEGIA COMPETITIVA APLICADO EN LA LOGÍSTICA Y EL TRANSPORTE DE PYMES.

Número de páginas: 102.

OBJETIVOS Y MÉTODO DE ESTUDIO: El principal objetivo de este proyecto fue acercar a las pymes al mundo del *Data Science* y brindarles una herramienta que les permitiera realizar un análisis completo de manera intuitiva a través de la cual analizar la información, mejorar su procesos de abastecimiento así como generar el pronóstico de su demanda, con lo cual replantear su estrategia competitiva y convertirlos en impulsores de datos.

La metodología propuesta contribuye al desarrollo de proyectos de minería de datos y en nuestro caso específico presenta los pasos a seguir para el desarrollo de una herramienta a través de la cual es posible analizar los datos y generar pronósticos de los servicios y de las refacciones requeridas para la realización de los mismos, que al ser implementada podrá mejorar el desempeño logístico de la organización.

CONTRIBUCIONES Y CONCLUSIONES: El presente trabajo representa un apoyo para la mejora en la toma de decisiones en la organización. Las empresas requieren hacer hincapié en el fortalecimiento y construcción del análisis de información y en la toma de decisiones con base en los datos.

En este caso se buscó que el análisis de la información tuviera impacto en las operaciones logísticas de la empresa para ser exactos en el procedimiento de abastecimiento de refacciones, almacén y compras para mejorar el tiempo de respuesta y entrega de unidades al cliente ya que como se comentó en el capítulo cuatro la empresa presentaba reportes de quejas por parte de los clientes en cuanto a los tiempos de entrega de unidades.

Con la reducción en los tiempos de servicio y entrega de las unidades derivadas de un eficiente abastecimiento la empresa logrará reducir las quejas por incumplimiento, brindándole a su vez la posibilidad de crear mejores acuerdos comerciales con los proveedores que le permitan reducir los costos al comprar en volumen, mantener e incrementar su cartera de clientes y al reducir los costos por abastecimiento poder invertir en nuevos proyectos o ampliaciones en la nave lo que representaría un aumento en su capacidad instalada.

Firma del asesor: _____
Dra. Jania Astrid Saucedo Martínez

CAPÍTULO 1

PLANTEAMIENTO Y FUNDAMENTACIÓN

1.1 INTRODUCCIÓN

En los últimos años la logística y el transporte se han convertido en actividades clave de la Cadena de Suministro, gestionan todas las actividades referentes al movimiento, transporte y flujo de materia prima, productos, servicios e información, convirtiendo a este último recurso la clave para garantizar un correcto servicio y comunicación a través de toda la cadena de valor.

Actualmente la información generada a través de los sistemas que conectan los procesos de la organización, puede ser analizada, estudiada y procesada para contribuir a mejorar la toma de decisiones, lo que le permite a las organizaciones, anticiparse a las necesidades de los clientes y el mercado al ritmo que estos lo demandan.

McAfee *et al.* (2012) exponen algunos ejemplos en los que mediante un correcto análisis de datos, es posible adquirir un mayor conocimiento del cliente así como generar cambios importantes en beneficio de la organización.

Por ejemplo, en el caso de las librerías tradicionales que solían contar únicamente con tiendas físicas, se dificultaba realizar el rastreo de las ventas con claridad,

una de las soluciones implementadas, fue establecer un sencillo programa de lealtad, con el cual pudieron vincular las compras realizadas por cada uno de sus clientes, tiempo después al recolectar la información, lograron visualizar el comportamiento de las ventas individuales y conocer los libros y temas que tenían mayor preferencia entre sus clientes.

Si bien este fue un gran paso en el camino a la comprensión del cliente en este sector, se queda muy por detrás de los resultados obtenidos una vez que las librerías entraron al mercado de las ventas en línea, la comprensión de los clientes aumentó dramáticamente ya que los minoristas podían rastrear no solo lo que los clientes compraban, sino también lo que miraban y cómo navegaban por el sitio; en poco tiempo, desarrollaron algoritmos para predecir qué libros les gustaría leer a cada cliente de manera individual.

Con el paso del tiempo los algoritmos comenzaron a funcionar mejor cada vez que el cliente respondía o ignoraba una recomendación.

Otro ejemplo de las ventajas que brinda el realizar prácticas de análisis de datos, se dio al conseguir la visualización y claridad de la efectividad en la relación comercial minorista/proveedor, debido a que el análisis predictivo no solo es capaz de analizar todo el proceso de la cadena de suministro minorista, sino que también puede garantizar el grado de cumplimiento del proveedor esto de acuerdo con Bentley *et al.* (2016).

La tecnología está transformando no solo la forma en que realizamos actividades cotidianas, sino también esta transformado la manera en que hacemos negocios, convirtiéndose en un factor clave para reducir costos, mejorar la experiencia del cliente y obtener conocimiento del negocio y el mercado.

A través del tiempo se han venido dado una serie de cambios, se ha transformado la forma en la que se conforman las empresas y se construyen y establecen los procesos, en este sentido entra la llamada cuarta revolución Industrial, cuyo principal objetivo es lograr la conectividad total y abrir paso a la generación de las fabricas

inteligentes.

La también denominada Industria 4.0, ha traído consigo una serie de tecnologías “prominentes” con diferentes características y aplicaciones, entre estas tecnologías se encuentra: *Big Data*.

Lhor (2012) definió el término *Big Data* como: “una nueva tendencia tecnológica que abre la puerta a una nueva manera de entender el mundo y tomar decisiones”.

Aun cuando desde hace ya muchos años se ha estudiado el comportamiento de los datos, en la última década el campo del *Data science* a despertado mayor interés el las organizaciones debido a los resultados obtenidos en diversos campos y giros empresariales.

Uno de los campos de mayor impacto se centra en la toma eficiente de decisiones en los negocios, que derivan y contribuyen a la creación de estrategias para mejorar el rendimiento de los diferentes actores en la cadena de suministro.

A pesar de encontrarnos en una época en la que esta en vogue el desarrollo tecnológico y digital, muchas empresas continúan realizando la toma de decisiones únicamente con base al “*Know How*” sin datos duros que los respalden.

Es de suma importancia que las empresas enfoquen sus recursos en obtener no solo una mayor visibilidad en las redes logísticas, si no en la administración y análisis de la información generada a través de estas actividades.

La correcta evaluación, visualización y determinación del comportamiento de los datos pueden contribuir a que las organizaciones logren anticiparse a las fluctuaciones y cambios inherentes del mercado, detectar problemáticas que puedan presentarse en el mediano y corto plazo y desarrollar ideas o estrategias a través de las cuales puedan aumentar su resiliencia así como desarrollar una ventaja competitiva, esto sin necesidad de realizar grandes inversiones económicas en sus sistemas de gestión de la información.

Acciones sumamente sencillas como la que proponemos en esta investigación lograr crear un gran impacto e las empresas en áreas como la logística, almacenes, transporte, que impactan de manera directa en la satisfacción del cliente, tiempo de respuesta y los costos.

1.2 DESCRIPCIÓN DEL PROBLEMA

Entre los principales retos a los que se enfrentan las empresas independientemente del giro o industria al cual pertenezcan, se encuentra el tratar de adaptarse a los cambios tecnológicos y tendencias del mercado, el implementar las nuevas herramientas, tecnologías y técnicas en sus operaciones diarias supone un gran reto.

Esta necesidad se acrecienta en las pequeñas y medianas empresas, que por sus características son las más vulnerables ante los cambios tecnológicos.

La mayoría de las pymes, no cuenta con el capital necesario para invertir en equipo, softwares y personal capacitado, carece de sistemas de administración por lo que la información generada a través de sus operaciones comúnmente no es registrada.

Si las empresas no cuentan con registros o bases de información por más sencillas que éstas sean, se convierte en una tarea casi imposible el realizar un análisis y medición del estado de sus operaciones así como del rendimiento de las mismas, perdiendo la oportunidad de mejorar sus procesos.

Otro de los principales problemas para este tipo de empresas surgen al tratar de construir o predecir su demanda, ya que al no contar con registros de sus ventas se vuelve imposible pronosticar su demanda.

Una de las principales causas o factores que intervienen en la poca practica de análisis de información, es la falta de acceso a los datos y herramientas que permitan y faciliten su interpretación.

Realizar de manera constante un análisis y consulta de información, así como *dashboards* mediante los cuales se pueda interpretar la información, puede contribuir a la mejora en la toma de decisiones y la creación de estrategias que permitan minimizar el grado de error e incertidumbre no solo en el pronóstico de la demanda si no en diversas áreas que impactan las operaciones en las organizaciones.

1.3 OBJETIVO

Generar ventaja competitiva de una cadena de suministro de una pyme dedicada al servicio de mantenimiento de transporte pesado, a través del uso e implementación de una herramienta basada en técnicas *Big Data*, como estrategia en las operaciones logísticas, mediante la cual puedan analizar, procesar la información y crear estrategias que les permitan no solo convertirse en impulsores de datos, sino también aumentar la capacidad de adaptarse a las necesidades de los clientes y el mercado al ritmo que lo demandan.

1.4 HIPÓTESIS

Mediante el desarrollo e implementación de una herramienta basada en técnicas de análisis *Big Data*, en el procesamiento, análisis y estudio del comportamiento de datos de los sistemas logísticos y de transporte, se podrá optimizar la toma de decisiones y eficientar los procesos logísticos brindándole a la empresa la capacidad de responder de manera ágil ante las necesidades y cambios en el comportamiento del mercado.

1.5 JUSTIFICACIÓN

En México; las micro, pequeñas y medianas empresas, son la columna vertebral de la economía, generan aproximadamente el 42 % del producto interno bruto, y el 78 % del empleo directo convirtiéndose en la fuente de empleo más importante del país (Arana, 2018).

Por otra parte, García Canseco (2015) menciona que entre las 5 principales causas del fracaso de una pyme, se encuentran:

1. La falta de ingresos para subsistir,
2. la falta de indicadores,
3. la falta de proceso de análisis de información,
4. la falta de planeación,
5. y una administración deficiente.

En esta investigación, trataremos de impactar a por lo menos dos de estos 5 factores, tratando de minimizar la falta de análisis de información y la planeación deficiente.

CEPAL *et al.* (2005) menciona que las pymes se manejan de una manera errática, puesto que la falta de control en las operaciones y la pobre estimación de la demanda eleva los costos de producción, además en su mayoría muchos de los errores de planeación son causados por la poca o nula gestión y análisis de la información interna y externa de las empresas a la hora de tomar decisiones.

En otro estudio realizado por la consultora Inc. (2016) “Logistic Cost And Service” resaltan que el transporte llega a representar cerca del 49 % de los costos logísticos de una empresa, y en las pymes estos pueden llegar a aumentar hasta

un 25 % más, por lo que, es de vital importancia desarrollar e implementar herramientas y técnicas que permitan mejorar estos procesos mediante estrategias con las cuales optimizar las operaciones, disminuir costos, elevar la eficiencia y mejorar la experiencia del cliente.

En el este mismo estudio desarrollado por la consultora internacional mencionan que otro gran tanto de los costos corresponde a las actividades relacionadas al almacenaje de mercancías y materias primas.

Las pymes enfrentan el difícil desafío de adaptarse a la era digital con una disponibilidad de capital típicamente baja para invertir en tecnología y, en algunos casos, también la falta de capital humano que cuente con las habilidades necesarias para emplear las técnicas, herramientas o softwares de análisis de información atípicos.

De acuerdo con las últimas cifras oficiales reportadas por la SCT (2017), la participación del sector de Transporte y Logística representó el 4.5 % del PIB durante el último trimestre del 2015. En este mismo periodo, el PIB mostró un aumento de 3.0 % respecto al mismo trimestre del año anterior, cifra superior al crecimiento de la economía mexicana actual que es cerca del 2.5 %, una adecuada administración de las actividades logísticas puede convertirse en la clave para mantener una ventaja competitiva en las empresas.

Datos duros como los presentados, nos refieren la importancia no solo logísticamente hablando que tiene el transporte y las actividades relacionadas al mismo, sino del impacto que puede traer a este sector realizar un análisis completo de la información generada por estos procesos.

La generación acelerada, uso y explotación de los datos en tiempo real esta transformado a las organizaciones en actores proactivos y no reactivos ante los cambios en la industria y las tendencias de consumo de los clientes, softwares de visualización pueden gestionar estos datos y transformarlos en reportes diarios y mostrarlos a través de dashboards de fácil interpretación, aplicaciones sencillas como esta son

emergentes en la industria del transporte.

Diariamente son generados una cantidad exorbitante de datos, tan solo en los últimos dos años se generaron más datos que en toda la historia de la humanidad y se estima que para 2020 cada individuo creará 1.7 megabytes de información nueva por segundo, las empresas exigen saber no sólo lo que sucede hoy, sino también lo que va a pasar en un futuro (McAfee *et al.*, 2012).

Las organizaciones están sujetas a factores externos que afectan directamente su estructura organizacional así como los costos, en el caso del transporte por ejemplo, el precio de la gasolina, refacciones, casetas, costos de seguros son algunos de los que tienen mayor fluctuación e impacto.

Los administradores necesitan de herramientas adecuadas para poder visualizar el estado de sus operaciones en tiempo real y así poder tomar decisiones basadas en información actual.

Big Data va más allá de la consolidación de datos, por sus características contribuye a la correlación de información en tiempo real ya sea en forma estructurada o no estructurada.

La mayoría de las pymes dedicadas a la logística y transporte de materias y/o productos de manera directa o indirecta, cuentan con poco personal y utilizan sistemas básicos para la administración de sus actividades y operaciones, por lo que cuentan con poca visibilidad de sus redes logísticas y suelen ser incapaces de pronosticar su demanda.

Al emplear técnicas de análisis *Big Data* pueden ser detectadas áreas de oportunidad y crear estrategias mediante las cuales realizar un mejor pronóstico de la demanda de manera eficiente.

En este sentido Choi *et al.* (2018) menciona que la previsión de la demanda se basa en gran medida en datos históricos, asesoramiento de expertos e información de mercado. En la era del *Big Data*, tenemos cada vez más herramientas, que poten-

cialmente pueden mejorar el rendimiento de los pronósticos, derivado de un mejor análisis y procesamiento de la información.

Con un mejor pronóstico de la demanda, las compañías también puede optimizar sus precios, especialmente para los productos o servicios recién lanzados o que nunca se han vendido antes (Choi *et al.*, 2018).

Uno ejemplo de las oportunidades que aporta la combinación de *Big Data* en la mejora de los pronósticos y la demanda lo mencionan Cui *et al.* (2018) que explora las previsiones de ventas diarias para un minorista en línea (incluidos datos de redes sociales), y usando varios métodos de aprendizaje automático, los autores muestran que los datos de las redes sociales pueden conducir a una mejora significativa de la previsión de la demanda de acuerdo a las reacciones y preferencias en las plataformas y los datos históricos.

Otro buen ejemplo de de las mejoras y efectividad que aporta el desarrollo de un proyecto *Big Data*, para mejorar los pronósticos lo presenta Sagaert *et al.* (2017) estudiaron los pronósticos de ventas utilizando “datos masivos temporales”, para su caso de estudio, los autores desarrollan un nuevo método que puede automatizar la identificación de factores críticos relacionados con las ventas.

Recientemente, Bertsimas *et al.* (2016) emplean una técnica de optimización basada en datos llamada optimización estocástica condicional para explorar el control de inventario.

Estos son solo algunos ejemplos de casos de éxito en los que se a empleado como solución *Big Data* en la mejora de los pronósticos y la gestión de inventarios.

1.6 METODOLOGÍA

Con base a la revisión de la literatura realizada durante la construcción de esta investigación, se estableció una metodología, que permite la creación y desarrollo

de una herramienta basada en técnicas de análisis *Big data*, con la que es posible realizar la recolección, limpieza, análisis e implementación de métodos de pronóstico para la estimación de la demanda mediante el análisis de series de tiempo de acuerdo a los datos históricos recabados en la base de datos de la empresa caso de estudio.

En la siguiente figura 1.1 se muestra el diagrama de los pasos a seguir para el desarrollo de esta investigación.

1.- Revisión sistemática de literatura: Se lleva a cabo la recopilación de los artículos relacionados con el tema de investigación, de acuerdo con la metodología propuesta por Lacey y Luff (2001).

2.- Establecer Metodología: De acuerdo con la revisión de literatura, son definidas las técnicas a utilizar en la construcción de la metodología para la recolección de información basados en el proceso KDD.

3.- Diseño y creación de la herramienta: De acuerdo a la metodología es diseñada la herramienta para la recolección, limpieza y análisis de datos, con el objetivo de mejorar los pronósticos y la toma de decisiones empresariales.

4.- Experimentación: En esta fase se realiza la experimentación y se exponen los resultados obtenidos así como el análisis de los mismos.

1.7 ESTRUCTURA DE LA TESIS

El desarrollo de este proyecto involucra un total de cinco capítulos. En el primer capítulo se establece la introducción que incluye la razón de ser de la investigación el objetivo, hipótesis y justificación, así como las delimitaciones de esta investigación.

En el segundo capítulo denominado antecedentes, se exponen a profundidad los componentes teóricos que dieron pie a la realización de la metodología, definiciones y conceptos que van de lo general a lo particular, así como casos de estudio

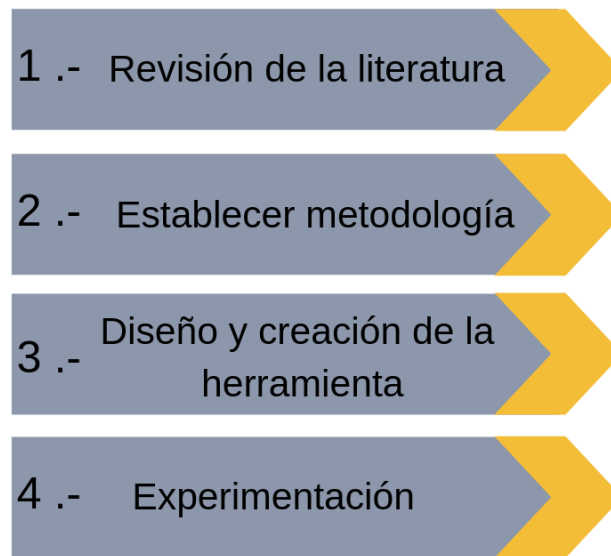


Figura 1.1: Metodología de la investigación

relacionados.

El tercer capítulo, metodología abarca la parte práctica y el desarrollo de la metodología como tal, se exponen los pasos que se llevaron a cabo para crear la herramienta así como el análisis de la información y la generación de pronósticos.

Posteriormente en el capítulo cuatro se describe la empresa caso de estudio, los problemas detectados en ella, como fue empleada la metodología y se muestran los resultados obtenidos, así como el análisis y comparaciones realizadas derivadas de la herramienta. También se incluye la interpretación de los resultados obtenidos durante la fase de experimentación.

Finalmente en el quinto capítulo se presentan las conclusiones y contribuciones explicando las aportaciones dadas gracias al desarrollo de esta investigación.

CAPÍTULO 2

ANTECEDENTES

La construcción de este documento se basa en la revisión bibliográfica de libros, revistas científico/comerciales y estudios especializados a nivel nacional e internacional.

La revisión y búsqueda de literatura fue realizada con las siguientes palabras clave:

- *Big Data*
- ventaja competitiva
- transporte
- logística
- pronósticos y
- series de tiempo

Estas palabras fueron combinadas entre sí en el idioma español e inglés para obtener mayores resultados.

Este capítulo se divide en cuatro partes, en la primera se presenta de forma general los conceptos de cadena de suministro y estrategia competitiva, así como los

componentes de la cadena de suministro en los que centramos nuestra investigación, la segunda parte se enfoca en las TIC's e industria 4.0 aplicadas en la cadena de suministro.

En la tercera parte se define y amplía el concepto de *Big Data* así como la presentación de los resultados en algunos casos de estudio que han tenido éxito, aportaciones de esta tecnología prominente en diversos sectores y se las diferentes técnicas y herramientas de análisis *Big Data*.

Finalmente se presenta las definiciones de serie de tiempo y pronósticos así como los principales métodos de pronósticos utilizados en series de tiempo.

Cada sección, se construye bajo la siguiente estructura:

- definición
- características
- ventajas e impacto en la cadena de suministro
- ejemplos y/o casos.

