

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

FACULTAD DE CIENCIAS E INGENIERÍA



PUCP

**REVISIÓN DE LA LITERATURA DE LA ADMINISTRACIÓN DE
SERVICIOS Y EL USO DE *MACHINE LEARNING* Y *DATA MINING* EN
EL SECTOR *RETAIL***

**Trabajo de investigación para la obtención del grado académico de
BACHILLER EN CIENCIAS CON MENCIÓN EN INGENIERÍA
INDUSTRIAL**

AUTOR:

Marco Alonso Jorge Wong

ASESOR:

Christian Santos Cornejo Sánchez

Lima, diciembre de 2020

Resumen

Este trabajo de investigación tiene dos objetivos: constatar que los servicios y el sector *retail* son campos de investigación relevantes en la actualidad a través de una revisión de la literatura y por su participación en la economía peruana y mundial; así como preparar el marco teórico necesario para el desarrollo de la tesis. En esta línea, primero se desarrolla el concepto del *retail* y los tipos de empresas que participan en este sector; para ello, se describe la adición de valor que las compañías de este rubro proporcionan dentro de sus servicios. Luego, se presenta la definición de los servicios, su aporte al PBI mundial y peruano, y también la descripción desde el enfoque de proceso. Se realiza una comparación entre la producción de un bien manufacturado versus la prestación de un servicio que resulta en una serie de diferencias donde destaca el rol altamente participativo del cliente dentro de los servicios. Por ello, se presenta también, el concepto de CRM (*Customer Relationship Management*) que busca generar relaciones de valor con los clientes. Además, se desarrollan las herramientas de diagnóstico que luego serán utilizadas en la tesis para definir la situación actual de la empresa a estudiar. Después, se describen los conceptos de *Machinle laerning* (ML) y *Data mining* (DM) y, en particular, el método de aprendizaje no supervisado: *clustering* (agrupamiento). Estas definiciones luego serán utilizadas en planteamiento del modelo de la tesis, junto con los *softwares* de aplicación cuyo detalle se desarrolla también en este trabajo de investigación. En el estado del arte se revisa la literatura relacionada con los servicios y el *retail*, que según la revisión efectuada son campos de estudio de interés tanto por su impacto en la economía, como por las investigaciones realizadas. Además, durante los últimos años, los servicios han localizado al cliente en el centro de sus operaciones y buscan generar una entrega de valor bilateral a través de estrategias de CRM. Dichas estrategias son: adquisición, retención y fidelización de clientes, las cuales pueden mejorar sus resultados al implementarse con herramientas analíticas como DM y ML. Finalmente, la conclusión del trabajo de investigación se resume en que, en la actualidad, las personas están expuestos a variadas y abundantes ofertas de distintas empresas que buscan convertirlo en su cliente. En este sentido, las empresas han reaccionado frente a un mercado mucho más competitivo, a través de herramientas que puedan mejorar la gestión de sus clientes, como el CRM, y algunas han complementado las estrategias de esta metodología con las técnicas analíticas de DM y ML. En particular, en el contexto peruano, hay precedentes de la utilización de todas las técnicas mencionadas en el sector *retail* que han generado beneficios sustanciales para la gestión de la clientela.

Tabla de contenidos

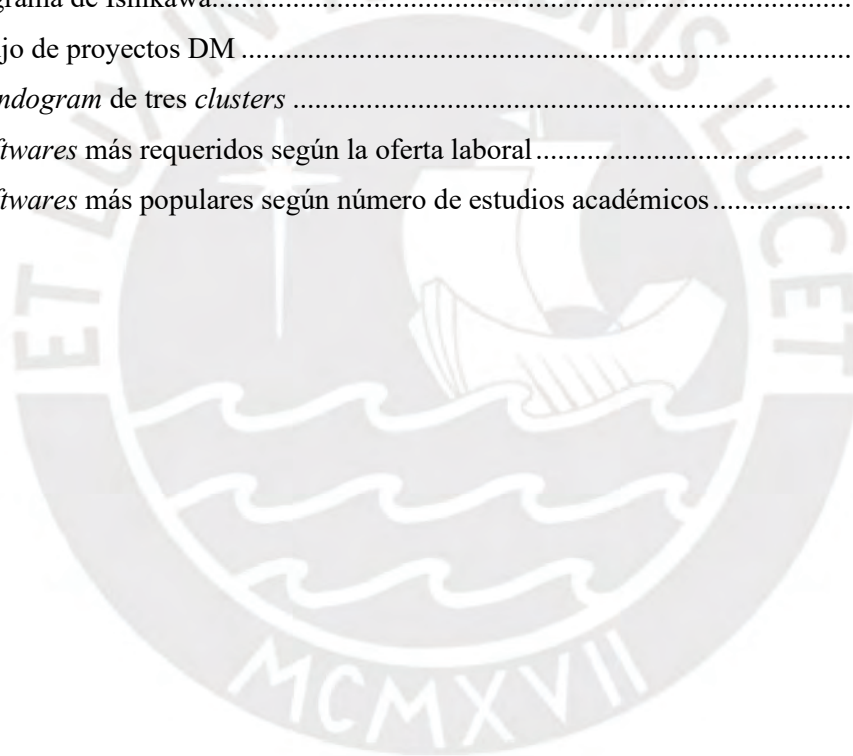
Resumen.....	i
Tabla de contenidos	ii
Índice de figuras.....	iv
Índice de tablas	v
Capítulo 1. Marco teórico	1
1.1 Sector <i>retail</i>	1
1.1.1 El modelo <i>retail</i>	1
1.1.2 Tipos de empresas <i>retail</i>	2
1.2 Administración de servicios.....	5
1.2.1 El servicio como un proceso	8
1.2.2 Gestión de Relaciones con los Clientes (CRM).....	10
1.3 Herramientas de diagnóstico.....	12
1.3.1 Cadena de valor.....	12
1.3.2 Mapeo de procesos.....	13
1.3.3 Flujogramas.....	15
1.3.4 Matriz QFD.....	16
1.3.5 Proceso de análisis jerárquico	17
1.3.6 Diagrama de Ishikawa.....	18
1.3.7 Matriz de probabilidad - impacto.....	19
1.3.8 5 por qué	19
1.4 Herramientas del modelo	20
1.4.1 Machine learning.....	20
1.4.2 Data Mining	22
1.4.3 <i>Clustering</i> (agrupamiento).....	24
1.5 <i>Softwares</i> de aplicación.....	27
1.5.1 R.....	28
1.5.2 Python	29
1.5.3 SQL	30

Capítulo 2. Estado del arte.....	31
Capítulo 3. Conclusiones.....	35
Referencias.....	36



Índice de figuras

<i>Figura 1.</i> Aporte al PBI peruano por sector en 2018.....	8
<i>Figura 2.</i> Participación de cada tipo de servicio dentro del aporte al PBI peruano de 2018 del sector servicios	8
<i>Figura 3.</i> Clasificación de los servicios según la naturaleza de los actos y el tipo de receptor.....	9
<i>Figura 4.</i> El continuo del CRM	10
<i>Figura 5.</i> Esquema de cadena de valor	13
<i>Figura 6.</i> Ejemplo de mapa de proceso	14
<i>Figura 7.</i> Simbología ANSI.....	15
<i>Figura 8.</i> Matriz QFD para negocios.....	16
<i>Figura 9.</i> Diagrama de Ishikawa.....	18
<i>Figura 10.</i> Flujo de proyectos DM.....	23
<i>Figura 11.</i> Dendogram de tres clusters	25
<i>Figura 12.</i> Softwares más requeridos según la oferta laboral.....	27
<i>Figura 13.</i> Softwares más populares según número de estudios académicos.....	28



Índice de tablas

Tabla 1	<i>Tipos de formato para retails de productos alimenticios.</i>	3
Tabla 2	<i>Tipos de formato para retails de productos generales</i>	4
Tabla 3	<i>Tipos de formato para retails de productos generales</i>	7
Tabla 4	<i>Algoritmo K - means clustering</i>	25
Tabla 5	<i>Agglomerative hierarchical algorithm</i>	26
Tabla 6	<i>Divisive hierarchical algortihm</i>	26



Capítulo 1. Marco teórico

En este capítulo se describe los conceptos que se utilizarán en este trabajo de investigación. Se desarrollan los fundamentos del sector *retail*, la administración de servicios, herramientas de diagnóstico y las herramientas informáticas que se utilizarán en el desarrollo de la tesis.

1.1 Sector *retail*

El término “*retail*” es una palabra originaria del idioma inglés cuya traducción aproximada al español es “comercio minorista” y agrupa a todos los negocios que comercializan bienes o servicios a consumidores finales. Este sector se caracteriza por la variedad de productos que puede ofrecer (e.g., útiles de escritorio o artículos deportivos para canotaje); por desarrollar una venta de gran cantidad de productos a un cuantioso número de compradores diferentes; por el bajo ingreso por venta que percibe y por la adquisición de productos en grandes lotes (Berman, Evans y Chatterjee, 2018).

1.1.1 El modelo *retail*

Dado que, el proceso *retail* puede entenderse únicamente como la etapa final de la cadena de distribución de un producto, puede resultar complicado, para el cliente, observar el valor agregado que las empresas *retail* añaden al producto. Para Zentes, Morschett y Schramm-Klein (2011), esta situación induce a las empresas del sector a concentrar sus esfuerzos en determinadas funciones que favorezcan la percepción de valor añadido que las compañías entregan. A continuación, se detallan esas acciones.

- **Clasificación de bienes**

Como se ha mencionado, las empresas *retail* atienden la demanda de un numeroso grupo de clientes que; sin embargo, no compra cuantiosas cantidades de productos. En ese sentido, según Coughlan, Anderson, Stern y El-Ansari (2006), estos clientes buscan diversidad de bienes y servicios, pero en poca cantidad. Por ello, la compañía añade valor al clasificar, ordenar y surtir sus productos; así como facilitar la compra de lotes pequeños.

- **Puente entre fabricantes y consumidores**

Los consumidores buscan satisfacer sus necesidades mediante bienes que son producidos por fabricantes ubicados en diferentes lugares. En tal sentido, las empresas *retail* muestran el valor añadido que otorgan al funcionar como intermediarias que acumulan variados bienes de diferentes fabricantes en un solo lugar al que los consumidores pueden acceder. Asimismo, los clientes ahorran tiempo al encontrar lo que demandan en un solo lugar.

- **Creación y estimación de demanda**

Mediante publicidad, análisis de sus clientes, campañas y promociones, las empresas pueden incentivar la demanda de sus productos. Además, mediante el conocimiento que obtienen de las ventas que realizan continuamente, podrían estimar los productos más solicitados y sus demandas. Esto favorece tanto a los proveedores -al solicitarles lotes de tamaños más ajustados a la realidad-, como a los consumidores al ofertar los productos que buscan.

- **Servicios adicionales**

Son todos los servicios extras relacionados con el producto vendido que la empresa puede ofrecer para agregar valor. Así, algunos negocios *retail* no solo ofrecen bienes estándar, sino que comercializan productos realizados con especificaciones del cliente; por ejemplo, la empresa Hockerty¹ ha vendido más de un millón de prendas confeccionadas a medida durante sus 10 años de operaciones (Samaniego, 2019). Por otro lado, las compañías *retails* suelen incluir los servicios de preparación, mantenimiento y reparación del producto que suelen ser requeridos cuando se venden electrodomésticos.

Las funciones descritas previamente caracterizan a los *retail* tradicionales; sin embargo, debido al incremento de poder de negociación de las empresas de ese sector, con respecto a los fabricantes, algunas compañías comenzaron a adquirir una mayor participación en toda la cadena de valor (Zentes et al., 2011). De ese modo, estas empresas han añadido nuevas acciones a la lista mencionada previamente. Dichas funciones se detallan a continuación.

- **Coordinación y producción de bienes**

Dada la cantidad de información que las empresas pueden generar a partir de la compra de sus clientes, algunas están impulsando la producción de productos de marca propia. Para estos bienes en particular, las compañías tienen injerencia en toda la cadena de valor, es decir, participan desde la creación del producto hasta la venta final. En algunos casos, solicitan a un proveedor la fabricación del bien de acuerdo con sus especificaciones (e.g., las marcas propias de Hiraoka y de Makro: “Miray” y “Aro” respectivamente) y, en otros, los mismos negocios *retail* se encargan de elaborar el producto (e.g, la comida preparada por Tambo).

1.1.2 Tipos de empresas *retail*

Pese a que todas las compañías *retail* guardan similitudes en cuanto al modelo de negocio, es importante mencionar que estas integran componentes *ad hoc* a sus planes estratégicos de acuerdo con los productos que ofrece y el público al que atiende.

¹ Hockerty es una empresa de origen español que produce ropa a la medida de sus clientes en el mercado europeo. Cuenta con tiendas en Barcelona, Zúrich y Shanghái.

En esta línea, Berman et al. (2018) clasifican a los *retail* en dos grupos: los orientados a la venta de productos alimenticios y los que comercializan mercancías generales (muebles, electrodomésticos, ropa, entre otros). Además, proponen una categorización de formatos de tienda para cada grupo basada en aspectos como la ubicación, la mercadería y servicio que ofrecen, los precios y la publicidad que usan. En las tablas 1 y 2 se muestran los formatos para los *retail* de productos comestibles y generales respectivamente que se pueden encontrar en el mercado peruano.

Tabla 1

Tipos de formato para retails de productos alimenticios.

Formato	Ubicación	Mercadería	Precios	Servicio	Publicidad
Tiendas de conveniencia	Vecindarios	Ancho y corto surtido de calidad media	Promedios a encima de la media	Promedio	Moderada
Supermercados/ hipermercados	Centros comerciales	Amplio y largo surtido de calidad media	Competitivos	Promedio	Extensiva en medios de comunicación, periódicos y afiches.
Mercados minoristas	Vecindarios, zonas concurridas	Surtido completo de supermercado con productos de salud, limpieza y otros	Bajos a promedio	Promedio o debajo de la media	Poca o ninguna
Tiendas de descuento / <i>minimarkets</i>	Periferias de la ciudad	Estrecho y corto surtido de calidad media	Bajos a promedio	Debajo de la media a promedio	Afiches
Bodegas	Vecindarios	Estrecho y corto surtido de calidad media	Promedios	Debajo de la media a promedio	Ninguna

Nota. Adaptado de Berman et al. (2018).

- **Tiendas de conveniencia:** corresponde al formato que busca ofrecer sus productos de amplio surtido, pero bajo volumen; en un espacio relativamente pequeño (entre 80 y 120 m²) y en un amplio horario de atención. Algunos arquetipos en el Perú son Tambo, Oxxo y Repshop.
- **Supermercados e hipermercados:** son formatos muy reconocidos del *retail* y ofrecen un diverso y cuantioso surtido de productos en una extensa área (supermercados entre 400 y 1000 m²; hipermercados, más de 1000 m²). Suelen enfocarse en productos comestibles, pero también ofrecen otros bienes debido a su gran tamaño. Los supermercados más conocidos en Perú son Plaza Vea, Tottus, Wong, Metro y Vivanda.