2.1 CADENA DE SUMINISTRO Y ESTRATEGIAS COMPETITIVAS

Sin importar el sector o giro al cual pertenezca una empresa, se ha detectado la necesidad de contar con una sólida y bien estructurada Cadena de Suministro.

A pesar de que no todas las industrias son iguales y responden a diferentes necesidades y mercados, se requiere que cuenten con procesos establecidos y que el nivel de integración de la cadena sea tal que les permita convertirse en socios estratégicos a largo plazo.

De acuerdo con el Council of Supply Chain Management la cadena de suministros se puede definir como «Todas aquellas actividades de la gestión de la logística, de la planificación y gestión de todas las actividades involucradas en la obtención y adquisición, conversión y la coordinación y colaboración con los socios, proveedores, intermediarios, proveedores de servicios externos y clientes», partiendo de este concepto, podemos destacar que las organizaciones ya no compiten uno a uno si no que la cadena de suministro tiene su propia justificación fuera de los intereses de la empresa individual.

Silva (2017) menciona que la administración de la cadena de suministro en una organización requiere de gran esfuerzo, debido a que no solo afecta la estructura logística, sino también la estructura organizacional y direccional.

Se requiere de la integración y participación de todos los departamentos que la integran para funcionar de manera exitosa.

La figura 2.1 que a continuación se presenta, muestra un esquema representativo de los principales actores dentro de la cadena de suministro tradicional.

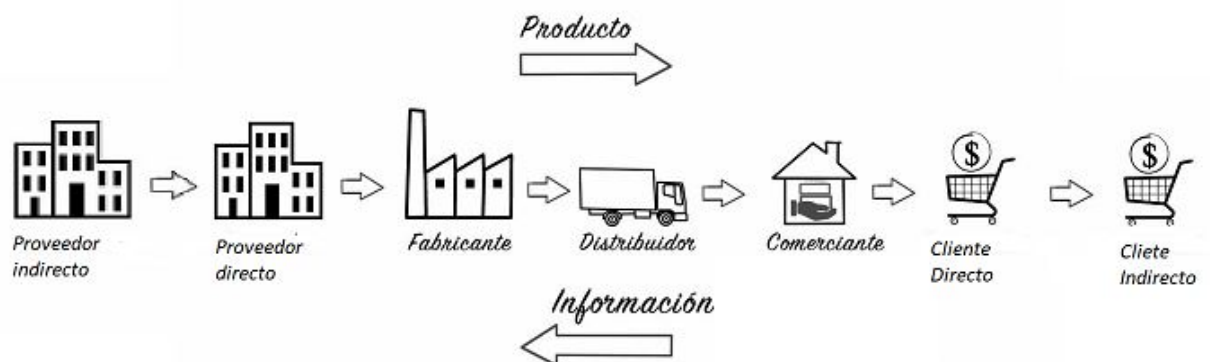


Figura 2.1: Cadena de suministro tradicional

Como podemos observar para que una organización pueda alcanzar una posición ventajosa en el mercado global, requiere de la colaboración de múltiples actores que en conjunto desarrollen habilidades y capacidades que le permitan enfrentar la

hiper competitividad que se vive en el actualmente.

Además, se requiere por parte de los gerentes y tomadores de decisiones la formulación e implementación de estrategias que les permitan dar respuesta y atender los desafíos de su entorno a mediano y largo plazo (Narváez y Fernández, 2008).

Para las empresas, el desarrollo de estrategias es fundamental, en gran medida determina la forma en que harán frente al mercado y su competencia, de ello dependerá su supervivencia.

Quero (2008) menciona que las estrategias competitivas pueden ser conceptualizadas como una herramienta fundamental para los directivos de las organizaciones, con estas buscan alcanzar una posición favorable en el mercado a través de su creación, ya sea ofreciendo un producto al costo mínimo (ventaja de costo) u ofreciendo productos o servicios con atributos únicos apreciables por los clientes (ventaja de valor).

De acuerdo a su estudio Quero (2008) también deduce que la competitividad tiene que ver con los siguientes indicadores: costos, precios, cantidad, calidad, presencia en el mercado, con la innovación, la flexibilidad y adaptación a los cambios, fortalecer y desarrollar la reflexión y el análisis.

Es de suma importancia que las empresas puedan identificar las áreas de oportunidad que les permitirán establecer los proyectos o estrategias de mejora para competir de manera eficiente (Monge, 2010).

Un gran referente en cuanto a estrategia competitiva se refiere, es Porter (1900) que, en su libro denominado “Ventaja competitiva: creación y sostenimiento de un desempeño superior”, menciona qué si una organización desea mejorar su ventaja competitiva, debe lograr el equilibrio entre costos y diferenciarse de la competencia.

Esto solo puede ser posible, analizando las operaciones, estudiando el comportamiento de los sistemas a lo largo del tiempo y estableciendo parámetros que permitan medir el nivel de servicio, para identificar áreas de oportunidad y desarro-

llar propuestas de mejora.

Harrison y New (2002) presentaron los resultados de una importante encuesta internacional sobre la relación entre la estrategia corporativa, la estrategia de la cadena de suministro y la de gestión del rendimiento de la cadena de suministro, entre las principales conclusiones que obtuvo, mencionaba que las empresas solían considerar como principal ventaja competitiva la estructura en la cadena de suministro y como valor añadido la integración de las tecnologías de información en sus procesos.

“El agregar valor a un producto o servicio a lo largo de la cadena de suministros es esencial para la competitividad de las organizaciones y en algunas ocasiones vital para la supervivencia de las mismas” (Félix *et al.*, 2013).

Desde hace años las compañías han detectado la necesidad de integrar a sus operaciones las aplicaciones tecnológicas disponibles para llevar el control y registro de sus operaciones, además de favorecer gradualmente el rendimiento e integración de las cadenas de suministro.

La cadena de suministros está compuesta por diferentes actores, que en conjunto trabajan para brindar un bien o servicio, uno de los principales actores dentro de la cadena de suministro es la logística.

De acuerdo con Lambert *et al.* (1998) la logística, es la parte encargada de planificar y controlar el flujo eficiente de materiales, el almacenamiento de productos, así como toda la información asociada desde el punto de origen hasta el consumo. Es importante tener claro este concepto ya que realza la importancia que tiene la logística en el flujo eficiente no solo de los materiales y productos, sino de toda la información que se genera a lo largo de la misma.

Los retos en la administración logística han cambiado, se está constituyendo un nuevo factor de diferenciación competitiva y para conseguirlo es necesario contar con un desarrollo mucho más acelerado, estructurado y adaptable de manera cooperativa, a nivel de filosofía, estructura y tecnología en las organizaciones.

En un sistema típico, la estructura logística cuenta con las siguientes actividades: servicio al cliente, pronóstico de la demanda, distribución, control de inventarios, manejo de materiales, procesamiento de pedidos, almacenamiento, compras, empaque, manejo de bienes, eliminación de mercaderías aseguradas y desperdicios, tráfico y transporte (Gibson *et al.*, 1998).

Algunas de las principales actividades logísticas, que se consideran críticas para muchos se encuentran: el transporte, el almacenamiento.

Acorde a esto, el transporte por su parte permite lograr una mayor competitividad a las empresas, por ello, es importante desarrollar estrategias a través de las cuales puedan generar una mayor agilidad en tiempos de respuesta y minimizar los tiempos de entrega.

En México el auto transporte, es el principal medio de transporte de bienes, moviliza alrededor de 55.9% del volumen de la carga en el país (CANACAR, 2015).

La mayor parte de los costos asociados al sector autotransporte suelen verse afectados por diversos factores externos como por ejemplo: el precio del combustible, el mantenimiento de la flota, las refacciones, peajes y casetas han ido al alza, impactando los costos de operación y reduciendo los márgenes de ganancia de las empresas que operan en este giro.

Al continuar a la alza el incremento en los costos relacionados a las operaciones logísticas, las empresas se ven obligadas a incrementar sus precios, afectando a los clientes, por ende, se encuentran en constante búsqueda de generación de estrategias que les permitan reducir costes, aminorar las demoras, paros por fallas e incrementar la rentabilidad de sus unidades.

Los pronósticos por su parte permiten mejorar el proceso de abastecimiento y control de inventarios ya que a través de una eficiente construcción y previsión de la demanda es posible mantener los recursos en la cantidad y momento que se requieren.

2.1.1 TECNOLOGÍAS DE LA INFORMACIÓN E INDUSTRIA 4.0 EN LA CADENA DE SUMINISTRO

En la siguiente sección se describen y ejemplifican los cambios, así como la evolución que se ha presentado en las cadena de suministros tradicionales con la incursión de las TIC's y algunas de las tecnologías prominentes relacionadas con Industria 4.0 que actualmente está teniendo gran impacto.

Las TIC's se han convertido en un medio facilitador para que la cadena de suministro y las partes que la integran, cumplan sus objetivos. Facilitan el almacenamiento y transmisión de información logística, misma que posteriormente, se convierte en la base para mejorar la toma de decisiones de manera eficiente.

Para lograrlo, las empresas invierten en sistemas de información y el desarrollo de nuevas herramientas tecnológicas (Espinal y Montoya, 2009).

Las TIC's, permiten lograr un desarrollo empresarial exponencial y en algunos casos se convierten en parte de la estrategia mediante la cual generan su propia ventaja competitiva así como trabajar en un ambiente globalizado, donde la información oportuna y de calidad se convierte en el mejor aliado (Espinal y Montoya, 2009).

La tecnología se a mantenido como el principal vehículo para mejorar la competitividad y rendimiento de la cadena de suministro al mejorar la eficacia global y la eficiencia de los sistemas logísticos.

Elegir la tecnología adecuada para apoyar las actividades o subprocesos, es crucial para que una empresa pueda obtener ventaja competitiva sobre la competencia (Introna, 1991).

Gran parte de los avances en este sector se han enfocado en el desarrollo de herramientas informáticas con un enfoque integral.

“Existen gran variedad de TIC's en los procesos logísticos de la cadena de su-

ministro. Para entender mejor como se involucran en las operaciones las dividiremos para su análisis en tres secciones.

Logística de entrada: encargada o parte clave en el abastecimiento y , logística interna: encargada de coordinar o ser parte de los procesos internos de la empresa.

Logística de salida: encargada de llevar a cabo y coordinar la distribución.

En la logística de entrada se encuentran los siguientes sistemas o tecnologías de la información: EDI, VMI, CRP, *e-procurement* y *e-sourcing*. , en la logística interna se utilizan sistemas tales como el ERP, MRP I, MRP II, WMS y en la logística de salida se considera que suele interferir o ser de mas uso sistemas como el TMS, EDI, EPC” (Espinal y Montoya, 2009).

Muchas operaciones se agilizaron y sistematizaron de manera acelerada gracias a la adopción de estas tecnologías.

A continuación, definiremos algunas de las ventajas con las que cuentan los sistemas encargados de administrar y coordinar las actividades en que hemos enfocando nuestra investigación.

- ERP: Software que suele utilizarse en la planificación y coordinación de recursos dentro y fuera de la empresa, así como el flujo y generación de información.

Las ventajas de los ERP residen principalmente en la utilización de una única base de datos, lo que facilita la comunicación e intercambio de información entre los departamentos de la empresa y evita la redundancia y duplicidad de la información (Espinal y Montoya, 2009).

- WMS: Se emplea para la administración y gestión de almacenes en la red logística, algunas de las ventajas potenciales con la utilización del WMS son el contar con conocimiento en tiempo real de la utilización de los recursos del almacén; la reducción en costos debido a la optimización de operaciones y la

mejora en la calidad del servicio, el cual implica, exactitud en el cumplimiento y la fiabilidad en los tiempos de entrega.

- TMS: Se enfoca en mejorar la administración y optimizar los recursos de transporte conciliando su menor coste con los estándares necesarios de servicios al cliente, y los requisitos de otros agentes de la cadena de suministro.

El TMS permite a las organizaciones contar con interacciones entre un sistema de gestión de pedidos y el centro de distribución o su propio almacén.

Si bien los sistemas de información han madurado, cada vez generan más información en diversas modalidades, debido al uso de sensores y a la captación de la información de fuentes externas como el clima, tráfico y satélites.

Es necesario adaptar los nuevos requerimientos a los sistemas de transporte y herramientas para lograr una mejor conectividad entre sensores y máquinas, ya que los datos que arrojan estos mecanismos pueden ser almacenados y transformados para su posterior análisis.

En la organización de una empresa, el responsable de logística está tomando cada vez más relevancia en los procesos de toma de decisiones estratégicas del negocio.

Deberá tener en cuenta qué nivel y grado de adaptación a nuevos sistemas tiene su organización. No se trata de elegir acorde al sistema de mayor facilidad de integración, sino elegir lo más conveniente para sostener el desarrollo de los procesos y modelo de negocio de cada empresa.

Algunos de los principales cambios dados en el sector industrial son vistos como un paradigma integral, al cual denominan como la cuarta revolución industrial esto de acuerdo con Barreto *et al.* (2017).

Industria 4.0 abarca el desarrollo e integración de tecnologías innovadoras de información y comunicación en la industria. El objetivo principal es fomentar la

interconexión inteligente de productos y procesos a lo largo de la cadena de valor, lo que permite utilizar de manera eficiente los procesos de la organización, en la creación de bienes y servicios en beneficio del cliente.

Saucedo-Martínez *et al.* (2017) determinaron con base a su investigación, los siguientes bloques que considera componen las Tecnologías prominentes de Industria 4.0:

- *Big Data* y análisis
- Robots autónomos
- Simulación
- Integración de sistemas horizontal y vertical
- Internet de las cosas
- Ciberseguridad
- Fabricación aditiva
- Realidad aumentada

Es importante que las organizaciones logren adaptarse a los cambios e innovaciones, debido a que impactan la manera en que se realizan los procesos dentro de las organizaciones.

Buscan mejorar y optimizar las actividades y procesos internos en las organizaciones, además de automatizar las actividades diarias que no agregan valor.

El concepto de Logística 4.0, se emplea para referirse a la combinación del uso de la logística con las innovaciones y aplicaciones agregadas.

Logística 4.0 está relacionada con las mismas condiciones que los servicios y productos inteligentes que permite la comunicación entre seres humanos y máquinas en tiempo real (Barreto *et al.*, 2017).

“Logística 4.0 emplea las siguientes aplicaciones tecnológicas: 1) Planificación de recursos, 2) Sistemas de gestión de almacenes, 3) Sistemas de gestión de transporte, 4) Sistemas de transporte inteligentes y 5) Seguridad de la información” (Barreto *et al.*, 2017).

Logística 4.0 utiliza datos en tiempo real para lograr aumentar la eficiencia y eficacia en los procesos logísticos.

Las organizaciones están profundamente interesadas en la búsqueda de nuevas iniciativas tecnológicas a bajo costo, lo que les permitiría ofrecer mejores e innovadores servicios, dándoles mayor ventaja respecto a la competencia.

Zmud *et al.* (2013) refieren que el tiempo que se invierte en el transporte de mercancías ha empeorado para casi todos los sectores, además de ir en aumento los costes relacionados con esta actividad, razón por la que las empresas buscan emplear acciones a través de las cuales puedan reducir tiempos de espera y minimizar los costos, principalmente.

En este aspecto las tecnologías han logrado impactar no solo en el funcionamiento de los sistemas y las formas en que los individuos organizan sus viajes, si no en la administración y creación de rutas en tiempo real con el fin de reducir estos tiempos.

Parte de la investigación se centra en el análisis del comportamiento de los cambios y mejoras de manera indirecta al transporte mediante el uso de tecnologías, así como la necesidad de administrar gestionar datos para mejorar la información sobre el transporte de carga y los servicios que integran las actividades logísticas, como la distribución, almacenamiento y servicios de mantenimiento respectivamente.

Al inicio de este capítulo en la figura 2.1 incluimos un esquema de la cadena de suministro tradicional, a continuación, colocamos un esquema representativo de los cambios e impacto con la inclusión de algunas tecnologías prominentes de acuerdo con la Figura 2.2

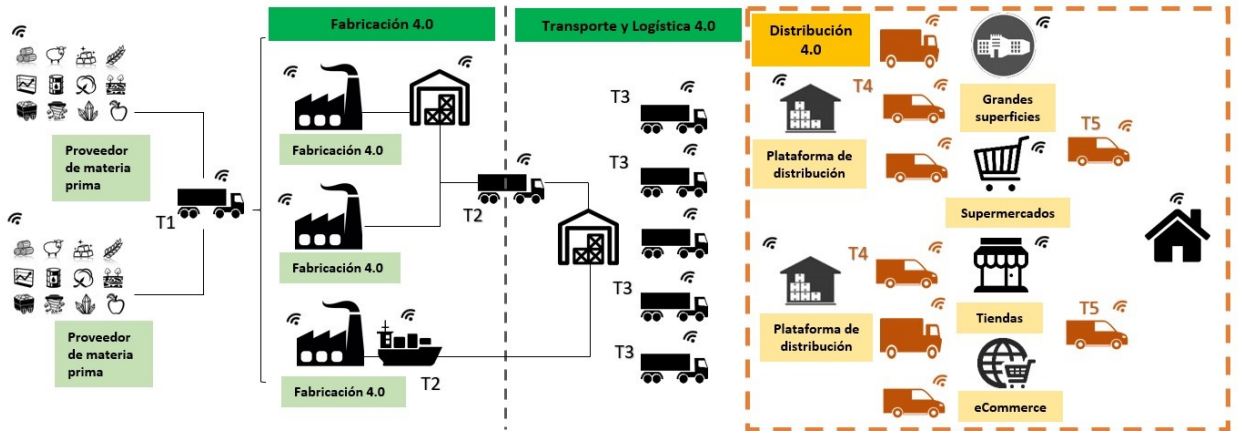


Figura 2.2: Cadena de suministro en entorno 4.0

Se han incorporado diversas herramientas tecnológicas a lo largo de la cadena de suministro desde la Logística de entrada hasta la Logística de salida, se ha buscado conectar los sistemas informáticos de la empresa con los sistemas de comunicación de los proveedores, transportes y canales de distribución.

Se han interconectado los sensores con los sistemas de almacenamiento para controlar la entrada y salida de mercancía, mejorando la administración y eficiencia de los almacenes, además de adoptar el uso de equipos tecnológicos como lentes de realidad virtual para facilitar y minimizar los tiempos en el proceso de picking, en la distribución de última milla la más reciente innovación es uso de drones para la distribución de paquetería.

Otro cambio notorio es la extracción de información recabada por medio de los sensores y sistemas que permite decodificar los patrones previos a la demanda y a la venta de bienes, mejorando la predicción de las necesidades de los clientes.

Este tipo de soluciones contribuyen a la automatización la cadena de suministro, una clara diferencia con respecto a la cadena de suministro tradicional, es que, si bien se buscaba aumentar la agilidad en las operaciones, no ponía gran interés en el incremento de la rapidez en la automatización y digitalización de la información y cuyo principal objetivo es lograr la autonomía de sus equipos y procesos, sin

intervención humana.

2.1.2 APLICACIONES *Big Data* EN LA LOGÍSTICA Y TRANSPORTE

En este marco una de las principales tecnologías prominentes vinculadas con el concepto de Industria 4.0, es *Big Data* ya que la cantidad de información y datos generados disponibles y recopilados a través de las empresas está creciendo a un ritmo acelerado (McAfee *et al.*, 2012).

“El concepto de *Big Data* se refiere fundamentalmente a volúmenes masivos y complejos de información tanto estructurada como no estructurada, que es recogida durante cierto período de tiempo y que requiere de métodos computacionales para extraer conocimiento” (Arcila-Calderón *et al.*, 2016).

Casado y Younas (2015) mencionan que generalmente en el procesamiento de los datos se siguen cuatro fases: adquisición de datos, almacenamiento de datos, análisis de datos y la explotación de datos, como se muestra en la figura 2.3.

Cabe señalar que el modelo de ciclo de vida no sigue estrictamente la secuencia en la que aparecen las fases. Es decir, algunas fases pueden tener un vínculo hacia atrás a la fase anterior. Por ejemplo, la fase de almacenamiento de datos puede tener un enlace hacia atrás a la fase de adquisición de datos. Esto es posible en la transmisión de datos donde los datos se adquieren y almacenan continuamente.