- **Mercados minoristas:** son los mercados tradicionales que se establecen en diferentes vecindarios para ofertar sus productos a los consumidores de los alrededores. Suelen abarcar áreas extensas y ofrecer productos de surtido similar a los supermercados a precios bajos.
- **Tiendas minoristas y *minimarkets*:** son las tiendas que se establecen en las zonas periféricas de la ciudad con una oferta de productos a precios bajos. Basan su modelo en la alta rotación de productos y escaso contrato de personal para atención. La tienda de este formato con mayor renombre en el país es Mass.
- **Bodegas:** corresponde a las tiendas independientes que se encuentran en todo el país. Usualmente poseen un surtido estrecho y corto y su principal oferta de valor es la proximidad a sus consumidores.

Tabla 2

Tipos de formato para retails de productos generales

Formato	Ubicación	Mercadería	Precios	Servicio	Publicidad
Tiendas especializadas	Zonas concurridas	Surtido largo y muy estrecho de calidad media a alta	Competitivos a encima de la media	Promedio a excelente	Mediante fuerza de ventas
Tiendas por departamento	Centros Comerciales, zonas concurridas	Amplio surtido y calidad media a alta	Promedios a encima de la media	Buena a excelente	Extensiva en medios de comunicación, periódicos, correos electrónicos.
Tiendas en Outlet	Periferias de la ciudad, Centros Comerciales	Surtido escaso y calidad media	Bajos	Promedio o debajo de la media	Moderada
Tiendas en galerías textiles	Centros de comercio	Poco surtido y calidad media	Muy bajos	Debajo de la media a promedio	Poca o ninguna
Tiendas independientes	Vecindarios	Surtido escaso y calidad media	Promedios a encima de la media	Debajo de la media a promedio	Ninguna

Nota. Adaptado de Berman et al. (2018).

- **Tiendas especializadas:** corresponde a las tiendas que adquieren pericia en la comercialización de determinada familia de productos. Los consumidores que las usan no solo buscan un producto, sino un servicio adicional asociado al bien. A modo de ejemplo, orientación sobre las características y

beneficios de los productos o servicio posventa. Algunos ejemplos de estas tiendas en Perú son Tiendas EFE (electrodomésticos) y Nutrishop (implementos deportivos).

- **Tiendas por departamento:** son tiendas de amplia extensión que ofrecen un surtido bastante variado y abundante de productos de diferente índole. Usualmente se encuentran dentro de centros comerciales o en zonas de bastante afluencia de personas. Las más conocidas en Perú son Saga Falabella, Ripley y Paris.
- **Tiendas en outlet:** son aquellas que se establecen en las zonas periféricas a la ciudad y ofertan productos de temporadas pasadas a un precio más bajo que el regular.
- **Tiendas en galerías textiles:** corresponde a las tiendas cuyos dueños frecuentemente son los fabricantes de los productos que comercializan y que se agrupan dentro de emporios comerciales. Por esta razón, los precios que ofrecen son muy bajos. Las galerías de más renombre en el país son las que se ubican en el emporio comercial de Gamarra.
- **Tiendas independientes:** son el símil de las bodegas, pero para productos no alimenticios. El principal valor que ofrecen es la ubicación cercana a los consumidores.

Los formatos de tiendas descritos corresponden al canal tradicional del *retail*, es decir, a las tiendas físicas que los consumidores visitan para comprar los productos que requieren. No obstante, en la actualidad, las ventas mediante el canal *online* del *retail* representan una parte relevante de los ingresos de las empresas en el mundo (Gallino & Moreno, 2014). De forma particular, en el Perú, de acuerdo con la consultora BlackSip (2019), este subsector ha mostrado un crecimiento anual de 26% todos los años desde el 2014 hasta el 2019.

Si bien, algunas empresas *retail* han conformado todo su modelo de negocio en torno al *e-commerce* o ventas por internet (e.g, Linio, Lumingo), la tendencia global es que las compañías busquen ofertar sus productos y servicios a través del canal tradicional y digital a la vez, es decir, volverse empresas multicanales. En ese sentido, en el contexto peruano, el sector *retail* ha incrementado sus ventas *online* considerablemente desde 2012, apalancado en el aumento de acceso a internet por parte de la población (BBVA Research, 2018).

1.2 Administración de servicios

Para Fitzsimmons y Fitzsimmons (2011), los servicios se pueden definir de distintas maneras; no obstante, todas coinciden en dos características que definen su naturaleza: el carácter intangible y el

consumo simultáneo con su creación. A modo de ejemplo, dentro de la revisión de la literatura de Fitzsimmons y Fitzsimmons, los servicios son hechos, procesos y actuaciones (Zeithaml & Bitner, 1996); y son también, una serie de actividades en las que usualmente participan los clientes y los trabajadores para crear una solución mayormente intangible para los clientes (Gronroos, 1990). En particular, la última definición guarda una estrecha relación con la definición proporcionada por Lovelock, Reynoso, D'Andrea, Huete y Wirtz (2011), los cuales añaden a dicha tesis el concepto de valor entregado al cliente.

Ahora bien, al revisar la literatura de la historia de los servicios, sus orígenes se remontan a definirse como actividades que complementan la producción de bienes y que las empresas manufactureras solían tercerizar. A modo de ejemplo, la banca, la distribución o la adquisición de suministros eran operaciones que se subcontrataban, por ello los fabricantes no poseían ningún control sobre su funcionamiento (Lamoreaux, Raff & Temin, 2002). No ocurre, sino hasta la llegada de la Segunda Revolución Industrial (segunda mitad del siglo XIX), que las empresas manufactureras, impulsadas por la creciente demanda de sus productos, comienzan a interesarse y estudiar cómo obtener una participación más activa dentro de la cadena de suministros. En consecuencia, Schmenner (2008) refiere que algunas compañías fabricantes decidieron realizar integraciones verticales con sus proveedores y comenzaron a ofrecer bienes que, en lugar de reducir su valor a la utilidad de estos, formasen parte de servicios agrupados que generen más valor para sus consumidores. Dichas empresas, años más tarde, dominaron sus industrias por décadas.

Lovelock et al. (2011) complementan esta postura al enlistar los factores que contribuyeron a que la concepción de los servicios cambie de la ya mencionada al resultado final de un proceso que puede ser ofertado en el mercado por su propio valor. Los factores en cuestión son la reducción de regulaciones gubernamentales y la privatización de los servicios; los cambios sociales que muestran una mayor capacidad de gasto por parte de los consumidores apalancado en la mayor tasa de empleo; las nuevas tecnologías de información que permite una mejor toma de decisiones; la tendencia de los negocios hacia ofrecer servicios por derecho propio (e.g., consultorías, capacitaciones) y la globalización de las empresas.

Una vez definido el concepto de los servicios, es relevante mencionar las diferencias que implica para una empresa comercializar bienes manufactureros o servicios. En esa línea, Fitzsimmons y Fitzsimmons (2011) sostienen que existen cinco diferencias centrales entre estas producciones, las cuales se resumen en la tabla 3.

Tabla 3***Tipos de formato para retails de productos generales***

Característica	Manufactura de productos	Prestación de servicios
Participación de clientes en el proceso	Los clientes no intervienen en ninguna parte del proceso, por ende, el local donde se produce el bien no está acondicionado para que los clientes se sientan cómodos.	Los clientes suelen ser partícipes activos en la creación y entrega del bien. Por lo tanto, toda la parte frontal del lugar donde se presta el servicio debe estar diseñada para la comodidad del cliente.
Simultaneidad de producción y consumo	Los bienes manufactureros son consumidos una vez que su producción ha sido completamente terminada, por tal motivo, es posible almacenarlos.	Los servicios no se pueden almacenar, dado que su consumo ocurre durante su creación. En ese sentido, no se les puede almacenar; sin embargo, es posible registrar un inventario de todos los insumos que se requieren para entregar el servicio.
Expiración del producto	En una planta manufacturera, se puede controlar el nivel de producción de acuerdo con la demanda estimada. Por ejemplo, en un mes se puede producir demás para almacenar inventarios.	Las empresas que ofrecen servicios no pueden almacenarlos; por lo tanto, se dice que los servicios son perecibles. A modo de ejemplo, si un restaurante cuenta con mesas vacías por una hora, no podrá ofrecer nada de la comida elaborada en esa hora ociosa en otro momento.
Intangibilidad del producto	Al elaborar productos de manufactura, es posible patentar una innovación que resulte en un nuevo producto, ya que, estos son tangibles.	Los servicios no necesariamente están ligados a un producto; sin embargo, es posible proteger y rentabilizar una metodología de desarrollo de un servicio a través de las franquicias. Para ejemplificar este punto, cabe mencionar las cadenas de comida rápida que mediante las franquicias se expanden en distintos mercados sin hipotecar sus metodologías.
Heterogeneidad del proceso	Debido a que el proceso es lineal: los proveedores abastecen a las fábricas; las fábricas producen los bienes y los distribuyen a sus clientes o sus intermediarios. Todo el flujo suele girar en torno al producto <i>per se</i> .	La característica de intangible de los servicios sumado a la participación de los clientes en su creación resulta en que los servicios sean desarrollados en un proceso heterogéneo. En tal línea, estos flujos suelen orientarse hacia las personas, y no hacia los productos que son parte del servicio.

Nota. Adaptado de Fitzimmons y Fitzimmons (2011)

En la actualidad, los servicios son un eje fundamental de la economía mundial, en esa línea, el Banco Mundial (2019), afirma que el sector de servicios representó el 61% del PBI mundial en el 2018; mientras que, en el Perú, dicho sector significó el 54% del PBI en el mismo año. Por su parte, el Instituto Nacional de Estadística e Informática (INEI, 2019) refiere que este sector aportó un 59% al PBI del año 2018, y que, dentro de los servicios, el comercio representó el 20% del total de PBI aportado (véase las figuras 1 y 2 respectivamente). Asimismo, el Ministerio del Trabajo y Promoción del Empleo (2018) sostuvo que, para ese año, el sector servicios le brindó empleo al 40.1% de la PEA (población económicamente activa). Por estas razones, estudiar los servicios ha tomado mayor relevancia. A continuación, se detallarán dos conceptos dentro del marco de la administración de servicios que serán utilizados a lo largo de este estudio.

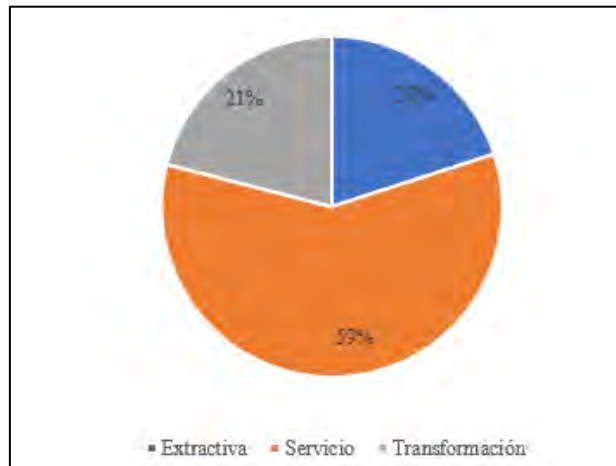


Figura 1. Aporte al PBI peruano por sector en 2018
Fuente: Adaptado de INEI (2019)

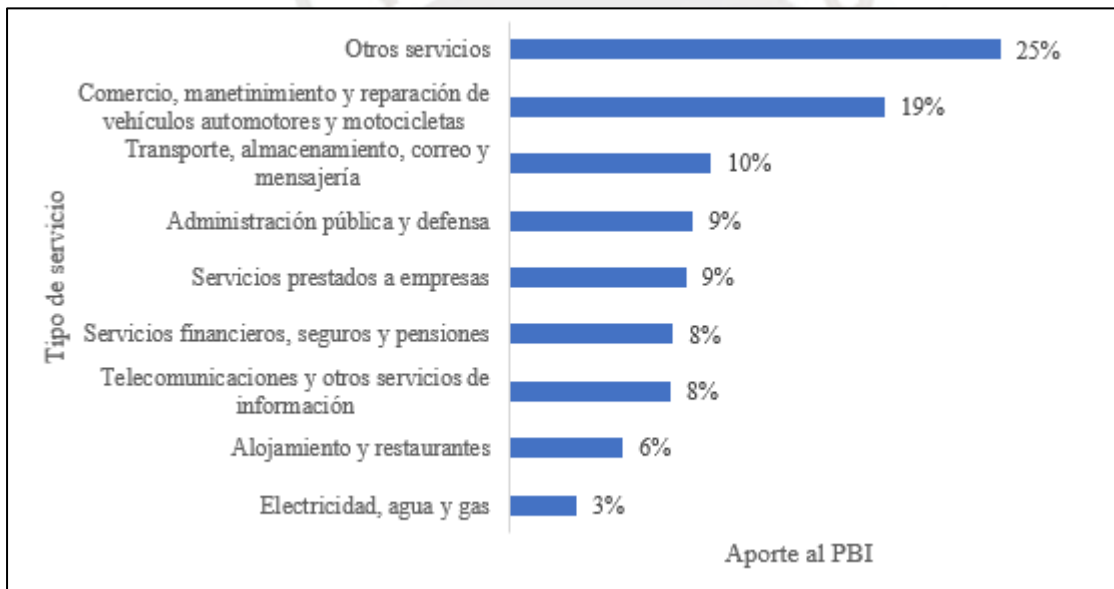


Figura 2. Participación de cada tipo de servicio dentro del aporte al PBI peruano de 2018 del sector servicios
Fuente: Adaptado de INEI (2019)

1.2.1 El servicio como un proceso

De acuerdo con Fitzsimmons y Fitzsimmons (2011), existen diferentes definiciones para los servicios; no obstante, estos autores consideran que estos pueden ser estudiados desde la perspectiva de los procesos. En efecto, Lovelock et al. (2011) y Bitner, Ostrom y Morgan (2008) concuerdan en dicho aspecto. Esto se debe a que los servicios se pueden resumir como una serie de pasos que deben realizarse de forma eficiente para satisfacer al cliente y generar valor. Para ejemplificar lo mencionado, se puede mencionar el caso de un servicio de comida rápida. El servicio se lleva a cabo mediante una secuencia de acciones (subprocesos) que tanto el cliente como la empresa, mediante sus colaboradores, realizan.

Posee un inicio, un fin, recursos asignados para la realización, un objetivo que es satisfacer al cliente y, en última instancia, un producto final: la comida.

En el ejemplo mencionado, aunque el servicio ofrecido estaba fuertemente asociado con un bien tangible y el receptor del valor entregado era una persona, este no es el estándar para todos los tipos de servicios. Así, Fitzsimmons y Fitzsimmons (2011), afirman que existen cuatro tipos de servicios categorizados por la naturaleza de las acciones de valor durante el proceso (tangibles o intangibles) y por el tipo de receptor del servicio (personas o propiedades). El resumen de dicha clasificación se presenta a continuación.