Adquisición de datos: Esta es la primera fase en todos los paradigmas principales de procesamiento de datos.

Almacenamiento de datos: Esta fase se refiere al almacenamiento de *Big Data*. La mayoría de los datos adquiridos deben almacenarse en algún lugar (físico o digital) para su posterior procesamiento o análisis.

Análisis de los datos: Esta fase involucra varios modelos, técnicas y algorit-

mos que se utilizan para procesar y analizar *Big Data* para varias aplicaciones, por ejemplo, pronosticar tendencias de negocios, análisis de ventas, análisis de datos de tráfico o información relacionada con el clima.

Explotación de datos: Esta fase se refiere a la explotación o resultados de *Big Data*. En otras palabras, involucra los resultados u observaciones del análisis realizado en las fases anteriores.

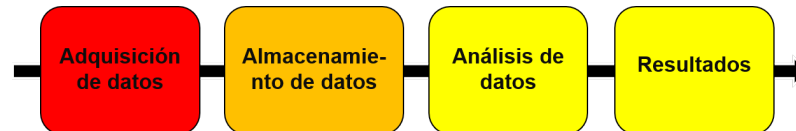


Figura 2.3: Adaptado de Casado y Younas (2015), fases de *Big data*

Esto plantea un desafío para las empresas, ya que complica la identificación y extracción de la información relevante requerida para mejorar la administración de la Cadena de Suministro.

Big Data Analytics marca el surgimiento de nuevas oportunidades en la explotación y utilización de grandes conjuntos de datos (Gupta *et al.*, 2012).

Big Data no hace referencia únicamente a los datos, sino que también comprende todo el espectro de técnicas, métodos, herramientas y tecnologías alternativas, que permiten resolver problemas que involucran cierta complejidad, de una forma más eficaz que los métodos de análisis tradicionales” (Tabares y Hernandez, 2014).

Wang y Moriarty (2018) refieren que la era de *Big Data* esta sobre nosotros, esta tecnología prominente supone una oportunidad para analizar los datos de manera libre, amplia y administrar diversas fuentes de datos sin un esquema común.

Durante el desarrollo de esta investigación, específicamente para la comprensión, análisis de técnicas y herramientas empleadas en implementaciones de proyectos del tipo *Big Data* se recopilaron un total de 41 artículos científicos, publicados entre los años 2011 y 2018, en diferentes revistas indexadas, con base al análisis de estos artículos se determinaron algunos de los campos en que ha tenido mayor aplicación

Big Data.

En la figura 2.4 podemos ver la gráfica que describe las principales áreas en las que se han desarrollado e implementado proyectos *Big Data*:

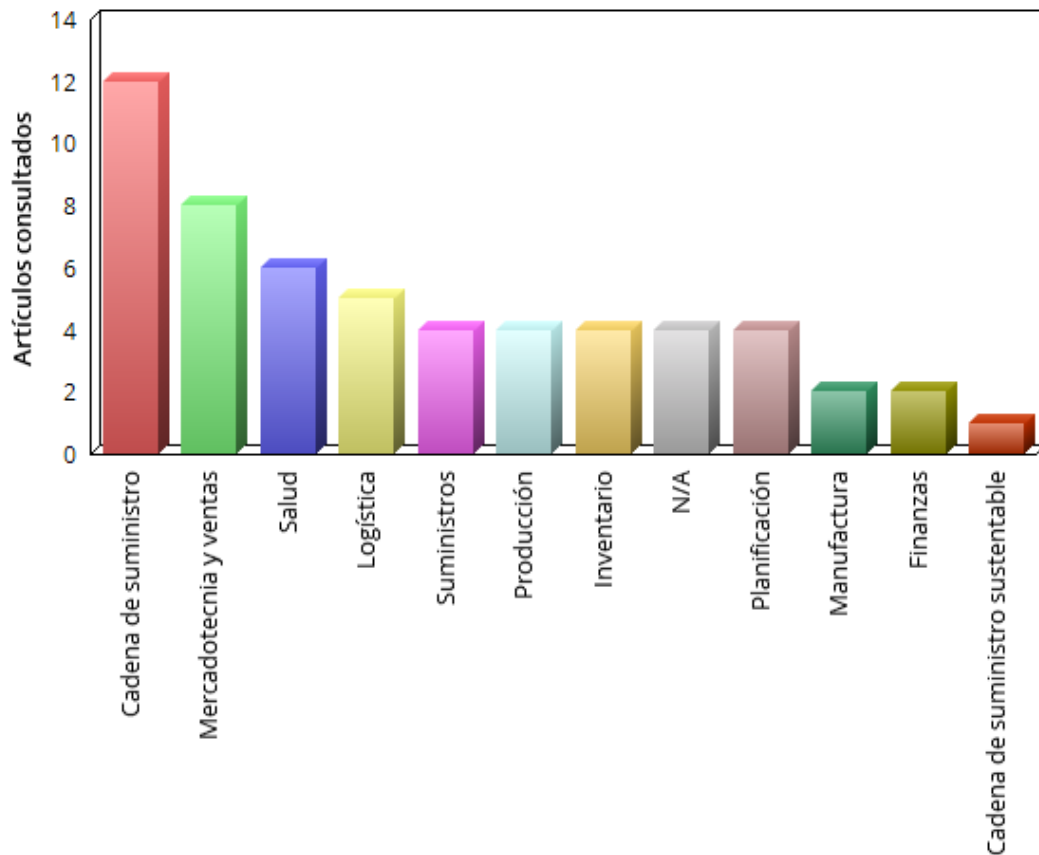


Figura 2.4: Elaboración propia

Big Data aporta nuevos conocimientos y correlaciones útiles, que proporcionan conocimiento útil para descubrir y generar estrategias y mejorar la toma de decisiones, su aplicación en los sistemas logísticos aporta una mejor comprensión, visibilidad y trazabilidad.

Uno de los más recientes casos de estudio analizados referentes a la aplicación de *Big Data* para mejorar las operaciones de las redes de transporte fue el estudio realizado en la ciudad francesa de Lyon.

Se llevo a cabo la captura de datos mediante sensores repartidos por toda la ciudad, con los cuales se analizo la congestión de tráfico, con el objetivo de predecir el momento exacto en el cual podía producirse un atasco, de tal manera que los sistemas de monitoreo de transito cambiaran automáticamente las señales de los semáforos mejorando el flujo circulatorio (Lordi, 2017).

Algunas agencias de movilidad, emplean *Big Data* para responder preguntas y tomar decisiones en un tiempo menor. Son muchas las ventajas que puede aportar *Big Data* en la planificación, gestión y movilidad.

No es sólo mirar la evolución histórica del comportamiento del tráfico, sino también ver la reducción y optimización de los tiempos de transporte, la disminución de las emisiones de CO₂, la planificación de rutas, áreas de congestión de tráfico, hábitos de transporte y muchas otras variables que pueden ser monitoreadas a través de sistemas *Big Data* en tiempo real.

Otros casos exitosos de la aplicación de *Big Data*, los encontramos en innovadores sistemas analíticos de compañías como American Airlines que buscan mejorar y agilizar las reservaciones efectuadas de manera electrónica, por su parte, United Parcel Service lo a empleado para mejorar la identificación de patrones en el comportamiento de compra de sus clientes, de acuerdo al historial de compras identifican la probabilidad de que realicen una reservacion o la compra de algún vuelo.

Amazon por su parte, hace gran uso de la información en la creación y mejora de “algoritmos de recomendación” la finalidad; aumentar las probabilidades de compra de sus clientes con base a su historial de compras y navegación en su pagina web.

Otro grande que hace uso de la información a través de metodologías basadas en *Big Data* es Netflix, empleándolo en gran medida en la creación de modelos de elección para el consumidor, generación de recomendaciones y creación de nuevo contenido con base a preferencias e historial de visualizaciones de sus clientes.

Viglioni *et al.* (2007) implementaron un proyecto de minería de datos para

mejorar la predicción de demandas ferroviarias, para facilitar la operación y la planificación de la mano de obra, por su Greis y Nogueira (2011) emplearon esta técnica para mejorar la identificación de envíos de alto riesgo que llegan a los puertos de EE. UU.

Rebón, Castander, Argandoña, Gerrikagoitia y Alzua-Sorzabal (2015) mejoraron la técnica de análisis en el proceso de toma de decisiones de detección de transacciones de fraude crediticio.

Estos son solo algunos ejemplos de los campos en los que se a empleado alguna de las técnicas de análisis big data y las mejoras que ha proporcionado de acuerdo con los estudios presentados.

Algunas de las aplicaciones basadas en *Big Data* han mejorado el rendimiento de las empresa al optimizar la administrar de la cadena de suministro, contribuir en la creación de nuevos productos y mejorar las relaciones con los clientes.

La economía de datos crea miles de exabytes nuevos a través de las plataformas digitales, incluidos dispositivos móviles y redes sociales, dando la oportunidad a las empresas de que puedan utilizar esta información a su favor (Cheah y Wang, 2017).

En la figura 2.5 presentamos algunas de las principales ventajas que aporta *Big Data* a la cadena de suministro de acuerdo con Tabares y Hernandez (2014).

El reto para muchas empresas; aumentar la rentabilidad del negocio, para esto se necesita usar la información a su favor, ahí es donde actúa *Big Data*, elabora un examen interno, mediante el cual ofrece soluciones que lleven a las empresas a mejorar sus servicios y repensar su estrategia competitiva.

Big Data no solo contribuye a desarrollar y mejorar los productos y servicios, sino que también permite a las empresas identificar nuevas e inesperadas fuentes de competencia en todas las industrias.

Un gran ejemplo del desarrollo de nuevos servicios y compañías, con base al

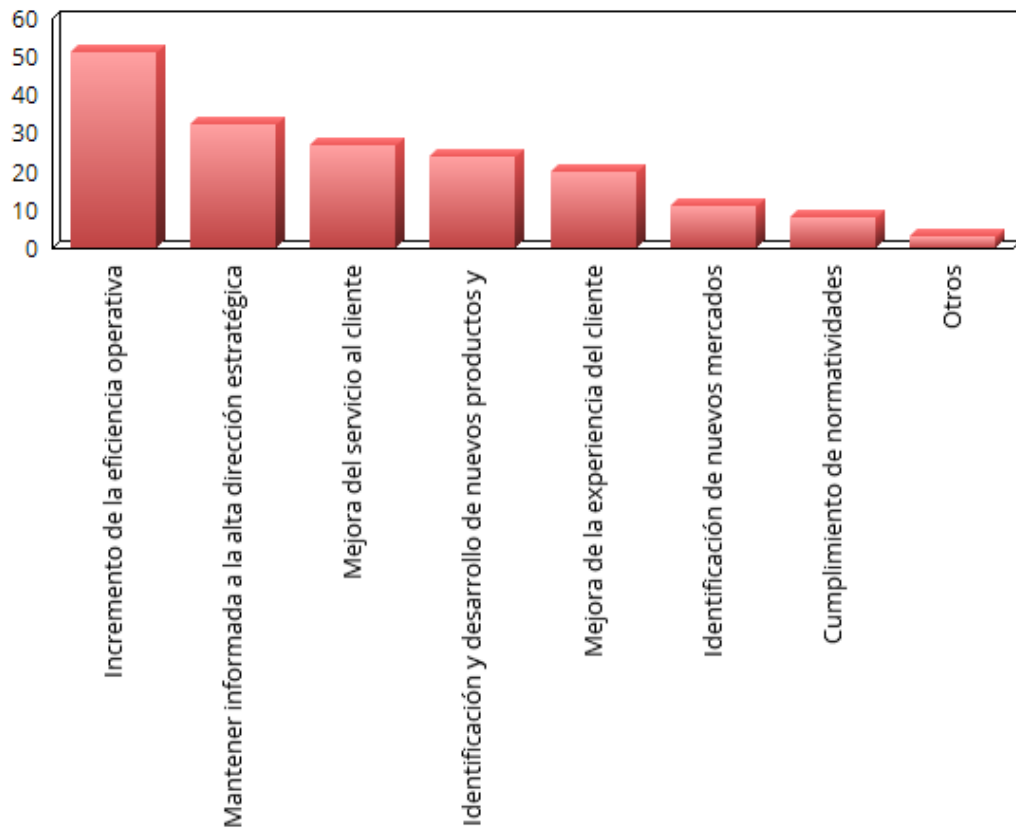


Figura 2.5: Ventajas que proporciona el análisis de *Big Data*

análisis de la información es el caso de Uber o Didi, este tipo de compañías comenzó a ganar terreno en un nicho que se creía tradicional haciendo uso de la tecnología. *Big Data* permite la oportunidad de cambiar el diseño del modelo de negocio (Tunguz y Bien, 2016).

Durante la fase del análisis de la información es posible revelar patrones ocultos y correlaciones secretas llamadas analíticas de *Big Data*, esta información está siendo útil para muchas compañías y organizaciones que busca obtener una ventaja sobre la competencia (Afacan-Seref *et al.*, 2018).

Emplear técnicas de análisis *Big Data* en la logística y el transporte puede proporcionar una gran oportunidad no solo de mejorar y optimizar las actividades si no de crear nuevos modelos de negocios, conocer mejor al cliente, mejorar la predicción de la demanda.

2.1.3 TÉCNICAS DE ANÁLISIS *Big Data*

En esta sección, examinamos varias técnicas de uso común para el análisis de información. Estas técnicas no se excluyen mutuamente y se superponen en cierta medida de acuerdo con el ciclo de vida de *Big Data* (figura 2.3).

«El análisis de big data implica el procesamiento de datos de diferentes fuentes y en diferentes formatos, pueden provenir de la web, los medios sociales, los sistemas ERP y las plataformas en la nube, y se pueden proporcionar en formatos de texto, gráficos, audio y vídeo» (Choi *et al.*, 2018).

Es importante destacar que nuestra investigación se centra mas no es de carácter exclusivo en las empresas del tipo pyme, por lo cual, el costo sera una de las restricciones a considerar a la hora de realizar la selección de la tecnología a emplear. Se desea que esta sea de uso libre, es decir que no genere costo de instalación o sea necesario comprar una licencia de uso.

A continuación se enlistan una serie de técnicas para realizar el análisis de datos.

- Estadística

La Estadística normalmente se ocupa de la recolección, agrupación, presentación, análisis e interpretación de datos. Contrario a lo que piensa, la estadística no implica solamente los resultados de encuestas, ni el cálculo de porcentajes, la estadística es un método científico que permite sacar conclusiones a partir de observaciones hechas o de colecciones de datos recolectados.

La estadística descriptiva es el primer paso en el análisis de datos, con la finalidad de conocer la estructura de los datos, verificar la calidad de los datos y validarlos. También, permite conocer los primeros *insights* del negocio, con los cuales es posible dirigir nuestra atención hacia una problemática, así como obtener la representación de los datos de manera visual, descriptiva y categórica.

El análisis estadístico multivariado también es una herramienta poderosa para el análisis de negocios. Se sabe que las estadísticas son rápidas y, por lo tanto, pueden usarse para superar los requisitos de cómputo rápido relacionados con el análisis de *Big Data*.

El único problema de usar esta clase de técnicas de análisis de datos es que se basan en probabilidad y tienen cierto grado de error en sus predicciones.

Por ejemplo, si los datos tienen un patrón muy extraño, como una curva no normal o una gran cantidad de valores atípicos, la desviación estándar no ilustrará la realidad, al no poder aportar toda la información que se necesita.

Sin embargo sigue siendo una técnica ampliamente utilizada en el análisis de datos y sirve como base de otras técnicas como la minería de datos.

- Regresión

En informática y estadística la regresión es una herramienta muy útil para realizar proyecciones futuras o estimaciones, sirve para determinar tendencias a lo largo del tiempo, ya que modela las relaciones entre variables dependientes y explicativas, que generalmente se grafican en un diagrama de dispersión, indicando si los vínculos que existen son fuertes o débiles.

- Visualización de datos

La visualización de datos es una de las técnicas de análisis de datos más utilizada hoy por hoy, por lo fácil que resulta a través de un gráfico o imagen detectar patrones en los datos. Es especialmente útil cuando buscamos entender el comportamiento de grandes volúmenes de datos de forma rápida y simplificada.

- Análisis de escenarios

El análisis de escenarios consiste en analizar una variedad determinada de eventos futuros con resultados alternativos. Es bueno utilizarlo cuando no estamos seguros sobre qué decisión tomar o que curso de acción perseguir.

- Aprendizaje automático

De acuerdo con Choi *et al.* (2018), los métodos bien establecidos, como las redes neuronales, las máquinas de vectores de soporte y el aprendizaje estadístico de máquinas, se clasifican como aprendizaje automático. El aprendizaje automático proporciona algoritmos para que las computadoras descubran conocimientos y tomen decisiones aprendiendo de los datos.

- Redes neuronales

Esta tal vez sea una de las técnicas de análisis de datos más complejas que existen. Las redes neuronales tratan de simular el proceso de decisión e información del

cerebro o grupos de neuronas. El objetivo de estas redes es simular el proceso de aprendizaje de un cerebro humano en una computadora para facilitar la toma de decisiones, o automatizar procesos con base a la experiencia y prueba y error.

- Proceso KDD (Knowledge data discovery) de minería de datos

Se ha definido como “el proceso no trivial de identificación en los datos de patrones válidos, nuevos, potencialmente útiles, y finalmente comprensibles” (Fayyad *et al.*, 1996).

De acuerdo a la línea de investigación de Han, Kamber y Pei (2006), el proceso KDD se puede dividir en las siguientes fases:

1. limpieza de datos
2. integración de los datos
3. selección de datos
4. transformación de los datos
5. minería de datos
6. evaluación de patrones
7. presentación del conocimiento.

Los conceptos KDD y minería de datos suelen confundirse o tomarse como un sinónimo pero realmente la técnica de minería de datos es uno de los pasos a seguir en el proceso KDD.

Esta mas que una técnica es una metodología que marca los pasos a seguir para realizar un proyecto de análisis de datos desde la recolección hasta la entrega de resultados.

- Data mining

«La minería de datos tiene una larga historia, con fuertes raíces en la estadística, inteligencia artificial, aprendizaje automático e investigación de bases de datos» (Hand, 2006).

En la actualidad, la minería de datos está altamente especializada con muchas áreas y enfoques funcionales diferentes. Por ejemplo, tenemos minería secuencial y temporal, minería espacial, minería de procesos, minería de preservación de la privacidad, minería de redes, minería web, etc., todos los cuales están asociados con el análisis de *Big Data* (Choi *et al.*, 2018).

Esta técnica se usa para detectar patrones, relaciones o información relevante que pueda mejorar el desempeño de operaciones relacionadas con el cliente. La minería de datos hace uso de un conjunto de técnicas para encontrar información relevante dentro de un conjunto de datos.

En el diagrama 2.6 se presentan algunas de las principales técnicas utilizadas en proyectos de minería de datos.

- Clustering

Son técnicas que parten de una medida de proximidad entre individuos y a partir de ahí, buscar los grupos de individuos mas parecidos entre si, según una serie de variables medidas.

- Series Temporales

A partir de la serie de comportamiento histórica, permite modelizar las componentes básicas de la serie, tendencia, ciclo y estacionalidad y así poder hacer predicciones para el futuro, tales como cifra de ventas, previsión de consumo de un producto o servicio, etc.



Figura 2.6: Técnicas de análisis *Data Mining*

- Árboles de decisión

Permiten obtener de forma visual las reglas de decisión bajo las cuales operan los consumidores, a partir de datos históricos almacenados. Su principal ventaja es la facilidad de interpretación.

- Simulación Monte Carlo

Esta técnica de probabilidad matemática es usada para medir el riesgo aproximado de que un hecho determinado tenga lugar. Es muy útil para entender las implicaciones que puede tener un determinado curso de acción derivado de una decisión.

- Programación y optimización matemática

También conocida como optimización lineal, es un método para identificar cual es mejor resultado posible dadas unas restricciones concretas a nuestra situación. Se

utiliza mucho para resolver problemas dados en procesos de producción y determinar cómo minimizar los costes o maximizar los beneficios.

- Predicción matemática

Es un conjunto de técnicas estadísticas que emplea datos de series temporales para predecir cuál es el resultado más probable que se puede dar en el futuro cercano. La base de estas técnicas de análisis de datos es fijarse en que es lo que ha ocurrido en el pasado para saber qué ocurrirá en el futuro. Es muy utilizado en proyecciones macroeconómicas.