		Receptores directos del servicio	
		Personas	Propiedades
Naturaleza del acto del servicio	Tangibles	Servicios dirigidos al cuerpo de las personas <ul style="list-style-type: none"> - Cuidado de la salud - Transporte de pasajeros - Restaurantes 	Servicios dirigidos a las propiedades físicas <ul style="list-style-type: none"> - Transporte de carga - Mantenimiento y reparación - Lavandería y limpieza
	Intangibles	Servicios dirigidos a la mente de las personas <ul style="list-style-type: none"> - Educación - Entretenimiento - Sesiones psicológicas 	Servicios dirigidos a activos intangibles <ul style="list-style-type: none"> - Banca y seguros - Asesoría legal - Contabilidad

Figura 3. Clasificación de los servicios según la naturaleza de los actos y el tipo de receptor
Fuente: Adaptado de Fitzsimmons y Fitzsimmons (2011)

Como se observa en la figura 3, la clasificación de Fitzsimmons y Fitzsimmons (2011) se ilustra en una matriz de doble entrada. Para estos autores, identificar la clasificación de los servicios que ofrece una empresa es muy importante para administrar correctamente el tipo de entrega de valor que se le hará al cliente. En ese sentido, es distinto un servicio dirigido al cuerpo de una persona que uno dirigido a activos intangibles porque, en el primer caso, el cliente necesariamente debe estar presente en el lugar donde se desarrolle el servicio; por el contrario, en el segundo, la presencia del cliente no es indispensable. Cabe mencionar que Lovelock et al. (2011) proponen una clasificación muy similar a la explicada anteriormente.

1.2.2 Gestión de Relaciones con los Clientes (CRM)

Una vez que se sentaron las bases de la administración de servicios, el estudio de esta se desarrolló en dos frentes. Por un lado, se realizaron investigaciones sobre mejoras de la cadena de suministros para diferentes tipos de servicios; mientras que, por otro, comenzó a crecer el interés por entender a los clientes y sus necesidades con mayor profundidad.

Según la revisión de la literatura de Boulding, Staelin, Ehret y Johnston (2005), desde fines del siglo XX el objetivo de los servicios se enfocó en la generación mutua de valor entre clientes y empresas (e.g., Bagozzi, 1974, Levitt, 1969); así como en crear y mantener buenas relaciones con los clientes (e.g., Berry, 1983; Grönroos, 1994; Morgan y Hunt, 1994). En ese contexto, surge el concepto de “Gestión de Relaciones con los Clientes”, o CRM (por sus siglas en inglés *Customer Relationship Management*), que se refiere a las soluciones informáticas que permitían a las empresas recopilar y ordenar la data que tenían disponible de sus procesos y clientes para luego analizarla.

Aunque dicho concepto es usualmente confundido por otros similares debido a sus diferentes definiciones, Payne y Frow (2005) proponen una definición que se compone de tres factores: solución táctica, tecnología de amplio alcance y centrada en el cliente; los cuales se interrelacionan y se muestran en la figura 4. En consecuencia, la definición que obtuvieron se resume a que el CRM es un enfoque estratégico que busca crear valor a partir de las relaciones con los clientes mediante la combinación de estrategias de marketing y tecnologías de información para entender a los clientes.

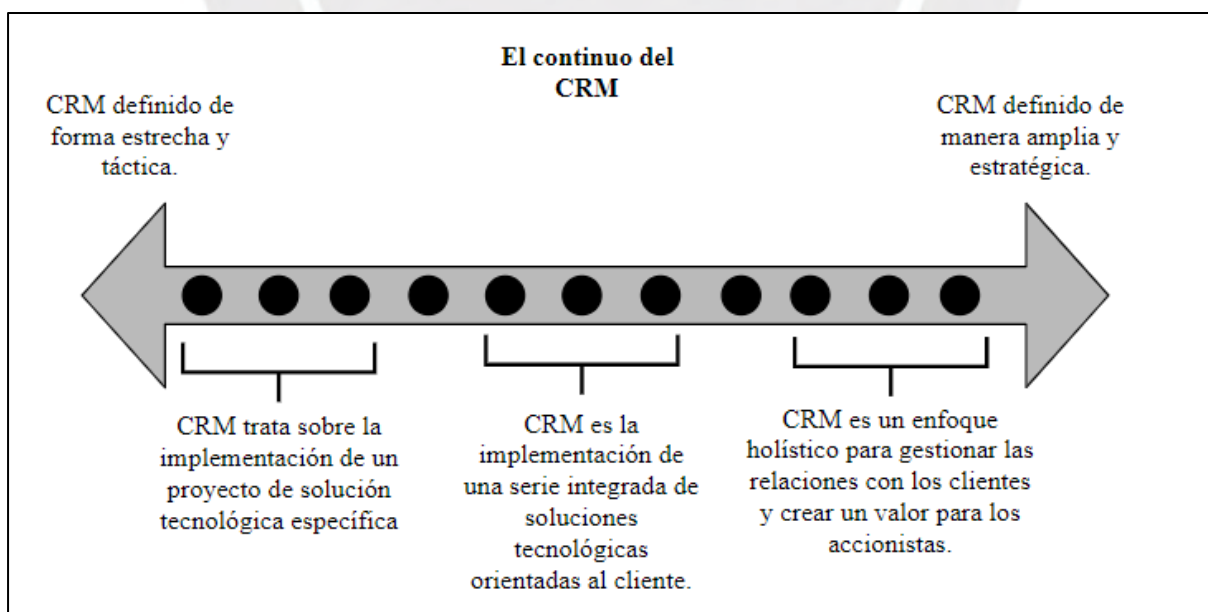


Figura 4. El continuo del CRM
Fuente: Adaptado de Payne y Frow (2005)

Mediante la ejecución del CRM, se establece una creación de valor tanto para los clientes como la empresa. Así, el valor que los clientes reciben radica en la oferta diferenciada que la compañía le ofrece; mientras que el valor que adquiere la empresa se basa en el conocimiento obtenido de sus clientes para mejorar sus estrategias. Este conocimiento es generado a partir del análisis que la compañía puede realizar a la data de compras que sus clientes efectúan en sus plataformas. En ese sentido, Payne y Frow (2005) mencionan dos macro estrategias que se pueden implementar a partir del CRM: la adquisición y la retención de clientes, las cuales se describen a continuación.

- **Adquisición de clientes:** es el primer paso que se implementa con CRM y consiste en capturar información de clientes potenciales (*leads*) que luego se convertirán en clientes. Las tácticas que se suelen para adquirir información son campañas de obtención de datos (usualmente correos electrónicos o teléfonos) para consolidarlos en una base de datos de clientes en potencia. Por otro lado, para que dichas personas realicen su primera compra se realizan ofertas genéricas de productos que suelen comprar los clientes por primera vez.
- **Retención de clientes:** una vez que se adquieren clientes nuevos, es decir, se realizó la primera relación cliente-empresa, se ejecutan acciones orientadas a que el cliente realice una segunda compra y la relación se mantenga en el tiempo. Las tácticas que se suelen aplicar son venta cruzada o *cross sell* (ofrecer promociones de productos de diferentes categorías), venta hacia arriba o *up sell* (colocar ofertas en productos premium o de mayor costo de una misma categoría) y venta hacia abajo o *down sell* (promocionar descuentos en los productos de menor costo de una categoría).

Sin embargo, es posible añadir una macro estrategia más a lo aportado por Payne y Frow (2005). Como concluye Zineldin (2006), la lealtad que un cliente adquiere hacia una empresa se injerta en las relaciones continuas que ambas tienen, es decir, está íntimamente relacionado con el CRM.

- **Fidelización de clientes:** consiste en generar relaciones continuas y de tanto valor con los clientes que estos se vuelvan leales a la empresa. Las tácticas que se usan plantean ofrecer a los clientes los productos que buscan, dado que la empresa está en capacidad de entender al cliente hasta ese punto. Para el caso de los clientes más valiosos para el negocio se busca blindarlos y para los demás, guiarlos hasta obtener el máximo valor de ellos.