Como se enfatizo en el inicio de este capítulo la elección de las técnicas es inherente y estas no se interpone una ente otra por lo que se realiza una combinación de estas para dar respuesta a diversas interrogantes, el proceso de análisis de datos dependerá de lo que se desea identificar, analizar o conocer, para cada caso.

A manera de resumen a continuación en la figura 2.7 se muestra un esquema de funciones y algoritmos de algunas de las técnicas de análisis de datos de acuerdo con Serrano-Cobos (2014).

Como pudimos observar, en la mayoría de los proyectos *Big Data* se hace uso de diversas metodologías y técnicas o una combinación de ellas para extraer conocimiento.

Nuestro caso no es la excepción, por lo que no se hará uso de una sola técnica de análisis, si no que para poder analizar adecuadamente los datos sera necesario hacer una combinación de las mismas.

La combinación de técnicas de análisis a su vez nos asegura obtener mejores resultados así como hacer mas robusta nuestra metodología ya que con cada técnica se aplica un filtro y análisis diferente que permite concentrar toda la información correcta en un solo sistema.

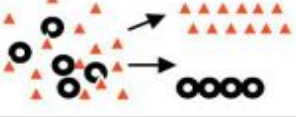


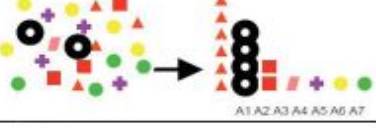
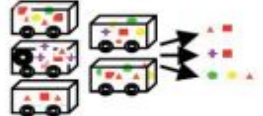
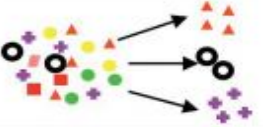
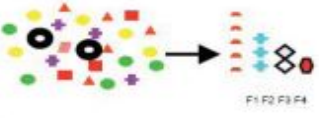
	Operaciones y técnicas	Algoritmos	Aplicaciones
Clasificación		-Regresión logística (modelos lineales generalizados, GLM) -Árboles de decisión -Bayes naïf -Máquinas de vectores soporte (support vector machine, SVM)	-Técnica estadística clásica -Popular / reglas / transparencia -App embebida -Amplios / datos estrechos / texto
Regresión		-Regresión múltiple (GLM) -Máquinas de vectores soporte (SVM)	-Técnica estadística clásica -Amplios / datos estrechos / texto
Detección de anomalías		-SVM de una clase	Falta de ejemplos del campo objetivo
Importancia de atributos		-Longitud de descripción mínima (minimum description length, MDL)	-Reducción de atributos -Identificación de datos útiles -Reducción del ruido de los datos
Reglas de asociación		Apriori	-Análisis del cesto del mercado -Análisis de enlaces
Clustering (agrupación en racimos)		-Jerárquico K-media -Jerárquico O-cluster (orthogonal partitioning clustering, de Oracle)	-Agrupación de productos -Minería de textos -Análisis de genes y proteínas
Extracción de características		Factorización no negativa de matrices (non negative matrix factorization)	-Análisis de textos -Reducción de características

Figura 2.7: Funciones y algoritmos de técnicas de análisis de datos (Serrano-Cobos, 2014)

2.1.4 HERRAMIENTAS DE ANÁLISIS *Big Data*

Una vez que a sido seleccionada la técnica o el conjunto de técnicas a utilizar durante el desarrollo del proyecto se debe elegir el sistema que permitirá el procesamiento de la información, en este sentido existen una serie de software con diferentes características y paquetes pre-instalados útiles en el procesamiento de datos.

En este apartado revisaremos algunas de las herramientas mas populares actualmente utilizadas en proyectos *Big Data*.

Existen productos comerciales como MatLab que son escalables (permite su ejecución en clústers y nubes), sin embargo, desde la limpieza de datos hasta la visualización final los científicos de datos prefieren la utilización de lenguajes de software libre como R o Python (Arcila-Calderón *et al.*, 2016).

Existe una gran variedad de herramientas que contribuyen y facilitan el analisis de datos, mismas que cuentan con paquetes de mineria de datos, matematicos, estadisticos, bussines intelligence ente otros.

Muchas herramientas avanzadas para *Big Data* están disponibles como software de código abierto o comercial.

Bartschat *et al.* (2019) realizo una comparación entre las principales herramientas de código abierto gratuitas existentes, estas se presentan en la tabla 2.1.

Esta tabla (2.1) incluye nueve tipos diferentes de herramientas: software de minería de datos, paquetes de BI, paquetes matemáticos, paquetes de INT, extensiones, especialidades, prototipos de investigación, bibliotecas de minería de datos y soluciones.

La selección de la herramienta depende en gran medida de las necesidades del usuario, en nuestro caso, el costo fue uno de los criterios mas importantes por lo cual la herramienta seleccionada fue de la lista de herramientas de código abierto.

Herramienta	Tipo	Enlace
Apache Spark	MAT	https://spark.apache.org
Apache Flink	LIB	https://flink.apache.org
H2O	SPEC	www.h2o.ai
ImageJ	SOL	https://rsbweb.nih.gov/ij
ITK	SOL	www.itk.org
KEEL	INT	www.keel.es
KNIME	DAS	www.knime.org
Orange	DAS	https://orange.biolab.si
OpenCV	LIB	www.opencv.org
Pegasus	RES	www.cs.cmu.edu/pegasus
Pentaho	BI	www.pentaho.com
Python	MAT	www.python.org
R	MAT	www.r-project.org
RapidMiner	DAS	www.rapidminer.com
Scikit - learn	LIB	www.scikit-learn.org
TensorFlow	SPEC	www.tensorflow.org
WEKA	DAS, LIB	www.cs.waikato.ac.nz/ml/weka

Tabla 2.1: Herramientas de análisis de datos de código abierto

Si el usuario desea cubrir un conjunto grande de casos de uso diferentes, se recomiendan los paquetes MAT con una comunidad de usuarios grande como R o Python.

Las nuevas tendencias en el análisis de datos llevan al desarrollo de nuevas herramientas o prototipos. Un ejemplo reciente son las redes neuronales profundas que conducen al surgimiento de varias herramientas nuevas como TensorFlow o computación distribuida con herramientas como SPARK.

Por estas razones, el campo de las herramientas de *Big Data* cambia rápidamente y mantener una visión general es casi imposible. Por lo tanto, se recomienda centrarse en herramientas establecidas.

Tendencias como AutoML se esfuerzan por hacer que las tareas de rutina en *Big Data* sean fácilmente accesibles para los no expertos, junto con la descentralización y la conexión de dispositivos distribuidos con diferentes sistemas operativos, también la estandarizan y las interfaces serán más importantes.

En resumen, hay muchos campos de investigación activos y nuevos requisitos para *Big Data*.

Por lo tanto, el aumento de herramientas que permitan el análisis de la información continuará en el futuro y será tarea del científico de datos encontrar la herramienta adecuada para un caso de uso específico, esta continuara siendo una tarea desafiante, en nuestro caso la elección se baso en la accesibilidad, la facilidad de procesamiento y los paquetes mediante los cuales se pueden programar y realizar diferentes aplicaciones.

2.2 ANÁLISIS DE SERIES DE TIEMPO EN EL PRONÓSTICO DE LA DEMANDA

Tradicionalmente, los pronósticos dependen en gran medida de los datos históricos, el asesoramiento de expertos y la información del mercado. En la era de los grandes datos, tenemos cada vez más fuentes de información disponibles, que potencialmente pueden mejorar el rendimiento de los pronósticos (Choi *et al.*, 2018).

En este sentido la demanda futura tiene un papel muy importante para las organizaciones en la planificación de la producción y la gestión del inventario, para ello se necesitan hacer uso de los pronósticos.

En esta sección se ha incluido información que nos servirá para identificar, clasificar y determinar el tipo de serie de tiempo al que pueden pertenecer los datos recopilados a través de un periodo de tiempo determinado, además de definir el método de pronóstico más adecuado de acuerdo al comportamiento de los datos.

Para ello primero definiremos el concepto de serie tiempo y pronóstico, así como los tipos de pronósticos tradicionales en series de tiempo.

Los pronósticos son una herramienta de gran utilidad que proporciona una estimación cuantitativa de la probabilidad de que ocurra un evento en el futuro.

El principal propósito de un pronóstico es “reducir el rango de incertidumbre dentro del cual deben hacerse las estimaciones relacionadas con la administración” (Hanke y Wichern, 2006).

Los pronósticos posibilitan las decisiones sobre el futuro y proporcionan una evaluación del riesgo involucrado en la toma de decisiones, la importancia y utilidad de los pronósticos radica en que permite la reducción de la incertidumbre en la previsión de la demanda.

Si bien los pronósticos no son totalmente confiables si aportan una muy buena estimación y brindan un panorama mas cercano a la realidad lo que ayuda a disminuir el riesgo en la toma de decisiones gerenciales u operacionales.

Es importante que se haga uso de toda la información disponible actual y del pasado, y no solo basarse en supuestos, con el fin de evitar malas decisiones; también se debe considerar que los modelos de toma de decisiones se apegan a la realidad de la empresa y no a la del entorno .

Por otra parte para la formulación y aplicación de diversos métodos de pronostico se necesita hacer uso de las series de tiempo.

2.2.1 SERIES DE TIEMPO

Una serie de tiempo se define como una secuencia de observaciones, medidos en determinados momentos del tiempo, ordenados cronológicamente y, espaciados entre sí de manera uniforme, así los datos usualmente son dependientes entre sí.

El principal objetivo de una serie de tiempo X_t , donde : $t = 1, 2, \dots, n$ es estudiar y analizar el comportamiento de sus datos para poder realizar una correcta estimación del pronostico de la demanda. (Villavicencio, 2010).

Uno de los pasos mas importantes para estimar la previsión de la demanda, se lleva a cabo al momento de realizar la exploración de los datos.

La exploración de los datos puede llevarse a cabo mediante el análisis de tablas y gráficas ya que nos permiten apreciar observaciones inusuales y patrones únicos de manera visual.

Para elegir una técnica de pronósticos adecuada es necesario realizar la identificación de los patrones de datos que existe dentro de una serie de tiempo.

En estos casos puede elaborarse un pronóstico con un método de series de

tiempo o un método causal.

El objetivo de los métodos de pronóstico con series de tiempo es descubrir un patrón en los datos históricos y luego extrapolarlo hacia el futuro; el pronóstico se basa sólo en valores pasados de la variable que tratamos de pronosticar o en errores pasados.

En cambio, los métodos causales o cualitativos por lo general involucran el uso del juicio experto para elaborar pronósticos.

Una ventaja de los procedimientos cualitativos es que pueden aplicarse cuando la información sobre la variable que se está pronosticando no puede cuantificarse o son escasos y para ello se utilizan métodos tales como:

1. Método Delphi
2. Juicio experto
3. Redacción de escenarios
4. Enfoques intuitivos

En nuestro caso las variables bajo las que realizamos el pronóstico son del tipo cuantitativas y el análisis se realiza a partir de la extracción de datos históricos.

En el análisis de las series de tiempo, las mediciones pueden hacerse diariamente, semanalmente, mensualmente, anualmente o en cualquier otro intervalo regular de tiempo.

Para seleccionar un método de pronóstico adecuado de acuerdo a la serie de tiempo, se deben considerar las distintas clases de patrones de datos ya que este es uno de los aspectos mas importantes.

De acuerdo con Abril (2011) existen dos aspectos en el estudio de las series de tiempo: el análisis y el modelado.

El objetivo del análisis es resumir las propiedades de una serie y remarcar sus características.

En cambio en el modelado, no se realiza intento alguno para formular una relación de comportamiento entre la serie de tiempo considerada y otras variables.

El análisis clásico de las series temporales se basa en la suposición de que los valores que toma la variable de observación es la consecuencia de cuatro componentes, cuya actuación conjunta da como resultado los valores medidos, estos componentes son los que podemos observar en la figura 2.8.

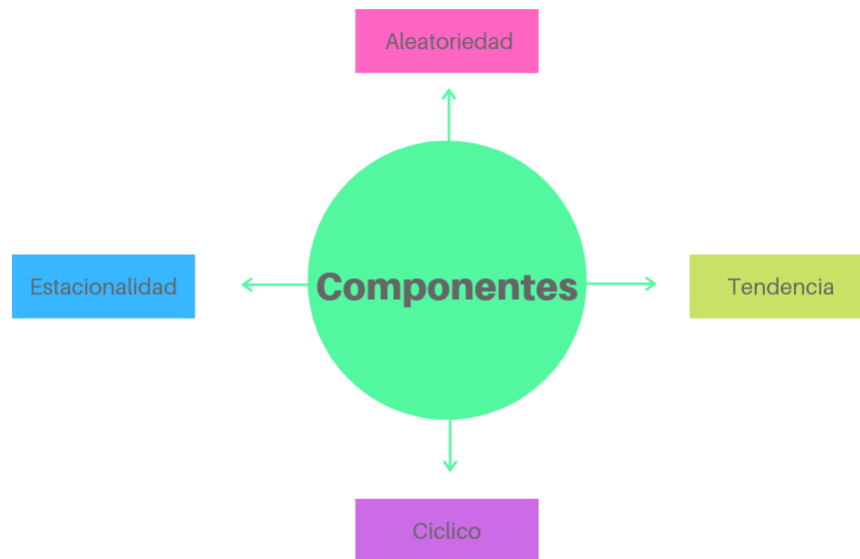


Figura 2.8: Componentes de las series de tiempo

- **Componente estacional:** una serie es estacionaria con características estacionarias cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo.

Esto se refleja gráficamente cuando los valores de la serie tienden a oscilar

alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.

La estacionalidad se presenta en muchas series temporales que presentan cierta periodicidad de forma anual.

- **Componente tendencia:** son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.
- **Componente aleatorio:** los datos de las series de tiempo suelen mostrar fluctuaciones, las series de tiempo también muestran un desplazamiento o movimiento gradual hacia valores relativamente altos o bajos a través de un lapso largo.
- **Componente cíclico:** cuando las series de tiempo presentan secuencias alternas de puntos abajo y arriba de la línea de tendencia se les denomina cíclicas.

Entre las principales ventajas que aporta el realizar los pronósticos mediante las series de tiempo es que solo se requiere conocer una cantidad limitada de datos para hacer pronóstico sin importar el horizonte de tiempo.

No requiere tener información de las variables exógenas que afectan la serie para su análisis, pero pueden ser univariadas o multivariadas.

Además las series de tiempo pueden ser estimadas más fácilmente que otros métodos de análisis de datos.

Los pronósticos basados en series de tiempo son bastante efectivos en el corto plazo (1 a 2 años) en comparación con otros métodos.

2.2.2 SERIES DE TIEMPO INTERMITENTES

Existe otro tipo de serie de tiempo, menos común a la cual se le denomina serie de tiempo intermitente.

Este tipo de series presentan un comportamiento que podrían confundirse con el de las series de tiempo con componente aleatorio.

Sin embargo de acuerdo a Griñón (2017) las series temporales intermitentes tienen un gran número de valores que son cero. Estos tipos de series ocurren comúnmente en la industria automotriz, aeroespacial y el sector retail o cuando el ciclo de vida de un producto se está acercando al final.

Las series de tiempo intermitente presentan un comportamiento como el que se muestra en la figura 2.10.

Como podemos observar presenta varios periodos con valor cero entre una demanda y otra.

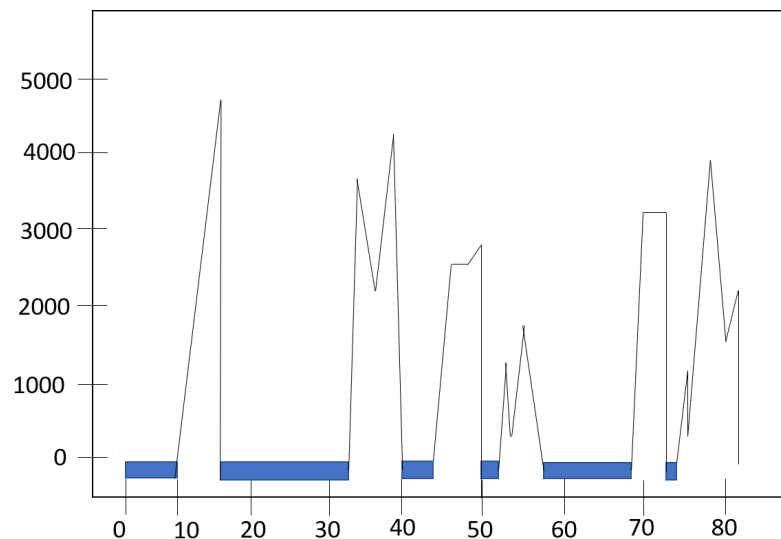


Figura 2.9: Serie de tiempo con comportamiento intermitente

De acuerdo con Rodríguez y Corrêa (2017) las series de tiempo intermitentes se pueden clasificar en cuatro grupos: i) errática, cuando el tamaño de presenta

elevada variabilidad; ii) intermitente, cuando la serie presenta varios valores nulos de la demanda; iii) granulada o irregular también conocida como lumpy cuando la variabilidad del tamaño de la demanda y los periodos entre dos demandas no nulas son altos; y iv) atenuada, cuando la variabilidad del tamaño de la demanda y el periodo entre dos demandas no nulas son bajas.

Ghobbar y Friend (2002) menciona que si se desea identificar el tipo de intermitencia que presenta la serie, deben ser calculados dos indicadores: un coeficiente de variación cuadrática (CV_2), y el promedio de intervalo entre demandas (ADI).

De acuerdo a los resultados obtenidos se puede categorizar de acuerdo con la figura 2.11.

Para cada caso es diferente debido al cálculo de los indicadores anteriormente mencionados, y se ubicarán de acuerdo a los cuadrantes mostrados en el esquema.

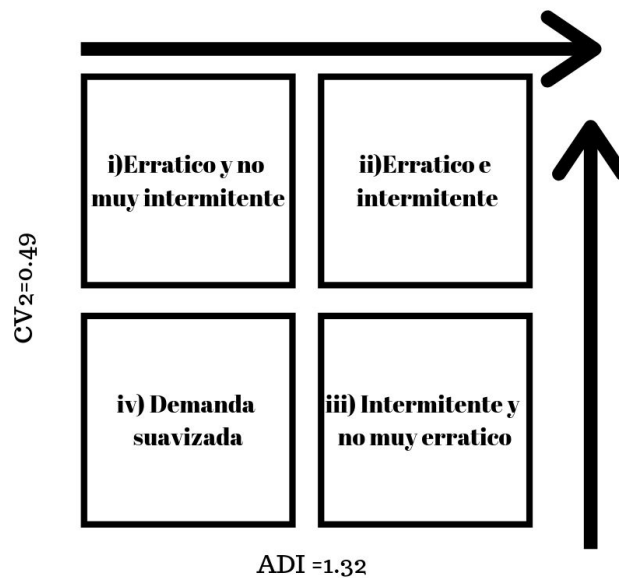


Figura 2.10: Clasificación de demanda intermitente

2.2.3 MÉTODOS DE PRONÓSTICO EN SERIES DE TIEMPO

Algunos de los principales métodos de pronóstico en las series de tiempo tradicionales se describen en el esquema 2.12.

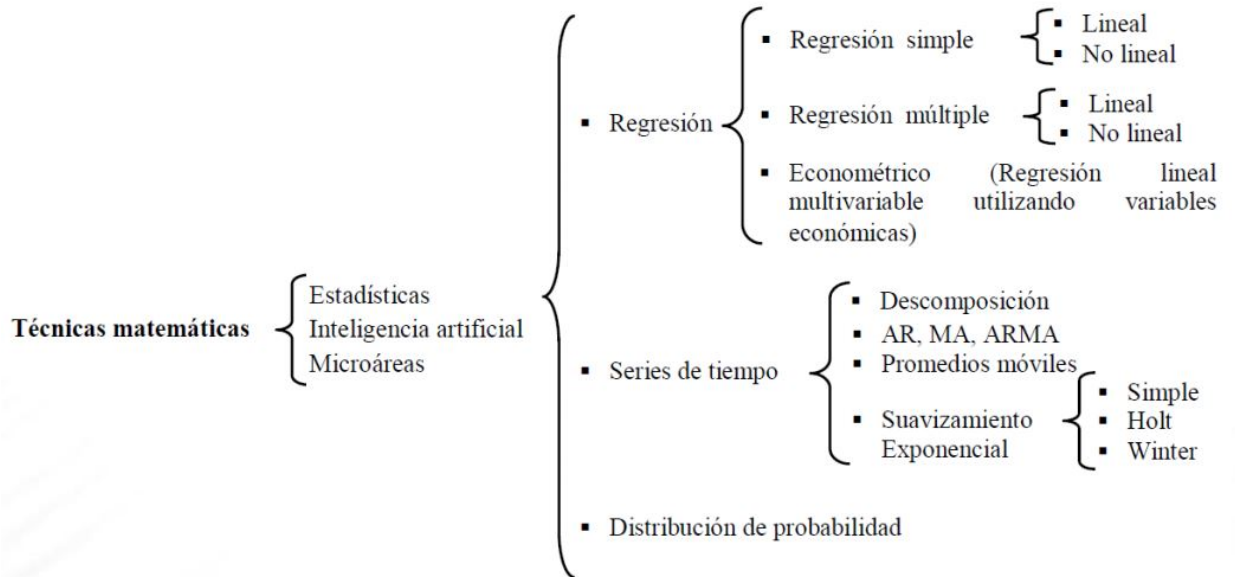


Figura 2.11: Métodos de solución y análisis en las series de tiempo

Anteriormente definimos a las series estacionarias como aquellas cuyo valor medio no cambia a lo largo del tiempo. En su forma mas simple, el pronóstico de una serie estacionaria implica utilizar la historia disponible de la serie para estimar su valor medio, lo que a su vez se convierte en el pronóstico de periodos futuros.

2.2.4 MÉTODOS DE SUAVIZACIÓN

En esta parte incluimos tres métodos de elaboración de pronósticos: promedios móviles, promedios móviles ponderados y suavización exponencial.

Este tipo de métodos lo que buscan es "suavizar" las fluctuaciones causadas por irregularidades en las series de tiempo de ahí es que se les conoce como métodos de suavización.

Consideraciones:

Este tipo de métodos tienen un buen rendimiento en series de tiempo estables, que no cuenta con efectos significativos de tendencia, cíclicos o estacionales.

Los métodos de suavización son sencillos de aplicar y proporcionan un alto nivel de precisión para pronósticos de corto alcance.

2.2.5 PROMEDIOS MÓVILES

El método de promedios móviles usa el promedio de k valores de datos más recientes en la serie de tiempo como el pronóstico para el siguiente periodo.

El término móvil indica que, mientras se dispone una nueva observación para la serie de tiempo, reemplaza a la observación más antigua de la observación anterior y se calcula un promedio nuevo.

Como resultado el promedio cambiará o se moverá conforme sigan nuevas observaciones en donde:

Y_t : observación en el periodo t

F_t : pronóstico para el periodo t

Este método emplea la siguiente fórmula dada (2.1).

$$F_{t-1} = \frac{Y_t + Y_{t+1} + \dots + Y_{t-k+1}}{k} \quad (2.1)$$

El pronóstico de promedio móvil es óptimo para patrones de demanda aleatorios o nivelados donde se pretende eliminar el impacto de los elementos irregulares históricos mediante un enfoque en períodos de demanda reciente.

2.2.6 MÉTODOS AUTO REGRESIVOS

Los métodos de pronóstico de regresión se engloban en: regresión simple, múltiple y económico multivariada sea lineal o no lineal en los tres casos.

El análisis de regresión es pertinente cuando se evidencia una tendencia en los datos históricos del pronóstico. ¿Cómo? simplemente analiza la tendencia de los datos, ¿es creciente o decreciente?

Para conocer la relación entre la demanda y el tiempo se hace uso del coeficiente de correlación en la literatura denominado como r

Este coeficiente, permite entender qué tanta correlación existe entre la demanda y el tiempo, existen tres tipos de correlación que son los siguientes:

Correlación perfecta: Cuando el resultado de coeficiente es igual a 1 o -1. En este caso existe una relación directamente proporcional entre la demanda y el tiempo.

Correlación fuerte: Cuando el resultado es mayor a 0.5 y menor que 1 (correlación positiva) o menor a -0.5 y mayor que -1 (correlación negativa).

Correlación débil: Valores que están entre -0.5 y 0.5. Entre más cercano se encuentre el coeficiente de correlación a +1 o -1 más fuerte será la tendencia y más apropiado será aplicar un modelo de regresión lineal. Por ejemplo si la correlación es igual a 1, observaremos que la relación entre las variables es directamente proporcional, en el sentido que si uno aumenta, la otra también lo hará.

Este método es más útil cuando se enfoca en periodos de demanda a largo plazo. Esto aunado a su utilidad para estimar la demanda en función de variables independientes.

Para el caso de las series de tiempo del tipo estacionales, una forma de desarrollar pronósticos implica la selección de un método de descomposición, multiplicativo o aditivo y enseguida la estimación de los índices estacionales de la serie.

Por lo cual las técnicas que deben considerarse al pronosticar este tipo de series incluyen los modelos de descomposición clásica, suavización exponencial de Winter, regresión múltiple y ARIMA.

Finalmente el efecto cíclico definido anteriormente como la oscilación alrededor de la tendencia. Para este tipo de series se recomienda utilizar los modelos de descomposición clásica, de regresión múltiple y ARIMA.

Siempre y cuando cumplan con las siguientes condiciones:

1. el ciclo del negocio influya en la variable de interés.
2. haya cambios en los gustos populares.
3. se den cambios en la población.
4. ocurran cambios en el ciclo de vida del producto.

2.2.7 MÉTODOS DE PRONOSTICO EN SERIES DE TIEMPO INTERMITENTES

Para el caso de las series de tiempo intermitentes de acuerdo a la revisión de la literatura los métodos de pronóstico tradicionales no se ajustan dadas sus características.

Algunos de los métodos de pronóstico recomendados son los siguientes:

El método de suavización exponencial simple es uno de los métodos que pueden ser aplicados a este tipo de series de tiempo.

Este método es apropiado para los datos que no tienen tendencia predecible hacia arriba o hacia abajo. El objetivo es estimar el nivel real y esta después es utilizada como el pronóstico de valores futuros.

Utiliza un factor de atenuación denominado alpha (α) que varia de $0 < (\alpha) < 1$.

En este caso si la variabilidad de la serie de tiempo es considerable, es preferible utilizar un valor pequeño para la constante de suavización. La razón de esta elección se debe gran parte del error de pronostico se debe a la variabilidad aleatoria y para no ajustar los pronósticos demasiado rápido elegimos el valor de (α) que minimiza el error del pronostico.

Este método emplea la siguiente formula dada (2.2).

$$F_t = F_{t-1} + \alpha(D_{t-1} - F_{t-1}) \quad (2.2)$$

Donde:

$\hat{F}_t + 1$: es el nuevo valor suavizado o el valor para el siguiente periodo t ;

α : es la constante de suavizamiento $0 < \alpha < 1$

Y : es la nueva observación o valor real de una serie en el periodo t

\hat{Y}_t : es el antiguo suavizado o pronóstico para el periodo t

El suavizamiento exponencial es un procedimiento para revisar de forma continua un pronóstico a la luz de experiencias mas recientes (Hanke y Wichern, 2006).

Por su parte (Croston, 1972) fue uno de los primeros autores que presento que el uso de los métodos clásicos como el de suavizacion exponencial generaba desvíos importantes en la previsión de demandas intermitentes.

El método de Croston utiliza como base el metodo de suavización exponencial simple pero a diferencia de este metodo, croston lo que hace es separar la serie de tiempo en dos partes: la primera, una serie con valores positivos de demanda, y la segunda, con los tiempos entre demandas consecutivas no nulas. En cada una de esas series se estima la previsión por medio de suavización exponencial y luego, ambos

valores son actualizados cuando existe un valor no nulo de demanda. En ambos casos se utiliza el mismo factor de atenuación denominado de igual manera como parámetro de atenuación alfa (α).

Croston notó un sesgo asociado al método de suavización exponencial simple ya que colocando el mayor peso en la fecha de la demanda más reciente. Para abordar este sesgo, Croston propuso un nuevo método para pronosticar bultos demanda, utilizando tanto el tamaño promedio de cero ocurrencias de demanda y el intervalo promedio entre tales ocurrencias (Gutierrez *et al.*, 2008).

El método de Croston emplea el siguiente modelo expresado en la figura 2.13

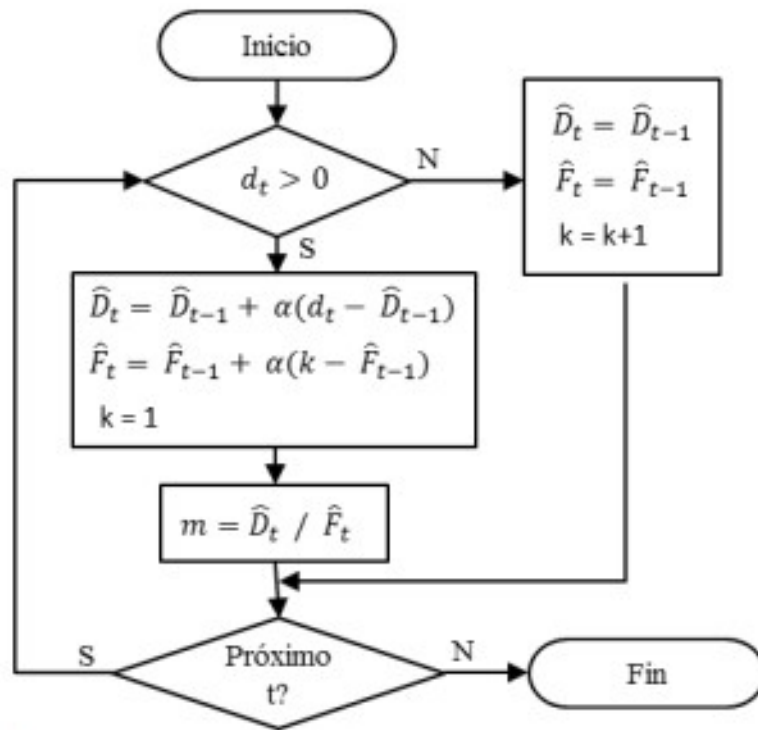


Figura 2.12: Diagrama de flujo método Croston

Donde:

d_t : demanda en el periodo t ;

D_t : Previsión de la demanda no nula para el período siguiente t ;

f_t : tiempo entre dos demandas no nulas;

F_t : previsión del intervalo de demanda;

k : intervalo desde la última demanda no nula;

α : parámetro de atenuación, $0 \leq \alpha \leq 1$;

m : demanda promedio en el período.

Así, la previsión para el próximo período es estimada como la razón entre D_t y F_t , y calculada mediante la expresión 2.2 :

$$m = \frac{D_t}{F_t} \quad (2.3)$$

Carlton (1969) presentaron un modelo basado en el método de Croston el cual sigue el mismo procedimiento pero a diferencia de este para realizar el calculo de la demanda media estimada debe realizarse el siguiente calculo dado de la expresión 2.3.

$$m = \left(1 - \frac{\alpha}{2}\right) \frac{D_t}{F_t} \quad (2.4)$$

2.2.8 EVALUACIÓN DE LOS MÉTODOS DE PRONÓSTICO

Existen diversas medidas de precisión para evaluar los diferentes métodos de pronóstico.

Estas medidas son: El MAD, MSE y MAPE.

Una forma de evaluar las técnicas de pronóstico utiliza la suma de los errores absolutos. En el caso de la desviación absoluta media (MAD) mide la precisión del pronóstico al promediar la magnitud de los errores de pronóstico, esta técnica es mas

útil cuando el analista requiere medir el error de pronóstico en las mismas unidades que la serie original.

La ecuación 2.1 muestra como se realiza el calculo del MAD.

$$MAD = \frac{1}{n} |Y_t - \hat{Y}_t| \quad (2.5)$$

El error cuadrático medio(MSE), es otro método para evaluar las técnicas de pronóstico, en este caso cada error de pronóstico o residual se eleva al cuadrado; luego se suman y se dividen el numero de observaciones. Este método penaliza los grandes errores de pronóstico debido a que los errores se elevan al cuadrado.

El MSE esta dado por la siguiente ecuación 2.2.

$$MSE = \frac{1}{n} (Y_t - \hat{Y}_t)^2 \quad (2.6)$$

Finalmente el error porcentual absoluto medio (MAPE) se calcula al encontrar el error absoluto en cada periodo, dividiendolo entre el valor real observado para ese periodo y luego promediando los errores porcentuales absolutos.

El MAPE se calcula mediante la ecuación 2.3.

$$MAPE = \frac{1}{n} \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (2.7)$$

La elección del método de pronóstico depende y requiere que produzca errores de pronóstico pequeños.

Una buena estrategia para evaluar los metodos de pronóstico acuerdo a describen los siguientes pasos:

- Se elige un método de pronóstico con base en el análisis y la intuición del pronosticador sobre la naturaleza de los datos.

- El conjunto de datos se divide en dos secciones: una de ajustes y otra de validación de la técnica de pronóstico.
- La técnica de pronóstico elegida se utiliza para obtener pronósticos de prueba o valores de ajuste con la primera porción de datos.
- Los pronósticos obtenidos se comparan con los datos a fin de evaluar el error de pronóstico.

El modelo de previsión más utilizado para patrones de demanda intermitente de acuerdo a la revisión de la literatura es el alisado exponencial simple. Sin embargo, este método suele subestimar sistemáticamente el valor de la demanda para cada periodo con demanda positiva; sobre estima el valor de la demanda media.

Sin embargo, ante un patrón de demanda intermitente, donde no se produce demanda en los $n-1$ periodos anteriores, la previsión dada para el periodo n será siempre inferior a la demanda real.

Con la información presentada en este capítulo se desarrollo la propuesta de una metodología en la que se incluyeron diversas técnicas de análisis de datos y técnicas para la estimación de los pronósticos, mismas con las cuales se desarrollo una herramienta adicional a las existentes en la literatura revisada para el caso de estudio y que cumple con las validaciones estadísticas apropiadas.

Dicha metodología así como su desarrollo sera presentada en el capítulo 3.

CAPÍTULO 3

METODOLOGÍA

3.1 DESARROLLO DE LA HERRAMIENTA

El objetivo principal del presente trabajo de investigación se centra en la creación de una metodología basada en técnicas *Big data*, a través de la cual desarrollar una herramienta mediante la cual sea posible realizar el análisis de información, de cualquier organización independientemente del tamaño y recursos con los que cuente.

Se pretende que, mediante su implementación, la empresa caso de estudio pueda mejorar su proceso de abastecimiento ya que con ayuda de los pronósticos es posible desarrollar estrategias que le permitan generar una ventaja competitiva.

La importancia del problema de investigación surge de la necesidad de dotar a las empresas con herramientas de código abierto y fácil acceso, que les permitan realizar sus propios análisis sin requerir de grandes inversiones económicas o de capacitación al personal.

Big data no es una moda, responde a una creciente necesidad de crear correlaciones y extraer información que, pueda ser de interés para mejorar la toma de decisiones estratégicas a través de el estudio de datos históricos.

Esta herramienta permite a las empresas entrar a la era de la inteligencia de

negocio e involucra directamente a la alta gerencia de las organizaciones además de que permite desarrollar un mejor conocimiento del cliente.

Como fundamento de este trabajo de investigación se realizó una extensa revisión de la literatura con la que se lograron reconocer los diferentes criterios que han sido tomados en cuenta para atender a esta problemática, además de conocer las diferentes técnicas o herramientas que se han desarrollado e implementado en las organizaciones para implementar proyectos big data.

El sistema fue diseñado basado en técnicas de análisis Data Mining, mediante la programación de algoritmos matemáticos y estadísticos que permite el realizar el pronóstico de series de tiempo en Visual Basic para Aplicaciones (VBA), incluido en Excel® 2010.

El desarrollo de la herramienta, fue basado en el proceso KDD ya que plantea todos los pasos a seguir para realizar la extracción del conocimiento, considera desde la extracción y manipulación de la base de datos hasta la creación de modelos y algoritmos para generar conocimiento.

En la siguiente figura se muestran los pasos considerados en el proceso KDD:

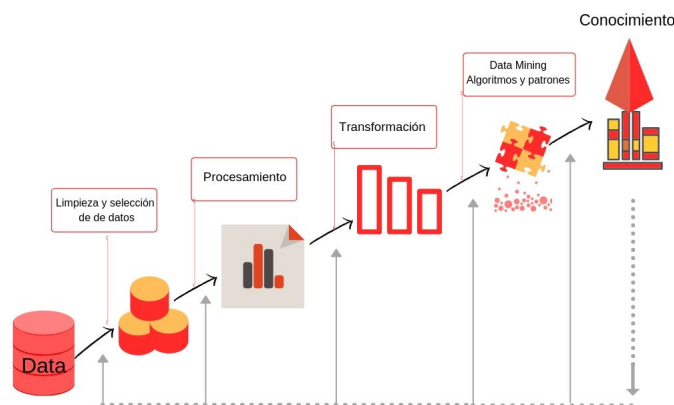


Figura 3.1: Elaboración propia

3.1.1 PSEUDOCÓDIGOS DE LA HERRAMIENTA

Esta metodología fue adaptada a nuestro caso de estudio, considerados todos los pasos del procesos KDD durante el desarrollo e implementación del proyecto y el caso de estudio.

La información así como la descripción de las actividades se sintetiza en el pseudocódigo mostrado a continuación:

Pseudocódigo General

Inicio

Entrada: base de datos

1. Recopilación de datos
2. Limpieza de datos
3. Integración de los datos*
4. Variables categóricas
5. Selección de datos de acuerdo a variables categóricas
6. Transformación de datos a serie de tiempo
7. Aplicación de método de pronóstico
8. Determinar porcentaje de error de pronóstico
9. Presentación y visualización de resultado

Salida: modelo optimizado de pronóstico de la serie

Fin

**Solo en el caso de necesitar integrar diferentes bases de datos en una sola es que se aplica este paso.*

Cada seudocódigo contiene los comandos y acciones realizadas para la creación y desarrollo de la herramienta mismos que fueron considerados durante la implementación.

Recopilación de datos

Inicio: 1.-Ingrese al sistema 2.-Acceda al módulo de interés 3.-Seleccione el período para analizar 4.-Descargar datos en formato CSV

Esta fase es la que describe el procedimiento, para obtener la base de datos a analizar y cargar que en nuestra herramienta. Se describen los pasos a seguir para realizar la extracción de la base de datos específicamente para nuestro caso de estudio.

Una vez que se ha accedido a la plataforma, se procede a consultar el modulo de interés, ya sea compras, ventas, clientes, servicios, entre otros de acuerdo a lo que se desee analizar. Al realizar la descarga de la base de datos debe seleccionar el periodo de tiempo que desea analizar (año, mes día).

Finalmente una vez seleccionado el modulo y el periodo de tiempo, se debe descargar la base de datos en formato CVS.

Seudocódigo de limpieza de datos

Entrada: Base de datos 1.-Cargar base de datos a herramienta VBA 2.-Ordenar datos lexicográficamente 3.-Identificar los datos que deben eliminarse: Duplicados Cancelados 4.-Unificar nomenclatura 5.-Eliminar datos incongruentes

Salida: base de datos depurada Fin

Se debe cargar la base de datos que se desea analizar a la herramienta VBA en la sección -Hoja1-. Antes de realizar la limpieza de datos ordene las filas lexicográficamente. Una vez ordenadas las filas, sera mas fácil identificar visualmente los datos que deben eliminarse. En la herramienta se encuentra el botón limpieza con el cual podrá y deberá:eliminar duplicados y unificar nomenclatura. Consideramos

datos incongruentes a todos aquellos que se encuentran contenidos en la base de datos y que, por necesidad del sistema (ERP de la empresa) están contenidos en la base de datos pero que, NO provén información, por ejemplo: en nuestro caso de estudio, cuando se abre una orden y se comete un error ya sea en precio, nombre del cliente o tipo de servicio registrado no es posible eliminarla, por lo cual se cancela y se almacena en el sistema. Otro tipo de error común son los errores tipográficos o en los que la información este incompleta es decir, que contenga filas sin llenar, lo que vuelve ese en particular invalido

Seudocódigo integración de datos

Inicio

Entrada: Nueva tabla

1.-Concatenar a base de datos una nueva tabla o base de datos 2.-Asegurar la integridad de la tabla 3.-Unificar categorías

Salida: Base de datos unificada

Fin

Este paso, solo debe ser realizado si se desea agregar una tabla o base de datos adicional a la ya programada. Se deberá concatenar la nueva base de datos o tabla asegurándose que las columnas concuerden con los encabezados. El aseguramiento de la integridad de los datos hace referencia a que si la base de datos que ha sido concatenada no ah pasado por el procedimiento de limpieza, debera realizarlo.