Finalmente, se puede concluir que las estrategias de gestión de clientes que el CRM comprende se apalancan en la información generada por sus clientes. En tal sentido, las empresas hacen uso de dicha data, así como de toda la información que tienen a su disposición para entender mejor a sus consumidores y, así, generar relaciones de valor.

Desde otra perspectiva, Payton y Zahay (2003) señalan que la cantidad de información que las empresas pueden gestionar aumenta considerablemente en el tiempo. Una solución informática que surgió en el mercado durante los primeros años del siglo XXI: DW (*Data Warehousing*). Esta propuesta se basa en aglomerar toda la información que una empresa maneje en un solo repositorio desde el cual se puedan ejecutar análisis eficientes a distinto nivel para mejorar la toma de decisiones (Khan, Ehsan, Mirza & Sheikh, 2012).

A pesar de que en los primeros años esta tecnología no era frecuentemente utilizada en marketing (Payton & Zahay, 2003); años después comenzó a ser más utilizada en diferentes empresas.

Esto se puede explicar por los beneficios que Khan et al. (2012) describen y que se ilustran a continuación.

- La información se almacena en un solo lugar y permite que la empresa obtenga información relevante en tiempo real para mejorar la toma de decisiones de una forma eficiente.
- Empodera a los colaboradores para que puedan desarrollar su creatividad a partir de la data disponible y no ejecuten solo operaciones repetitivas.

Columbus (2017) sostiene que, al 2017, el 87% de empresas de telecomunicaciones y el 76% de compañías financieras ejecutan estrategias a partir de análisis de grandes volúmenes de datos con el fin de desarrollar la inteligencia del negocio. Dichos proyectos son posibles gracias a la optimización proporcionada por la DW y, por ello, el uso de esta tecnología se está masificando en las grandes empresas (Columbus, 2017).

1.3 Herramientas de diagnóstico

Para este trabajo de investigación se utilizarán diferentes herramientas que permitan elaborar un diagnóstico de la situación de la empresa objeto de estudio. A continuación, se presenta dichas herramientas.

1.3.1 Cadena de valor

Las organizaciones a menudo suelen tener diversos procesos o actividades que se llevan a cabo en simultáneo para producir el bien o servicio que ofrecen. Por ello, puede resultar complejo entender cómo interactúan estos procesos y distinguir cuáles de ellos realmente crean valor para el cliente. Ante esta situación, Porter (1985) desarrolló la cadena de valor que es una herramienta que permite describir tanto las actividades que una organización realiza para generar valor como las que se ejecutan para garantizar la correcta realización de estas. En la figura 5 se ilustra la cadena de valor según Porter (1985)

y pueden observar dos tipos de actividades; las primarias y las de soporte. A continuación, se describen ambas.

- **Actividades de soporte:** tienen como objetivo dar apoyo a las actividades primarias para garantizar el correcto funcionamiento de la organización. Se clasifican en cuatro tipos: infraestructura de la empresa (procesos de índole financiero, legal e informático); gestión de recursos humanos; desarrollo la tecnología (soluciones tecnológicas que mejoran los procesos) y aprovisionamiento (planeación y compra de materias primas). Cabe mencionar que estas actividades pueden ser tercerizadas porque no generan valor a los clientes directamente.
- **Actividades primarias:** son los procesos que realmente crean valor para el cliente que reciben soporte de las actividades previamente descritas. Se dividen en cinco grupos: logística interior (procesos que se requieren para la recepción, almacenamiento y distribución de insumos); operaciones (procesos necesarios para la creación del bien o servicio); logística exterior (distribución de productos terminados); marketing y ventas, y servicios (actividades de atención posventa).



Figura 5. Esquema de cadena de valor
Fuente: Adaptado de Porter (1985)

1.3.2 Mapeo de procesos

Como se ha mencionado anteriormente, los servicios se comportan como procesos complejos que deben ejecutarse apropiadamente para generar valor al cliente. En tal sentido, cada una de las actividades debe desarrollarse bajo estándares de calidad que garanticen la satisfacción del cliente.

Para el *International Organization for Standardization* (ISO, 2015), implementar un sistema de gestión de calidad brinda diversos beneficios a una organización. Dicho sistema de gestión se implementa en diferentes áreas de la empresa a través de metodologías y herramientas establecidas.

Debido a la esencia de proceso que los servicios poseen, las herramientas recomendadas por la ISO son aplicables y permiten realizar inspecciones en diferentes niveles de análisis. Jacka y Keller (2009) sostienen que, para implementar un cambio en cualquier proceso, es necesario entenderlo plenamente y esto se puede lograr mediante la diagramación de este. Así, una de las herramientas de la ISO que permite diagramar un servicio de una empresa de manera global es el mapeo de procesos.

Según el ISO (2015), existen tres tipos de procesos, los cuales se describen a continuación.

- **Procesos estratégicos:** son aquellos que se encuentran directamente relacionados con los objetivos estratégicos de la empresa y representan los lineamientos sobre los cuales se ejecutarán los demás.
- **Procesos operativos:** son los procesos que ejecutan directamente el bien o servicio que la empresa ofrece. En estos procesos es en donde se entrega el valor que luego será dirigido a los clientes.
- **Procesos de soporte:** son aquellos que dan apoyo a los procesos operativos para su correcto funcionamiento.

Con base en esta clasificación en la figura 6 se muestra un arquetipo de mapa de proceso.

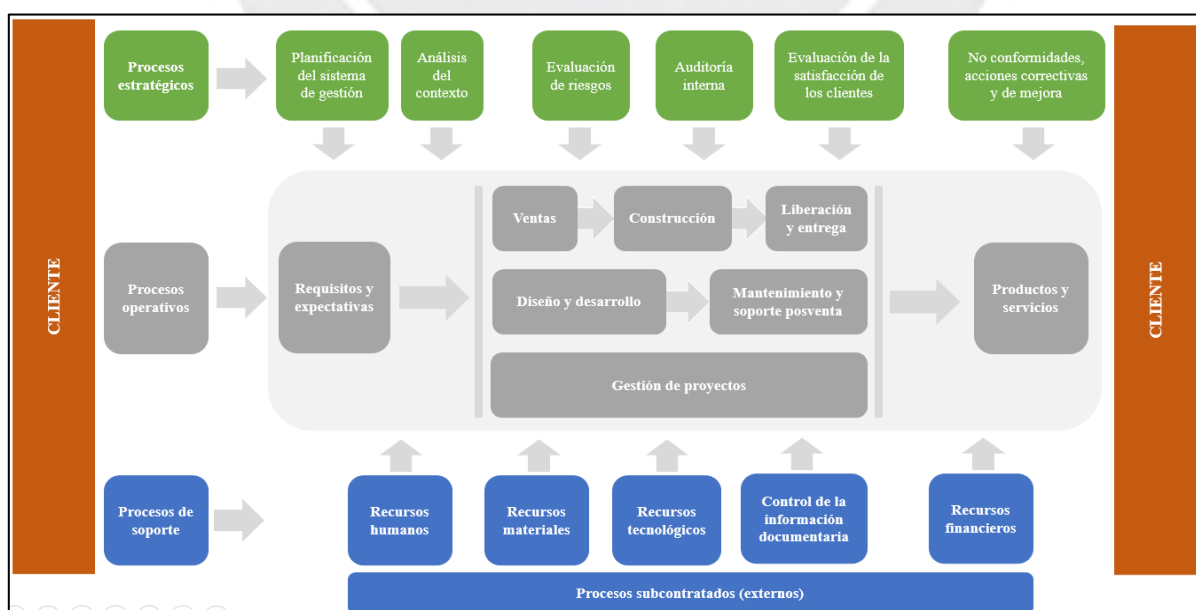


Figura 6. Ejemplo de mapa de proceso
Fuente: Adaptado de ISOTools Excellence (2016)

Finalmente, esta herramienta permite identificar y clasificar los procesos más importantes de una empresa, así como observar la interacción entre ellos.