Seudocódigo Definir variables categoricas

Inicio

Entrada: tabla de datos

1.-Seleccionar categorías relevantes 2.-Asignar nivel de importancia a las categorías aplicar métodos ABC, y Pareto 3.-Analizar/medir nivel de influencia de las

categorías en las variables de respuesta

Salida: tabla de datos filtrada

Fin

Es importante resaltar que la clasificación al igual como el nivel de importancia dado será establecido por el experto del proceso que es analizado. El experto en el proceso a analizar debe categorizar y dar orden de importancia a las secciones establecidas en la base de datos. Después se deberá medir el nivel de influencia de las secciones en las variables de respuesta, este nivel será dado mediante los resultados dados en la clasificación abc y el diagrama de Pareto. Este paso es muy importante, debido a que normalmente cuando se tiene un compendio de información a analizar no se sabe por donde empezar o que analizar, a menos de que ya se tenga un criterio o problemática en concreto a la cual dar solución.

Seudocódigo Selección de datos de acuerdo a variable categórica

Inicio

Entrada: Base de datos

1.-Definir prioridad de criterio de selección (cualitativo/cuantitativo) 2.-Identificar los propósitos que los datos podrían cumplir 3.-Definir periodo a analizar

Salida: Base de datos con criterio seleccionado a analizar

Fin

De acuerdo a los criterios definidos por el experto, se selecciona el ítem a analizar.

La selección de los datos dependerá de lo que se desea analizar en el momento, conocimiento de

los clientes, compras, ventas, entre otros.

Seudocódigo transformación de datos a serie de tiempo

Inicio

Entrada: Tabla de datos y periodo de tiempo significativo

1.-Agrupación de ítems por periodo de tiempo 2.-Selección de periodo de tiempo a pronosticar 3.-Descomponer fechas 4.-Extraer intervalos de tiempo nulos 5.-Transformar periodo a serie de tiempo 6.-Clasificar serie de tiempo 6.-Identificar si presenta intermitencia 7.-Clasificar tipo de Intermitencia

Salida: Serie de tiempo clasificada

Fin

Una vez que se conoce el producto, servicio, cliente... que se desea analizar o pronosticar, debe agruparlo por fechas. Una vez que ha agrupado por fechas el periodo de tiempo que desea analizar, procede a realizar la descomposición de fechas, es decir, dividir las para encontrar los periodos de tiempo nulos. Una vez que realice la descomposición de la fecha inserte cero donde no se encuentre coincidencia para extraer la serie de tiempo. Después de extraer la serie de tiempo se procede a identificar que tipo de serie es, en el caso de identificarse como serie de tiempo intermitente debe clasificar que tipo de intermitencia presenta.

Seudocódigo aplicación de método de pronostico

Inicio

Entrada: Serie de tiempo clasificada

1.-Seleccionar ítem con intermitencia 2.-Seleccionar periodo a pronosticar 3.-Aplicar métodos de pronóstico a ítem en el periodo seleccionado

Salida: Pronóstico de serie de tiempo

Fin

Una vez que se ha identificado y clasificado la serie de tiempo de acuerdo a sus características se elije el método de pronóstico apropiado para la serie, en el caso de ser una serie de tiempo de tipo intermitente se aplican los métodos de pronóstico: Croston, SBJ, SBA y suavización exponencial simple. En caso de ser una serie de tiempo de otras características se aplica uno de los métodos de pronóstico que mejor ajuste a la serie de acuerdo a la revisión de la literatura.

Seudocódigo error de pronostico

Entrada: Pronóstico de Serie de tiempo

Inicio

- 1.-Aplique Montecarlo a Alpha para iterar de 0 a 1 entre 1000
- 2.-Para cada Alpha generar un pronóstico
- 3.-Calcular el porcentaje de error de cada alpha
- 4.-Seleccionar el Alpha con menor error
- 5.-Repetir procedimiento para cada método de pronostico

Fin

Salida: Resultados de pronóstico con calculo de error optimizado

Una vez que se a realizado el pronostico se debe determinar el porcentaje de error con la finalidad de identificar cual es el que presenta un mejor resultado. Para encontrar el alpha optimizado ya que se debe utilizar un alpha entre los valores 0 y 1 se aplica el método monte carlo y se varia entre 1000. Este procedimiento debera ser realizado para cada método de pronostico aplicado.

Seudocódigo Presentación y visualización de resultados

Inicio

Entrada: Resultados de pronóstico con calculo de error optimizado

1.-Realizar comparación entre los métodos de pronóstico en el periodo establecido con el Alpha de mejor desempeño asignado 2.-Comparar margen de error entre los diferentes métodos de pronóstico 3.-Determinar cual presenta el mejor desempeño

Salida: Pronóstico de serie de tiempo optimizado

Fin

Como ultimo paso, para obtener la serie pronosticada así como visualizar el pronóstico se deberá graficar el pronóstico obtenido. Se realizara la comparación entre los diferentes métodos de pronóstico y sera elegido el que presente mejor desempeño.

La elección y programación en la herramienta de los métodos de pronóstico fue realizada con base al comportamiento de los datos y las características que presentaban las series de tiempo.

De acuerdo al pseudocódigo general se exponen a continuación los códigos para cada paso de manera general.

El procedimiento, implementación y resultados del caso de estudio serán expuestos en el capitulo de cuatro.

CAPÍTULO 4

CASO DE ESTUDIO

Como parte de esta investigación, se probará la herramienta con la finalidad de verificar su funcionalidad así como el grado de efectividad en la limpieza, selección, análisis de datos y elaboración de pronósticos.

En este capítulo se da un panorama general de la situación de la empresa elegida como caso de estudio, será descrita cada etapa del acercamiento e implementación de la herramienta para probar la efectividad de la misma, así como comprobar el impacto en los procesos logísticos de la organización, específicamente en el proceso de abastecimiento y el departamento de compras.

Por políticas de confidencialidad será omitido el nombre de la empresa, sin embargo esta brinda todas las facilidades y el apoyo para la consecución de este proyecto.

4.1 INFORMACIÓN DE LA EMPRESA

El caso de estudio se desarrolla en una empresa de carácter privado, dedicada al servicio de mantenimiento preventivo, correctivo y rescates en carretera de transporte pesado.

Se encuentra ubicada en el bajío de México, zona estratégica ya que además de colindar con 8 estados de la república, es cruce obligatorio para las unidades que llevan la ruta CDMX-Nuevo Laredo como se muestra en la Figura 4.1.

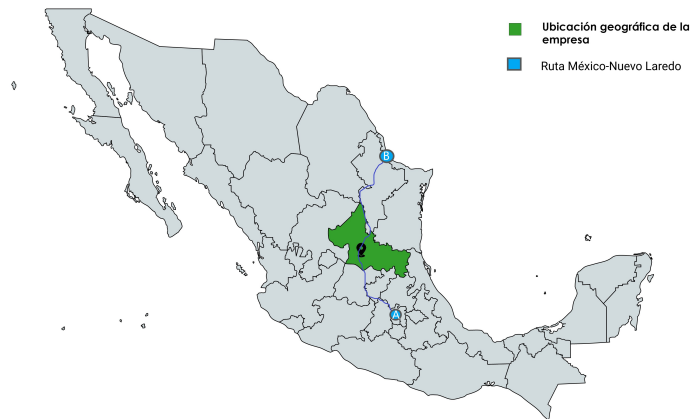


Figura 4.1: Ubicación geográfica de la empresa

La empresa, nació como una empresa familiar fue constituida en 1989 con la firme convicción de dar el mejor servicio de mantenimiento a la industria del transporte de carga, en San Luis Potosí.

La empresa inicia sus operaciones como un pequeño taller de mecánica diesel teniendo como instalaciones un pequeño terreno rentado y un pequeño almacén de herramientas y el equipo de trabajo estaba conformado por solo cuatro empleados.

En el año 2004 adquiere el terreno donde actualmente se ubica, manteniendo año con año un crecimiento aproximado al 10%, se construyen las primeras oficinas y la nave de servicios para ofrecer una mejor atención a los clientes.

A partir de 2012 se toma la decisión de cambiar su razón social y crea el logotipo que la distingue, desde entonces a la fecha ha venido modificando su modelo de negocio, las instalaciones y sus servicios lo que le ha permitido posicionarse y aumentar la cartera de clientes.

La empresa actualmente esta conformada por 30 trabajadores entre técnicos,

mecánicos y personal administrativo, cuenta con dos sucursales en la ciudad de San Luis Potosí y una en la ciudad de Zacatecas, por la cantidad de trabajadores así como los ingresos que mantiene se clasifica como una empresa pequeña.

Los servicios que actualmente presta la empresa son los siguientes:

- Mantenimiento preventivo y correctivo de unidades diesel
- Reparación de motores diesel.
- Reparación al tren motriz (transmisiones y diferenciales).
- Diagnostico por computadora.
- Muelles, suspensiones y soldadura.
- Reparaciones menores; sistema de frenos de aire, mantenimiento a remolques y cualquier tipo de fallas mecánicas en unidades.
- Rescates en carretera.
- Administración de garantías a motores Cummins.

4.2 DESCRIPCIÓN DEL PROCESO DE ATENCIÓN

El proceso para realizar el servicio de mantenimiento a camiones y tracto camiones comienza con la solicitud del servicio, que es atendida por el Gerente de Servicios el cual se encarga de llenar la O.T (Orden de Trabajo) con la información de la unidad, el tipo de servicio que se le brindara y la empresa transportista a la que pertenece; posteriormente se da ingreso a la unidad a la nave de servicios.

Una vez que la unidad ingresa a la nave uno de los técnicos es asignado a la reparación y procede a realizar una inspección y evaluación para determinar que tipo de mantenimiento necesita o determinar el tipo de falla que presenta, esto para

informar al cliente el estado de la unidad y que este, posteriormente autorice el servicio, y finalmente canalizar la unidad con el técnico adecuado según la falla que presente.

Una vez que determina el tipo de servicio a realizar o la falla se procede a solicitar las refacciones al área de almacén, en caso de que el almacén no cuente con las refacciones necesarias el comprador abre una orden de compra y la envía al proveedor.

Una vez concluido el servicio, el técnico informa al jefe de taller para registrar el alta y salida de la unidad.

El proceso de atención se muestra gráficamente en el siguiente diagrama de flujo.

4.3 EXPLORACIÓN DE LA PROBLEMÁTICA Y ÁREAS DE OPORTUNIDAD

Desde que inicio sus operaciones hasta el año 2015 la empresa no contaba con ningún sistema de información para el registro y respaldo de sus operaciones, unicamente se llevaba el registro de las ordenes de trabajo en Access.

Al ser una empresa familiar todas la operaciones eran realizadas con base al "*know how*", el almacén contaba con escasas refacciones de algunas partes o un exceso de otras, no se tenia un control sobre las compras, los pedidos se realizaban al día, hasta el 2016 la empresa había estado funcionado de esta manera, pero al paso de los años y debido al aumento en su cartera de clientes, es que comenzaron a tener quejas en cuanto a la entrega de unidades en tiempo y forma, ademas el almacén comenzó a llenarse de refacciones debido a que duplicaban las compras.

Para aminorar los problemas administrativos y comenzar a llevar un mayor y





















Actividad	Símbolo	Descripción de la Actividad
1		Llenar orden de servicio y dar entrada a la unidad
2		
2		Realizar Diagnostico previo a la unidad
3		
3		Informar al jefe de taller de servicio
4		
4		Cargar información a la plataforma
5		Ingresar la unidad a la nave de servicio
6		
6		Asignar técnico a la unidad para dar atención
7		
7		Ejecutar inspección visual y técnica a la unidad
8		
8		Realizar servicio a la unidad
9		
9		Realizar servicio a la unidad
10		Solicitar refacciones a almacen
11		Solicitar refacciones a compras
12		
12		Autorizacion de salida de unidad

Figura 4.2: Diagrama de Flujo del proceso de atencion de unidades

mejor control sobre las operaciones deciden implementar un sistema ERP hecho a la medida, a través del cual comenzaron a registrar las ordenes de servicio, compras, entradas y salidas de almacén, registro de clientes nomina y empleados.

Con el sistema en funcionamiento se comenzó a guardar y recolectar mas información de la empresa, sin embargo, no habían sido capaces de generar ningún tipo de informe o análisis, para saber el estado de sus operaciones.

Después de un año, pesar de que la administración ya tenia dominio del sistema y se habían corregido algunos de los problemas en cuanto al control y registro de las operaciones derivados de la implementación del sistema ERP, las quejas del cliente en cuanto a los tiempos de entrega continuaban, la mayoría de los retrasos debido a que los mecánicos no contaban con las refacciones necesarias para la realización de los servicios en tiempo y forma.

Por sus características fue elegida esta empresa para fungir como caso de estudio, se tomo en cuenta las opiniones del personal y se nos concedió acceso total a sus bases de datos, partimos de la problemática presentada pero no sesgamos nuestra investigación y dejamos que los datos hablaran por si mismos.

4.4 PAUTAS PARA EL DESARROLLO DE LA METODOLOGÍA Y LA APLICACIÓN DE LA HERRAMIENTA EN LA EMPRESA

El primer paso a seguir es la extraccion de la informacion es decir contar con los datos o la base de datos que se desea analizar, en nuestro caso de estudio especificamente se descargo la información del sistema ERP de la empresa.

Precedimos a acceder al sistema, para después entrar al modulo que se deseaba analizar, seleccionamos el periodo de tiempo que deseábamos analizar para después

descargar esta información en un formato excel CSV.

Es importante aclarar que es requisito que la base de datos o registro que se pretenda analizar debe estar en formato de hoja de calculo excel ya que nuestra herramienta fue diseñada para funcionar en este sistema, bajo el lenguaje de programación VBA.

Las empresas que no cuenten con un sistema ERP pero que al menos lleven un registro de sus operaciones en una hoja de calculo excel tambien podrán realiza el análisis de la información siempre y cuando cada columna cuente con un encabezado el cual identifique o mencione el contenido de la fila y como contenido o información mínima debe contar con los siguientes datos:

- tipo de producto, servicio, refacción etc.
- fecha de compra/venta segun sea el caso.

en nuestro caso la información se presenta de la siguiente manera de acuerdo a la figura 4.3, para su ejemplificación se omitieron datos claves de la empresa pero se resalta el contenido de la descripción/nombre de la refacción comprada en este caso y la fecha en la que fue solicitada.

4.5 FORMULACIÓN DE LOS CRITERIOS DE DECISIÓN

Teniendo la base de datos en el sistema se procedió a realizar la limpieza de datos de acuerdo a los criterios definidos en el algoritmo de limpieza de datos descrito en el capitulo de metodología.

Una parte importante en esta investigación fue la definición de criterios, ya que fue la que nos dio la pauta para la realización del análisis de la información, el cruce de información y la selección de los datos.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Nombre Proveedor	Autorizado	Abierta	Desc. Breve	Fecha Orden	Observacion	Activo Fijo	Moneda	Tasa Descue	Tasa Iva	Otros Impue	Compra	Recibe	Autorizadod	Comentari
2			SI	KIT DE FRENOS.	13/09/2016										
3			SI	TAMBOR	09/11/2016										
4			SI	BOMBA DE AGUA	09/12/2016										
5			SI	BOMBA DE ACEITE	15/12/2016										
6			SI	Matraca fina	15/12/2016										
7			SI	ACTUADOR VGT	21/12/2016										
8			SI	ARBOL DE LEVAS	21/12/2016										
9			SI	sin descripcion	28/12/2016										
10			NO	GRASA DE CHASIS	30/12/2016										
11			SI	sin descripcion	03/01/2017										
12			SI	sin descripcion	03/01/2017										
13			SI	sin descripcion	03/01/2017										
14			SI	FILTRO	09/01/2017										
15			SI	FILTRO	10/01/2017										
16			SI	FILTRO	16/01/2017										
17			SI	FILTRO	17/01/2017										
18			SI	EXTERNO	18/01/2017										
19			SI	INYECTOR	18/01/2017										
20			SI	sin descripcion	18/01/2017										
21			SI	sin descripcion	20/01/2017										
22			SI	ANTICONGELANTE	28/01/2017										
23			SI	ACTUADOR VGT	21/01/2017										

Figura 4.3: Ejemplo de datos mínimos para análisis de información de la herramienta

Como lo comentamos anteriormente la empresa tenia problemas en la entrega de unidades en tiempo al cliente. Se analizaron 100 ordenes de servicio diferentes de las cuales se detecto que en los mantenimientos del tipo A, que la empresa tiene definido un tiempo máximo de dos horas para ser realizado tardaban un promedio de hasta 6 horas en ser realizado.

Debido principalmente, a la falta de refacciones y en segundo lugar a la falta de personal, lo que representaba tres veces mas el estimado de tiempo para la realización del mantenimiento.

En consecuencia, el primer modulo que fue analizado fue la base de datos referente a los mantenimientos y servicios realizados versus las refacciones adquiridas.

Antes de realizar el cruce y análisis de la información, definimos los criterios, que de acuerdo con el gerente de compras, fueron el tiempo y el costo de las refacciones por lo cual el análisis se realizo con base a esos dos criterios, es decir que al realizar el análisis de la información se busco la relación que había entre el tiempo de entrega de las reacciones, el costo y la disponibilidad de las mismas.

En el proceso de análisis de acuerdo a los criterios establecidos se encontra-

ron cuatro productos principales acordes a estas características los cuales fueron seleccionados para realizar la experimentación.

4.6 ANÁLISIS DE RESULTADOS

De acuerdo con el análisis de la información realizado, al extraer las series de tiempo se procedió a realizar el análisis del comportamiento de las mismas, y de acuerdo con las características que presentaba se determinó que los 4 productos analizados de acuerdo a los criterios de selección establecidos por la empresa presentaban un comportamiento del tipo intermitente, que como fue definido en la sección de antecedentes una serie intermitente es aquella que a lo largo del tiempo presenta valores cero entre demandas no nulas.

Los datos se presentan en las tablas 4.1 muestran el comportamiento del producto 1 en los años 2017 y 2018 correspondientemente.

Año	Mes	Serie Producto #1	Año	Mes	Serie Producto #1
2017	Enero	0	2018	Enero	1
2017	Febrero	0	2018	Febrero	1
2017	Marzo	0	2018	Marzo	0
2017	Abril	1	2018	Abril	1
2017	Mayo	0	2018	Mayo	0
2017	Junio	1	2018	Junio	3
2017	Julio	0	2018	Julio	0
2017	Agosto	2	2018	Agosto	3
2017	Septiembre	0	2018	Septiembre	0
2017	Octubre	0	2018	Octubre	3
2017	Noviembre	1	2018	Noviembre	1
2017	Diciembre	0	2018	Diciembre	2

Tabla 4.1: Datos del producto 1.

La representación del comportamiento gráfico del producto 1 se muestra en la figura 4.4.