1.3.3 Flujogramas

Los flujogramas son representaciones gráficas de una secuencia de acciones o pasos (Manene, 2011). En ese sentido sirven para diagramar los procesos complejos de manera más sencilla, por eso esta herramienta será utilizada para describir el estado actual de la empresa a estudiar y se utilizará la simbología de *American National Standard Institute* (ANSI) cuyo detalle se muestra a continuación en la figura 7.





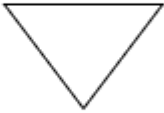
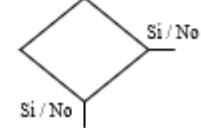
Símbolo	Significado
	Inicio / Fin
	Operación / Actividad
	Documento
	Datos
	Almacenamiento / Archivo
	Decisión

Figura 7. Simbología ANSI

Fuente: Adaptado de Ministerio de Planificación Nacional y Política Económica (2009)

1.3.4 Matriz QFD

La matriz QFD (*Quality Function Deployment*), o matriz de calidad, es una herramienta que permite relacionar los requerimientos específicos de los clientes dentro del diseño y fabricación de un bien o servicio con las características técnicas que la empresa dispone (Cristiano, Liker & White III, 2000).

Esta matriz cuyo origen se encuentra en la década de 1960 en Japón suele ser utilizada en programas de gestión de calidad; sin embargo, según Yacuzzi y Martín (2003), su aplicación no es exclusiva en la industria de la manufactura, sino también, en el análisis de servicios y formulación de estrategias corporativas.

En ese sentido, si bien hay diferentes versiones de la matriz según la aplicación requerida, la idea central de la herramienta es contrastar requerimientos (con una priorización establecida) versus características dadas para identificar cuáles atributos son los más importantes. En la figura 8 se muestra el modelo de matriz QFD aplicado específicamente para el análisis de procesos de un negocio.

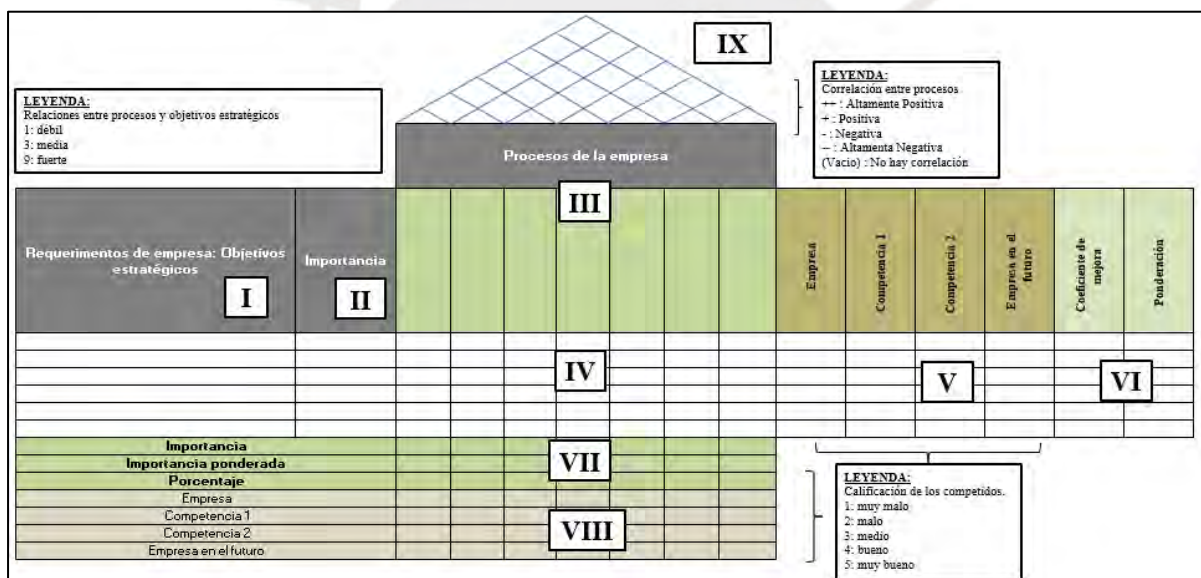


Figura 8. Matriz QFD para negocios

En la figura 8, se pueden notar los requerimientos alineados con los objetivos estratégicos de la empresa (I), la importancia de cada uno de estos (II), los procesos que la empresa tiene a disposición para crear valor a los clientes (III), un puntaje que califica cuán alineados están los procesos que crean valor con los requerimientos estratégicos (IV), comparaciones entre la calificación de la empresa analizada y su competencia basada en la calificación obtenida en cada requerimiento y en el desempeño de cada proceso (V y VIII respectivamente), el coeficiente de mejora de la empresa y el peso ponderado de los requerimientos (VI), la importancia simple y ponderada de cada proceso (VII) y la correlación entre procesos, es decir, el impacto que la mejora de uno produce en otro (IX).

Asimismo, esta herramienta permite analizar con más detalle los resultados obtenidos en su primera versión. A través de la creación de más matrices de menor complejidad, es posible profundizar qué subproceso, es decir, qué parte de todo el proceso elegido es de mayor relevancia.

1.3.5 Proceso de análisis jerárquico

Al realizar un diagnóstico a un sistema es común encontrar diferentes problemáticas que, debido a los recursos limitados, no se pueden solucionar todos a la vez. Frente a esta situación, surge la necesidad de una herramienta que permita decidir objetivamente cuáles se deberían priorizar.

El proceso de análisis jerárquico (AHP por sus siglas en inglés) cumple esta función y consiste en la medición relativa de las alternativas. Brunelli (2015) sostiene que dada una serie de alternativas sobre las cuales se debe crear una priorización, se puede utilizar AHP a través de la matriz de comparaciones pareadas. Este método consiste en enfrentar todas las alternativas y asignar un valor que represente cuánto más importante es una respecto a la otra con el fin de conseguir un objetivo.

Sean las alternativas agrupadas en el vector $X = \{x_1, x_2, \dots, x_n\}$, se creará la matriz de comparaciones pareadas $A = (a_{ij})_{n \times n}$ que se muestra a continuación.

$$A = \begin{matrix} & \begin{matrix} x_1 & x_2 & \dots & x_n \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_3 \end{matrix} & \begin{pmatrix} a_{11} & a_{12} & \dots & a_{n1} \\ a_{21} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \end{matrix}$$

En esta matriz se registran los resultados de los enfrentamientos que siguen la condición de

$$a_{ij} = 1/a_{ji}, \quad \forall i, j.$$

En otras palabras, existe una condición de reciprocidad. Por ejemplo, si la alternativa x_2 tiene una relevancia de 5 sobre la alternativa x_3 , entonces la importancia de x_3 sobre x_2 será de $1/5$.

Luego, se debe calcular la matriz normalizada M que se detalla a continuación.

$$M = \begin{pmatrix} \frac{a_{11}}{\sum_{i=1}^n a_{i1}} & \frac{a_{12}}{\sum_{i=1}^n a_{i2}} & \dots & \frac{a_{1n}}{\sum_{i=1}^n a_{in}} \\ \frac{a_{21}}{\sum_{i=1}^n a_{i1}} & \frac{a_{22}}{\sum_{i=1}^n a_{i2}} & \dots & \frac{a_{2n}}{\sum_{i=1}^n a_{in}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{a_{n1}}{\sum_{i=1}^n a_{i1}} & \frac{a_{n2}}{\sum_{i=1}^n a_{i2}} & \dots & \frac{a_{nn}}{\sum_{i=1}^n a_{in}} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ m_{21} & m_{22} & \dots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \dots & m_{nn} \end{pmatrix}$$

Finalmente, para obtener la priorización se calcula el vector v que se muestra a continuación.

$$v = \begin{pmatrix} \frac{\sum_{i=1}^n m_{1n}}{n} \\ \frac{\sum_{i=1}^n m_{2n}}{n} \\ \vdots \\ \frac{\sum_{i=1}^n m_{3n}}{n} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

La priorización se dará según los valores del vector v de forma descendente.