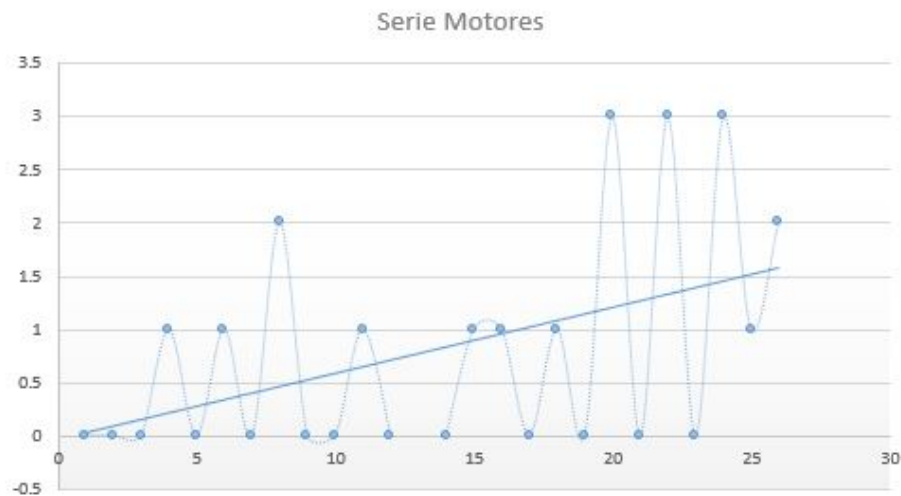


Figura 4.4: Gráfico de la demanda del producto 1

Como podemos observar en la gráfica los datos no presentan tendencia, por lo cual entra en la clasificación de serie de tiempo del tipo intermitente.

Se analizaron las compras realizadas en un periodo de dos años y el horizonte de la demanda se realiza a un año.

Una vez que se observó el comportamiento de los datos y se determinó que tenían un comportamiento intermitente se procedió a clasificar de acuerdo a los resultados obtenidos al calcular el ADI Y CV^2 como se muestra en la siguiente tabla:

Producto	ADI	CV ²
1	0.0833333	2.581924
2	0.83	2.489142
3	0.2	3.6

4	0.3333	2.5854
---	--------	--------

De acuerdo a Ghobbar y Friend (2002) se clasifican de la siguiente manera:

De acuerdo a los cálculos del ADI y el CV^2 los productos tendrían la siguiente clasificación o grado de intermitencia:

Producto	ADI	CV2	TIPO
1	0.0833333	2.581924	iii
2	0.83	2.489142	iii
3	0.2	3.6	iii
4	0.3333	2.5854	iii

Los cuatro productos se clasifican como intermitentes y no muy erráticos.

Después de clasificar el tipo de intermitencia se procedió a evaluar como el método de pronóstico a aplicar, de acuerdo a la revisión de la literatura los métodos de pronóstico que se ajustan mejor a este tipo de series son el método de SES, CROSTON, SBJ y SBA.

Los métodos de pronóstico seleccionados fueron aplicados y para el producto 1 se obtuvieron los siguientes resultados:

Método	Porcentaje de error	Valor de alpha utilizado
SES	188 %	0.82

SBA	138 %	0.75
SBJ	174 %	0.51
CROSTON	114 %	0.48

Como podemos observar el método de Croston presenta ligeramente mejores resultados en comparación con los otros métodos de pronóstico, sin embargo nos dimos cuenta que el porcentaje de error sobre el pronóstico presentaba valores muy altos, revisamos en la literatura y no existía una fuente en la cual te recomendaran el valor de alpha que debía ser usado para lograr mejores resultados, unicamente mostraban los rangos en los que se podía obtener un mejor rendimiento.

Fue entonces que se opto por utilizar el método de simulación monte carlo para variar los alphas en diferentes valores hasta encontrar el que tenia un menor porcentaje de error y los resultados obtenidos son los siguientes:

Método	Porcentaje de error	Valor de alpha utilizado
SES	92 %	0.644557
SBA	80 %	0.64653546
SBJ	127 %	0.471746
CROSTON	24 %	0.9883279

Se realizo el mismo procedimiento para los otros productos y la herramienta arrojó los siguientes resultados:

Para el producto #2:

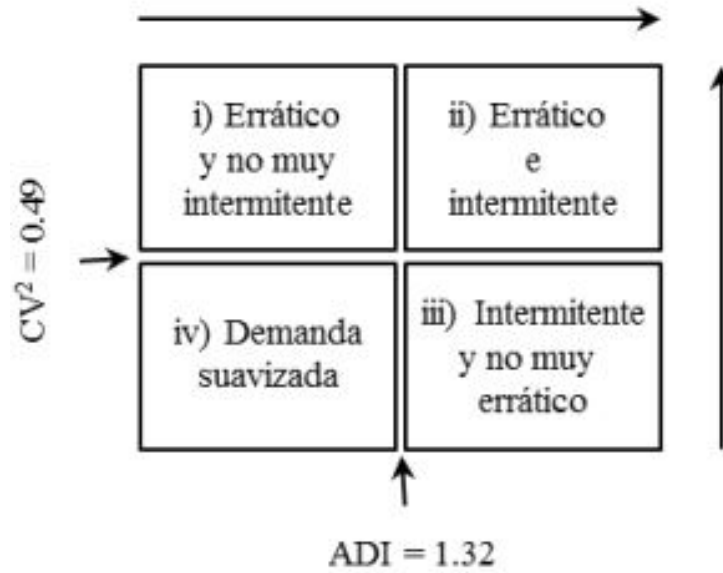


Figura 4.5: Clasificación de intermitencia

Método	Porcentaje de error	Valor de alpha utilizado
SES	66 %	0.8556
SBA	43 %	0.6127
SBJ	46 %	0.4172
CROSTON	40 %	0.6364

Para el producto #3:

Método	Porcentaje de error	Valor de alpha utilizado
SES	99 %	0.1876
SBA	79 %	0.9553
SBJ	91 %	0.6089
CROSTON	70 %	0.9994

Para el producto #3:

Método	Porcentaje de error	Valor de alpha utilizado
--------	---------------------	--------------------------

Método	Porcentaje de error	Valor de alpha utilizado
SES	57 %	0.2395
SBA	66 %	0.9966
SBJ	70 %	0.7178
CROSTON	64 %	0.9994

Como podemos observar Croston continua presentando un mejor desempeño en comparación con los otros métodos de pronóstico además que al utilizar el alpha de la simulación monte carlo se logro reducir el porcentaje de error, se realizo el calculo del error promedio utilizando el metodo de croston dando un resultado de 56 %.

CAPÍTULO 5

CONCLUSIONES

El capítulo que se expone a continuación contiene lo referente a las resoluciones finales que surgieron del estudio de la temática, así como del análisis de los resultados derivadas de la experimentación con la herramienta propuesta.

De manera adicional se exponen algunas recomendaciones generales para empresa y las contribuciones derivadas de la investigación.

5.1 CONCLUSIONES GENERALES

El presente trabajo representa un apoyo para la mejora en la toma de decisiones en la organización. Las empresas requieren hacer hincapié en el fortalecimiento y construcción del análisis de información y en la toma de decisiones con base en los datos.

En este caso se buscó que el análisis de la información tuviera impacto en las operaciones logísticas de la empresa para ser exactos en el procedimiento de abastecimiento de refacciones, almacén y compras para mejorar el tiempo de respuesta y entrega de unidades al cliente ya que como se comentó en el capítulo cuatro la empresa presentaba reportes de quejas por parte de los clientes en cuanto a los tiempos de entrega de unidades.

Al analizar las ordenes de servicio se detecto que una de las causas era la falta de refacciones en almacén por lo cual la propuesta de mejora fue para atacar esta problemática.

La empresa caso de estudio es parte de la cadena de suministro del sector auto transporte, este sector exige agilidad y rapidez en la atención de sus unidades. Ya que las empresas desean no solo para cumplir en tiempo y forma con las entregas de mercancía a sus clientes si no también mantener la menor cantidad de tiempo posible detenidas las unidades en el taller de servicio.

Para las empresas de transporte mantener sus unidades en taller se traduce en tiempo improductivo y perdidas, por lo cual el servicio que presta la empresa caso de estudio es de gran valor e importancia dentro de la cadena de suministros, brindar un servicio de calidad, ágil y en tiempo permite reducir demoras, paros y atrasos en la cadena de abastecimiento.

Con la reducción en los tiempos de servicio y entrega de las unidades derivadas de un eficiente abastecimiento la empresa lograra reducir las quejas por incumplimiento, brindándole a su vez la posibilidad de crear mejores acuerdos comerciales con los proveedores que le permitan reducir los costos al comprar en volumen, mantener e incrementar su cartera de clientes y al reducir los costos por abastecimiento poder invertir en nuevos proyectos o ampliaciones en la nave lo que representaría un aumento en su capacidad instalada.

La metodología propuesta contribuye al desarrollo de proyectos de minería de datos y en nuestro caso específico presenta los pasos a seguir para el desarrollo de una herramienta a través de la cual es posible analizar los datos y generar pronósticos de los servicios y de las refacciones requeridas para la realización de los mismos, que al ser implementada podrá mejorar el desempeño logístico de la organización.

Como se puede observar por medio de esta aplicación, el proceso de minería de datos ha demostrado ser una herramienta efectiva en el desarrollo de proyectos para extraer conocimiento de las bases de datos, contribuyendo a la mejora en la toma de

decisiones.

5.2 RECOMENDACIONES

Un punto a considerar para que este tipo de proyectos y herramientas pueden ser desarrollados por cualquier organización siempre y cuando cuente con los registros de sus operaciones en formato electrónico, es que se debe poseer o tener nociones previas de técnicas minería de datos o de alguna herramienta de análisis de información.

En particular para nuestro caso de estudio, gracias a esta investigación fue posible detectar una falla en el sistema ERP de la empresa ya que permite que el usuario a la hora de realizar el registro y creación de ordenes de servicio, compra o alta de clientes pueda dejar secciones en blanco, secciones que por comodidad el usuario no registra y por consiguiente se pierde información valiosa para la operación, además de que en muchos campos se puede agilizar y sistematizar el registro de información, estas observaciones fueron expuestas a la empresa para su consideración.

5.3 CONTRIBUCIONES

La aportación de este trabajo se centra principalmente en brindar una metodología estructurada con soporte en una herramienta cuantitativa para ofrecer a las empresas independientemente de su tamaño y poder adquisitivo la posibilidad de realizar todo el proceso de extracción de conocimiento en bases de datos con la cual podrán obtener una mayor visibilidad y conocimiento del negocio así como mejorar sus operaciones.

Tanto la metodología como la herramienta han sido validadas en una herramienta comercial, en nuestro caso fue el software R a través de la cual desarrollamos

los códigos que posteriormente fueron programados en un lenguaje VBA de excel.

Mayor definición y visibilidad al efectuar la toma de decisiones sobre los clientes y el abastecimiento, a través del uso de la herramienta es posible realizar la implementación de pronósticos para mejorar el abastecimiento de refacciones en la empresa de acuerdo a la demanda.

Así también, como parte de la investigación, se encontraron criterios base para un análisis integral y, con la aplicación de lo teórico a lo práctico, se enriquece el proyecto al tomar en cuenta factores importantes que se le atribuyen a la experiencia de la empresa en el despliegue de sus operaciones habituales en el proceso de abastecimiento.

Finalmente, se sugiere que una vez que la empresa defina la importancia de realizar de manera constante el análisis de la información generada mediante sus operaciones así como la implementación de pronósticos para realizar las compras de refacciones podrá establecer mejores acuerdos comerciales con sus proveedores, generar nuevas propuestas de trabajo a sus clientes y aumentar su capacidad de respuesta y minimizar sus tiempos de entrega con lo cual por consiguiente se logre crear una cadena de suministro integrada al fomentar y mejorar el vínculo entre cliente y proveedor.

Al contar con un mayor conocimiento sobre las fortalezas y debilidades de nuestra empresa, es posible crear estrategias a través de las cuales mejorar las operaciones y generar estrategias competitivas que contribuyan a mejorar el desempeño en general.

CAPÍTULO 6

ANEXOS

En esta sección se encuentran parte de los datos recabados y utilizados en el análisis del capítulo cuatro.

Tabla 6.1 de la serie de tiempo para el producto numero 2 de acuerdo a la información recolectada en los años 2017 y 2018.

Año	Mes	Serie Producto #2	Año	Mes	Serie Producto #2
2017	Enero	0	2018	Enero	2
2017	Febrero	2	2018	Febrero	1
2017	Marzo	2	2018	Marzo	0
2017	Abril	0	2018	Abril	0
2017	Mayo	0	2018	Mayo	1
2017	Junio	0	2018	Junio	1
2017	Julio	1	2018	Julio	1
2017	Agosto	1	2018	Agosto	1
2017	Septiembre	0	2018	Septiembre	3
2017	Octubre	0	2018	Octubre	1
2017	Noviembre	3	2018	Noviembre	0
2017	Diciembre	0	2018	Diciembre	0

Tabla 6.1: Datos del producto 2.

En la figura 6.1 podemos observar el comportamiento gráfico de la demanda del producto 2.

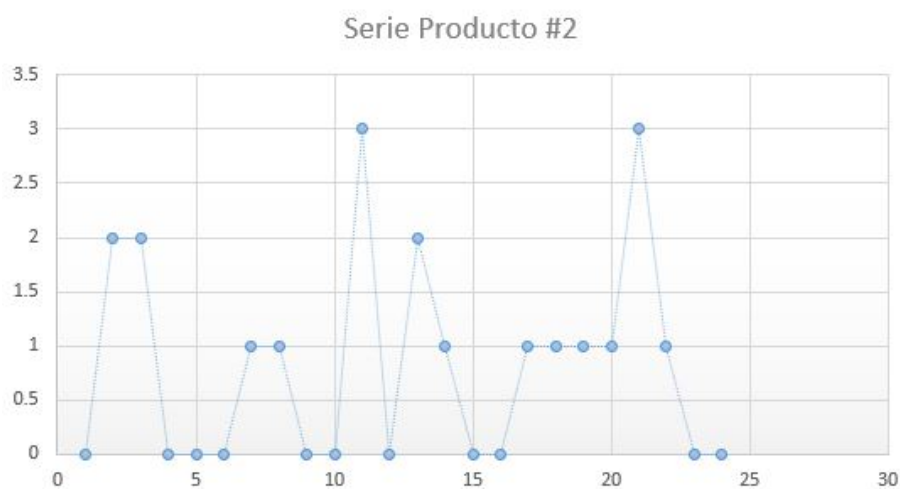


Figura 6.1: Gráfico de la demanda del producto 2.

Tabla 6.2 de la serie de tiempo para el producto numero 3 de acuerdo a la información recolectada en los años 2017 y 2018.

Año	Mes	Serie Producto #3	Año	Mes	Serie Producto #3
2017	Enero	0	2018	Enero	0
2017	Febrero	0	2018	Febrero	2
2017	Marzo	4	2018	Marzo	1
2017	Abril	0	2018	Abril	0
2017	Mayo	1	2018	Mayo	4
2017	Junio	0	2018	Junio	2
2017	Julio	2	2018	Julio	0
2017	Agosto	1	2018	Agosto	1
2017	Septiembre	0	2018	Septiembre	0
2017	Octubre	0	2018	Octubre	1
2017	Noviembre	1	2018	Noviembre	0
2017	Diciembre	0	2018	Diciembre	0

Tabla 6.2: Datos del producto 3.

En la figura 6.2 podemos observar el comportamiento gráfico de la demanda del producto 3.

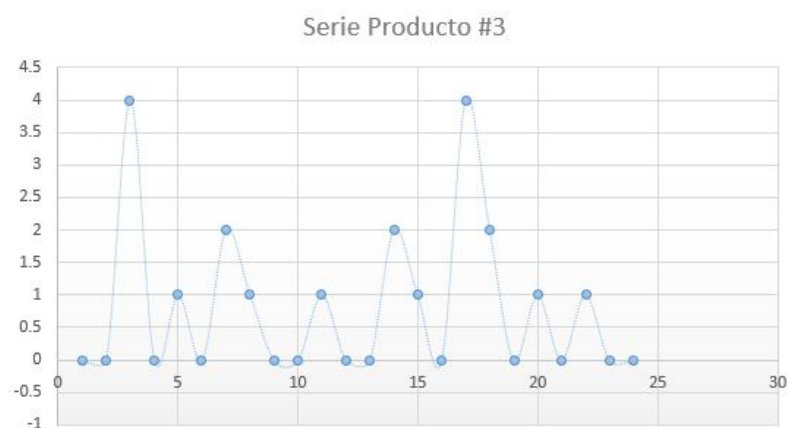


Figura 6.2: Gráfico de la demanda del producto 3

Tabla 6.3 de la serie de tiempo para el producto numero 4 de acuerdo a la información recolectada en los años 2017 y 2018.

Año	Mes	Serie Producto #4	Año	Mes	Serie Producto #4
2017	Enero	0	2018	Enero	0
2017	Febrero	0	2018	Febrero	0
2017	Marzo	0	2018	Marzo	1
2017	Abril	3	2018	Abril	2
2017	Mayo	0	2018	Mayo	2
2017	Junio	3	2018	Junio	1
2017	Julio	4	2018	Julio	0
2017	Agosto	1	2018	Agosto	0
2017	Septiembre	1	2018	Septiembre	0
2017	Octubre	3	2018	Octubre	1
2017	Noviembre	2	2018	Noviembre	0
2017	Diciembre	1	2018	Diciembre	0

Tabla 6.3: Datos del producto 4.

En la figura 6.3 podemos observar el comportamiento gráfico de la demanda del producto 4.

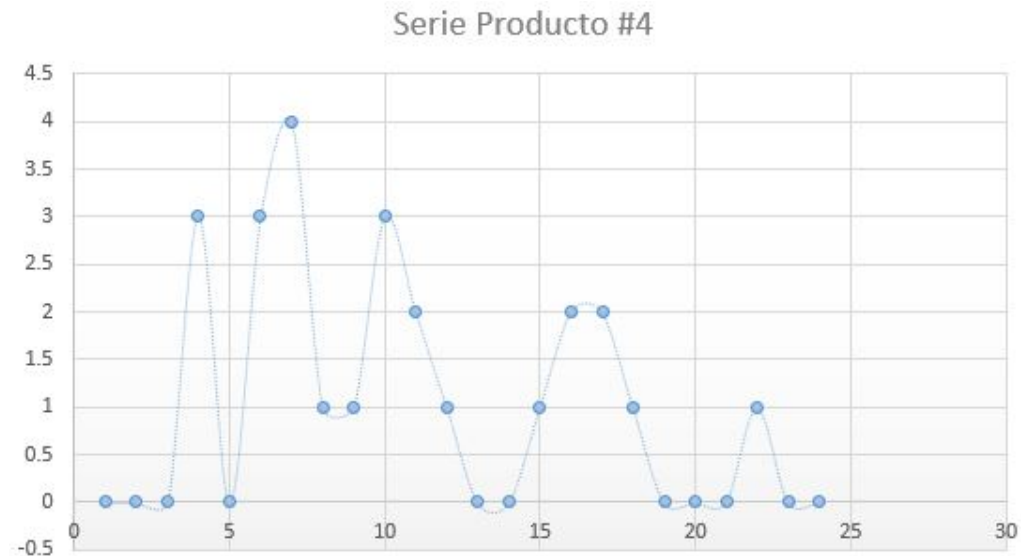


Figura 6.3: Gráfico de la demanda del producto 4

6.0.1 PSEUDOCÓDIGO DE LIMPIEZA:

```
Sub [Nombre del Procedimiento]()
```

```
Dim [Nombre de la variable 1] As String
```

```
Dim [Nombre de la variable 2] As String
```

```
Dim [Nombre de la variable 3] As Range
```

```
On Error Resume Next
```

```
Set [Nombre de la variable 3] = Cells
```

```
[Nombre variable 1] = InputBox (“[Instrucción al usuario]”, “Buscar”, “ “)
```

```
[Nombre variable 1] = “*” & [Nombre variable 1] & “*” [Nombre variable 2]  
= InputBox (“[Instrucción al usuario]”, “Renombrar”, “ “)
```

```
If xFind = "False" Or xRep = "False" then
```

```
Exit
```

```
Sub [Nombre de la variable 3].Replace [Nombre de la variable 1], [Nombre de la variable 2], xlPart, xlByRows, False, False, False, False
```

6.0.2 PSEUDOCODIGO PARA REALIZAR LA EXTRACCIÓN DE DATOS

```
Sub [Nombre del Proceso]
```

```
Dim [Nombre variable 1] As String
```

```
Dim [Nombre variable 2] As String
```

```
Dim [Nombre variable 3] As Long
```

```
Dim [Nombre variable 4] As Long
```

```
Dim [Nombre variable 5] As Long
```

```
Dim [Nombre variable 6] As Long
```

```
[Nombre variable 6] = InputBox("Nombre producto o servicio" , "producto o servicio nuevo")
```

```
[Nombre variable 6] = "*" & producto o servicio & "*"
```

```
[Nombre variable 3] = Sheets("numero o nombre de hoja").Range ("fila" & Rows.Count).End(xlUp).Row
```

```
If [Nombre variable 6] < 2 Then
```

```
Exit Sub
```

```

End If

For [Nombre variable 5] = 2 to [Nombre variable 3]

    If Sheets("numero o nombre de hoja").Cells([Nombre variable 5],4) Like [Nombre variable 6] Then

        [Nombre variable 1] = Sheets("numero o nombre de hoja").Cells([Nombre variable 5],4)

        [Nombre variable 2] = Sheets("numero o nombre de la hoja").Cells([Nombre variable 5],5)

        [Nombre variable 4] = Sheets("numero o nombre de hoja").Range("A" & Rows.Count).End(xlToLeft)
        Sheets("numero o nombre de hoja").Cells([Nombre variable 4] + 1 , 1) = [Nombre variable 1]
        Sheets("numero o nombre de hoja").Cells([Nombre variable 4] + 1 , 2) = [Nombre variable 2]

    End If

Next [Nombre variable 5]

End Sub

```

6.0.3 PSEUDOCÓDIGO PARA DESCOMPONER LAS FECHAS:

```

Sub [Nombre del Procedimiento]()

    Columns("columnas").Select

    Application.CutCopyMode = False

    Selection.TextToColumns Destination:=Range("columna"), _DataType:=xlDelimited,
    _TextQualifier:=xlDoubleQuote, _ConsecutiveDelimiter:=False, _Tab:=True, _Semicolon:=False,
    _Comma:=False, _Space:=False, _Other:=True, _OtherChar:= "/", _FieldInfo:=Array(Array(1,

```

```
1), _Array(2, 1), _Array(3, 1)), _TrailingMinusNumbers:=True
```

```
Columns(":").Select Selection.Copy Columns(":").Select Selection.PasteSpecial
Paste:=xlPasteFormats, _ Operation:=xlNone, _ SkipBlanks:=False, _ Transpose:=False
```

```
Application.CutCopyMode = False
```

```
End Sub
```

6.0.4 PSEUDOCÓDIGO PARA EXTRAER LA FECHA:

```
Sub [Nombre del procedimiento]()
```

```
Dim [Nombre variable 1] As Double
```

```
[Nombre variable 1] = Range ("fila" & Rows.Count).End(xlUp).Row
```

```
For j=1 to largo
```

```
For a = .año" To .año
```

```
For c = .año" To .año
```

```
If [Nombre variable 1] < 1 Then x = 0
```

```
Range(Cells(j, ), Cells(, )) = WorksheetFunction.Application.CountIfs(Range(":" &
[Nombre variable 1]), j, Range(":" & [Nombre variable 1]), a)
```

```
Range(Cells(j, ), Cells(, )) = WorksheetFunction.Application.CountIfs(Range(":" &
[Nombre variable 1]), j, Range(":" & [Nombre variable 1]), rango)
```

```
Next
```

```
Next
```

```
Next
```

End Sub

6.0.5 PSEUDOCÓDIGO PARA APLICAR MÉTODO DE PRONÓSTICO SUAVIZACIÓN EXPONENCIAL SIMPLE:

```
Sub [Nombre del Proceso]()  
  
Dim [Nombre variable 1] As Double  
  
Dim [Nombre variable 2] As Double  
  
Dim [Nombre variable 3] As Double  
  
Dim [Nombre variable 4] As Integer  
  
Dim [Nombre variable 5] As Integer  
  
Dim [Nombre variable 6] As Integer  
  
Dim [Nombre variable 7] As Integer  
  
Dim [Nombre variable 8] As Double  
  
Dim [Nombre variable 9] As Double  
  
Dim [Nombre variable 10] As Double  
  
Dim [Nombre variable 11] As Single  
  
Dim [Nombre variable 12] As Single  
  
Monte_carlo = 1000  
  
Dim [Nombre variable 13] As Double  
  
[Nombre variable 11] = Timer()
```

```

For [Nombre variable 4] = 1 to Monte_carlo

[Nombre variable 13]([Nombre variable 4]) = Rnd()

Next [Nombre variable 7] = Range("fila" & Rows.Count).End(xlUp).Row

[Nombre variable 1])(1) = Cells(,

For [Nombre variable 4]=1 to Monte_carlo

For [Nombre variable 5] = 2 to [Nombre variable 7]

[Nombre variable 1]([Nombre variable 5]) = [Nombre variable 13](j) + (1+[Nombre
variable 13](j)) * (Cells([Nombre variable 5]-,))

Next

For [Nombre variable 5] = 1 To [Nombre variable 7]

[Nombre variable 10]([Nombre variable 5]) = (Cells ([Nombre variable 5],7 +
[Nombre variable 1]([Nombre variable 5])) ^ 2

Cells([Nombre variable 5], 17) = Cells ([Nombre variable 5], 7)

Cells([Nombre variable 5], 20) = [Nombre variable 1]([Nombre variable 5])

Cells([Nombre variable 5], 21) = [Nombre variable 10]([Nombre variable 5])

Next

aux = Application.WorksheetFunction.Sum([Nombre variable 10])

[Nombre variable 3]([Nombre variable 4]) = aux / [Nombre variable 7]

Next

For [Nombre variable 6] = 1 To 50 Cells(3, 13) = .ErrorCells(3, 14) = [Nombre
variable 3](1) Next

```

```
[Nombre variable 12]= Timer()

MsgBox ("Tiempo de ejecucion del calculo:"& vbNewLine & Round([Nombre
variable 12]- [Nombre variable 13], 2) & "segundos")

aux = 0

Index = Application.Match(Application.Min([Nombre variable 3]), [Nombre
variable 3], 0) - 1

For [Nombre variable 6] = 1 To monte_carlo

Next

End Sub
```

6.0.6 PSEUDOCÓDIGO PARA APLICAR MÉTODO DE PRONÓSTICO CROSTON:

```
Sub [Nombre del Procedimiento]()

Dim [Nombre de la variable 1] As Double

Dim [Nombre de la variable 2] As Double

Dim [Nombre de la variable 3] As Double

Dim [Nombre de la variable 4] As Integer

Dim [Nombre de la variable 5] As Double

Dim [Nombre de la variable 6] As Double

Dim [Nombre de la variable 7] As Double

Dim [Nombre de la variable 8] As Double
```

```
Dim [Nombre de la variable 9] As Integer

Dim [monte_carlo(1000)] As Double

Dim [Nombre de la variable 11] As Integer

Dim [Nombre de la variable 12] As Double

Dim [Nombre de la variable 13] As Double

Dim [Nombre de la variable 14] As Single

Dim [Nombre de la variable 15] As Single

For [Nombre de la variable 9] = 1 to [monte_carlo]

alpha (Nombre de la variable 9) = Rnd()

Next

Dim [Nombre de la variable 16] As Integer

For [Nombre de la variable 9] = 0 To [Nombre de la variable11]-1

[Nombre de la variable 7] ([Nombre de la variable 9]) = 0

[Nombre de la variable 5] ([Nombre de la variable 9]) = 0

[Nombre de la variable 6] ([Nombre de la variable 9]) = 0

Next

[Nombre de la variable 11] = Range("G" & Rows.Count).End(xlUp).Row

[Nombre de la variable 14] = Timer()

[Nombre de la variable 1](0) = 99

For [Nombre de la variable 9] To [Nombre de la variable11]
```

```

[Nombre de la variable0]([Nombre de la variable9]) = Cells
([Nombre de la variable9],7) Next

For [Nombre de la variable 9] To [Nombre de la variable11]

Cells ([Nombre de la variable9],8) = [Nombre de la variable0]([Nombre de la
variable9])

Next

Index_nonega = Application.match(0, [Nombre de la variable 1], 0) - 1

[Nombre de la variable 7] = [Nombre de la variable 1](index_no_nega)

[Nombre de la variable 4] = index_no_nega

[Nombre de la variable 6](0) = index_no_nega

For al = 1 to 1000 For t = 1 to [Nombre de la variable11]

If [Nombre de la variable1] (t) > 0 Then

[Nombre de la variable7](t) = [Nombre de la variable11](t-1) + [Nombre de
la variable8](al) * ([Nombre de la variable1] (t) - [Nombre de la variable7] (t-1))
[Nombre de la variable6](t) = [Nombre de la variable11] (t-1) + [Nombre de la
variable8](al) * ([Nombre de la variable4] - [Nombre de la variable6](t-1)) [Nombre
de la variable4] = 1

Else [Nombre de la variable7](t) = [Nombre de la variable7](t-1)

[Nombre de la variable6](t) = [Nombre de la variable6](t-1)

[Nombre de la variable4] = [Nombre de la variable4] + 1

End IF

[Nombre de la variable5](t) = [Nombre de la variable7](t) / [Nombre de la

```

variable6](t) Next

For t = 1 to [Nombre de la variable11]

[Nombre de la variable12](t) = ([Nombre de la variable1](t) - [Nombre de la variable5](t))² Cells (1,10 + t) = [Nombre de la variable12](t)

Next

Aux = Application.WorksheetFunction.Sum([Nombre de la variable12]) [Nombre de la variable13](al) = aux / ([Nombre de la variable11] - 1

Cells(4,11) = aux

Cells(5,11) = [Nombre de la variable13](al)

Next

For [Nombre de la variable9] = 1 To 1000

Cells([Nombre de la variable9],2) = [Nombre de la variable13]([Nombre de la variable9])

Cells([Nombre de la variable9],3) = [Nombre de la variable8]([Nombre de la variable9])

Index = Application.match(Application.Min([Nombre de la variable13]), [Nombre de la variable13],0) - 1

Aux = 0

[Nombre de la variable13](0) = 999

Next

[Nombre de la variable15] = Timer() MsgBox (“Time taken to run code:” & vbNewLine & Round ([Nombre de la variable15] - [Nombre de la variable14],2) & “Seconds”)

BIBLIOGRAFÍA

- ABRIL, J. C. (2011), «Análisis de la evolución de las técnicas de series de tiempo: Un enfoque unificado», .
- AFACAN-SEREF, K., N. A. STEINEMANN, A. BLANGERO y S. P. KELLY (2018), «Dynamic interplay of value and sensory information in high-speed decision making», *Current Biology*, **28**(5), págs. 795–802.
- ARANA, D. (2018), «Pymes mexicanas, un panorama para 2018», .
- ARCILA-CALDERÓN, C., E. BARBOSA-CARO y F. CABEZUELO-LORENZO (2016), «Técnicas big data: Análisis de textos a gran escala para la investigación científica y periodística», *El profesional de la información (EPI)*, **25**(4), págs. 623–631.
- BARRETO, L., A. AMARAL y T. PEREIRA (2017), «Industry 4.0 implications in logistics: an overview», *Procedia Manufacturing*, **13**, págs. 1245–1252.
- BARTSCHAT, A., M. REISCHL y R. MIKUT (2019), «Data mining tools», *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, pág. e1309.
- BENTLEY, J. W., P. VAN MELE, M. HARUN-AR-RASHID y T. J. KRUPNIK (2016), «Distributing and showing farmer learning videos in Bangladesh», *The Journal of Agricultural Education and Extension*, **22**(2), págs. 179–197.
- BERTSIMAS, D., N. KALLUS y A. HUSSAIN (2016), «Inventory management in the era of big data», *Production and Operations Management*, **25**(12), págs. 2006–2009.

- CANACAR (2015), «Movimiento de carga por modo de transporte», Recuperado el 20 de mayo del 2019, disponible en <https://canacar.com.mx/stat/movimiento-carga-modo-transporte/>.
- CARLTON, A. (1969), «On the bias of information estimates.», *Psychological Bulletin*, **71**(2), pág. 108.
- CASADO, R. y M. YOUNAS (2015), «Emerging trends and technologies in big data processing», *Concurrency and Computation: Practice and Experience*, **27**(8), págs. 2078–2091.
- CEPAL, N. *et al.* (2005), «El aporte de FUNDES a las PYMES en América Latina: una evaluación de impacto», .
- CHEAH, S. y S. WANG (2017), «Big data-driven business model innovation by traditional industries in the Chinese economy», *Journal of Chinese Economic and Foreign Trade Studies*, **10**(3), págs. 229–251.
- CHOI, T.-M., S. W. WALLACE y Y. WANG (2018), «Big data analytics in operations management», *Production and Operations Management*, **27**(10), págs. 1868–1883.
- CROSTON, J. D. (1972), «Forecasting and stock control for intermittent demands», *Journal of the Operational Research Society*, **23**(3), págs. 289–303.
- CUI, R., S. GALLINO, A. MORENO y D. J. ZHANG (2018), «The operational value of social media information», *Production and Operations Management*, **27**(10), págs. 1749–1769.
- ESPINAL, A. C. y R. A. G. MONTOYA (2009), «Tecnologías de la información en la cadena de suministro», *Dyna*, **76**(157), págs. 37–48.
- FAYYAD, U. M., G. PIATETSKY-SHAPIRO, P. SMYTH y R. UTHURUSAMY (1996), «Advances in knowledge discovery and data mining», .

- FÉLIX, M. Z., G. M. CAVAZOS y J. M. Q. DÍAZ (2013), «Ventaja competitiva de la cadena de suministros: alianzas, asimetrías organizativas y conflictos», *Revista Nacional de Administración*, **4**(1), págs. 109–118.
- GARCÍA CANSECO, L. B. (2015), «Causas del fracaso en las pyme», *Emprendedores*, (151), págs. 22–25.
- GHOBBAR, A. A. y C. H. FRIEND (2002), «Sources of intermittent demand for aircraft spare parts within airline operations», *Journal of Air Transport Management*, **8**(4), págs. 221–231.
- GIBSON, B., M. GIBSON y S. RUTNER (1998), «Careers in logistics», *Council of Logistics Management, Oak Brook, IL*.
- GREIS, N. P. y M. L. NOGUEIRA (2011), «Use of data mining for validation and verification of maritime cargo movement», *Institute for Homeland Security Solutions (Research Brief)*. North Carolina: University of North Carolina.
- GUPTA, R., H. GUPTA, U. NAMBIAR y M. MOHANIA (2012), «Enabling active data archival over cloud», en *Services Computing (SCC), 2012 IEEE Ninth International Conference on*, IEEE, págs. 98–105.
- GUTIERREZ, R. S., A. O. SOLIS y S. MUKHOPADHYAY (2008), «Lumpy demand forecasting using neural networks», *International Journal of Production Economics*, **111**(2), págs. 409–420.
- HAND, D. J. (2006), «Data Mining», *Encyclopedia of Environmetrics*, **2**.
- HANKE, J. E. y D. W. WICHERN (2006), *Pronósticos en los negocios*, Pearson educación.
- HARRISON, A. y C. NEW (2002), «The role of coherent supply chain strategy and performance management in achieving competitive advantage: an international survey», *Journal of the Operational Research Society*, **53**(3), págs. 263–271.

- INC., E. (2016), «Logistic Cost and Service 2016», <https://static1.squarespace.com/static/57bf65a1c534a52224df643c/t/57f6a2aa6a4963b686c1bc3>
- INTRONA, L. D. (1991), «The impact of information technology on logistics», *International Journal of Physical Distribution & Logistics Management*, **21**(5), págs. 32–37.
- LACEY, A. y D. LUFF (2001), *Qualitative data analysis*, Trent Focus Sheffield.
- LAMBERT, D. M., M. C. COOPER y J. D. PUGH (1998), «Supply Chain Management: Implementation Issues and Research Opportunities», *The International Journal of Logistics Management*, **9**(2), págs. 1–20, URL <https://doi.org/10.1108/09574099810805807>.
- LHOR, S. (2012), «The Age of Big Data», (<https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html>).
- LORDI, H. A. (2017), «“BIG DATA Y SU APLICACIÓN EN EL TRANSPORTE”», Recuperado el 20 de mayo del 2019, disponible en <http://www.aacarreteras.org.ar/pdf/Lordi-BIG-DATA-AAC-VF.pdf>.
- MCAFEE, A., E. BRYNJOLFSSON, T. H. DAVENPORT, D. PATIL y D. BARTON (2012), «Big data: the management revolution», *Harvard business review*, **90**(10), págs. 60–68.
- MONGE, E. C. (2010), «Las estrategias competitivas y su importancia en la buena gestión de las empresas», *Revista de Ciencias Económicas*, **28**(1).
- NARVÁEZ, M. y G. FERNÁNDEZ (2008), «Estrategias competitivas para fortalecer sectores de actividad empresarial en el mercado global», *Revista Venezolana de Gerencia*, **13**(42), págs. 233–243.
- PORTER, M. E. (1900), *Ventaja competitiva: creación y sostenimiento de un desempeño superior*, Grupo Editorial Patria.

- QUERO, L. (2008), «Estrategias competitivas: Factor clave de desarrollo», *Revista Negotium*, (10), págs. 36–49.
- RODRIGUEZ, A. R. S. C. y C. CORRÊA (2017), «Previsión de demanda intermitente con métodos de series de tiempo y redes neuronales artificiales: Estudio de caso», *Dyna*, **84**(203), págs. 9–16.
- SAGAERT, Y. R., E.-H. AGHEZZAF, N. KOURENTZES y B. DESMET (2017), «Temporal big data for tactical sales forecasting in the tire industry», *Interfaces*, **48**(2), págs. 121–129.
- SAUCEDO-MARTÍNEZ, J. A., M. PÉREZ-LARA, J. A. MARMOLEJO-SAUCEDO, T. E. SALAIS-FIERRO y P. VASANT (2017), «Industry 4.0 framework for management and operations: a review», *Journal of Ambient Intelligence and Humanized Computing*, págs. 1–13.
- SCT (2017), «Sector Telecomunicaciones Primer Trimestre 2017», .
- SERRANO-COBOS, J. (2014), «Big data y analítica web. Estudiar las corrientes y pescar en un océano de datos», *El profesional de la información*, **23**(6), págs. 561–565.
- SILVA, J. D. (2017), «Gestión de la cadena de suministro: una revisión desde la logística y el medio ambiente», *Entre Ciencia e Ingeniería*, **11**, págs. 51 – 59.
- TABARES, L. F. y J. F. HERNANDEZ (2014), «Big Data Analytics: Oportunidades, Retos y Tendencias», *Universidad de San Buenaventura*, pág. 20.
- TUNGUZ, T. y F. BIEN (2016), *Winning with data: transform your culture, empower your people, and shape the future*, John Wiley & Sons.
- VIGLIONI, C., M. V. Q. CURY y P. SILVA (2007), «Methodology for railway demand forecasting using data mining», en *SAS global forum*, tomo 161, págs. 1–8.
- VILLAVICENCIO, J. (2010), «Introducción a series de tiempo», *Metodología de series de tiempo*, **4**.

WANG, S. J. y P. MORIARTY (2018), «Big Data for Sustainable Urban Transport», en *Big Data for Urban Sustainability*, Springer, págs. 81–103.

ZMUD, J., M. LEE-GOSSELIN, M. MUNIZAGA y J.-A. CARRASCO (2013), «Transport surveys: Considerations for decision makers and decision making», *Transport Survey Methods: Best practice for decision making*, págs. 3–19.

RESUMEN AUTOBIOGRÁFICO

Fátima Sánchez Ruiz

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

Universidad Autónoma de Nuevo León
Facultad de Ingeniería Mecánica y Eléctrica

Tesis:

Big data COMO ESTRATEGIA COMPETITIVA APLICADO EN LA
LOGÍSTICA Y EL TRANSPORTE DE PYMES

Originaria de la ciudad de San Luis Potosí, nací el 19 de abril de 1991, mi madre, Monica Ruiz Gallegos y mi padre Ruben Sanchez de la Rosa.

Curse la carrera de ingeniería en sistemas y tecnologías industriales en la Universidad Politécnica de San Luis Potosí. Realice mis practicas profesionales en empresas como: Continental, Mabe, AAP en areas como calidad, ergonómica y seguridad industrial respectivamente, en el 2004 comencé a laboral en la empresa Opción Diesel como gerente de operaciones y años mas tarde en la empresa ISALI como gerente de calidad.

Obtuve certificaciones como consultor interno de la norma ISO 1400 y OSHAS 1800 y administrador de garantías CUMMINS.