1.3.6 Diagrama de Ishikawa

Lleva su nombre en honor al Dr. Kouro Ishikawa y también es conocida como el diagrama de causa y efecto. Su objetivo es mostrar la relación entre un resultado y sus posibles causas (Magar y Shinde, 2014); en la figura 9 se presenta un modelo.

Esta herramienta plantea el siguiente procedimiento para su elaboración.

- Definir el resultado o efecto que se diagnosticará.
- Determinar las categorías principales que agrupan diferentes causas.
- Efectuar una lluvia de ideas para enlistar las posibles causas que provocan la situación actual.
- Realizar una segunda lluvia de ideas para determinar el origen de las causas halladas.
- Cuando se haya completado el diagrama, se deben identificar las causas raíz de mayor relevancia.

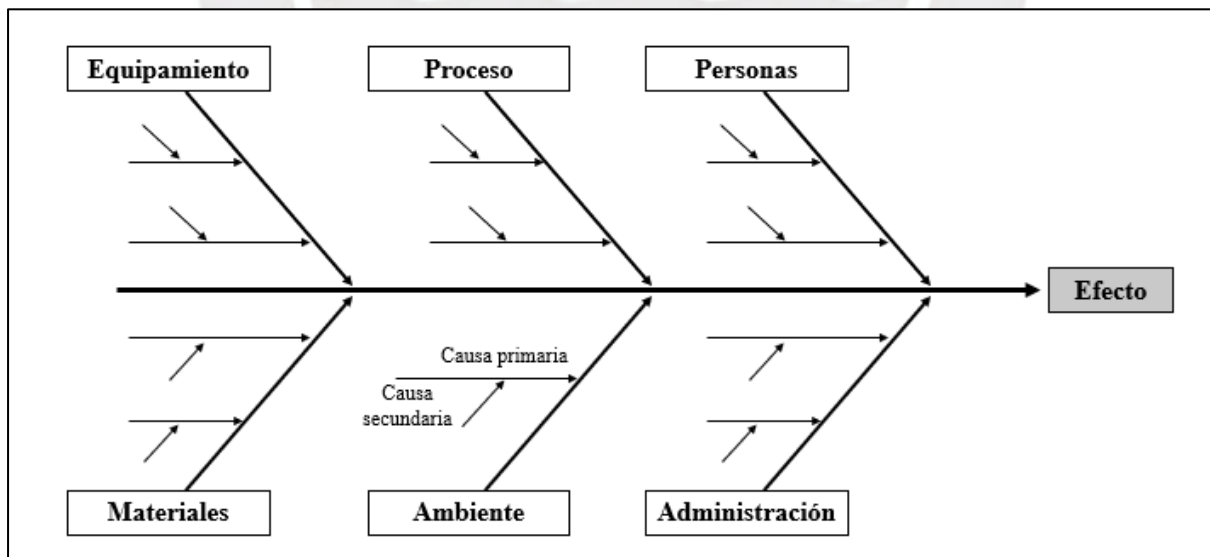


Figura 9. Diagrama de Ishikawa
Fuente: Adaptado de Magar y Shinde (2014)

1.3.7 Matriz de probabilidad - impacto

El *Project Management Institute* (PMI, 2017) define al riesgo como todo evento incierto que de ocurrir afectará positiva o negativamente los objetivos de un proyecto. Por ello, para efectos de la gestión de proyectos y, específicamente, en materia de riesgos, el PMI (2017) define diferentes herramientas que se pueden utilizar para gestionar dichos riesgos.

Entre las técnicas en cuestión se encuentra la matriz de probabilidad - impacto que, a través de un análisis de la probabilidad de ocurrencia y el impacto que provocaría la concretización del evento, cuantifica el riesgo. Dicho cálculo se presenta a continuación.

$$\text{Riesgo} = \text{Probabilidad} * \text{Impacto}$$

La probabilidad de que el evento ocurra se estima en una escala numérica entre los valores {0.1, 0.3, 0.5, 0.7, 0.9} y el impacto se califica entre los valores {0.05, 0.1, 0.2, 0.4, 0.8} de acuerdo con la experiencia del sujeto que realiza el análisis. Esta herramienta es comúnmente usada para evaluar riesgos en campos de ingeniería (Ruan, Yin & Frangopol, 2015) y puede ser usada en la toma de decisiones cuando se requiere priorizar una problemática sobre otras. En este último sentido se aplicará esta técnica en el desarrollo de la tesis.

1.3.8 5 por qué

Esta técnica es parte de las metodologías propuestas por Taiichi Ohno, es decir, el padre del Sistema de Producción de Toyota que sentó las bases del modelo de producción *Just in Time* (JIT) de la metodología Lean de mejora continua en la industria manufacturera (Alukal, 2007). Dado que los errores siempre van a ocurrir, la mejor forma de abordar estas situaciones es encontrar las causas raíz mediante la consulta sucesiva de por qué ocurre la situación. La herramienta en cuestión plantea identificar un error, consultar por qué ocurre esta situación y determinar la razón fundamental, luego se repite el proceso cuatro veces más hasta identificar la causa original del problema. Cabe indicar que los resultados que esta herramienta obtiene pueden ser la base de una planificación de mejora inmediata y a mediano o largo plazo. En ese sentido, Murugiah, Benjamin, Marathamuthu y Muthaiyah (2009) sostienen que si este análisis se realiza correctamente puede usarse tanto para toma de acciones correctivas como preventivas. En concreto, esta metodología será utilizada para el desarrollo de la tesis.

1.4 Herramientas del modelo

Con el objetivo de desarrollar una propuesta de mejora con base en el diagnóstico efectuado a la empresa se hará uso de diferentes herramientas para desarrollar el modelo matemático que permita generar dicha propuesta. Por ello, a continuación, se detallarán las herramientas en cuestión.

1.4.1 Machine learning

El aprendizaje automático o *machine learning* (ML) es la ciencia de la programación de sistemas de tal forma que estos son capaces de aprender a partir de la data que poseen (Géron, 2017). En la actualidad, esta metodología diferentes aplicaciones en diferentes industrias de bienes y servicios. Así, por ejemplo, se ha usado para predecir diagnósticos de enfermedades como el cáncer (e.g., Bibault, Giraud & Burgun, 2016; Cruz & Wishart, 2006) o inferir el comportamiento animal (e.g., Valletta, Torney, Kings, Thornton & Madden, 2017; Wang, 2019); para identificar el medio de transporte que un usuario está usando (e.g., Fang, Fei, Xu & Tsao, 2017; Jahangiri & Rakha, 2015) o en la mejora de campañas de marketing en empresas de servicio (e.g., Cui, Wong & Lui, 2006; Jordan & Mitchell, 2015; Ramageri & Desai, 2013).

De acuerdo con Hurwitz y Kirsch (2018), el ML posee dos enfoques: descriptivo y predictivo. El primero se basa en comprender la situación actual. En el campo empresarial, es fundamental comprender en dónde se encuentra la compañía en el presente y así, establecer objetivos realistas. El segundo dispone del conocimiento recogido en el análisis descriptivo para anticiparse al futuro mediante la predicción de escenarios y resultados utilizando algoritmos de ML.

En esa línea, el desarrollo de ML se puede definir como un proceso que inicia en la captura y procesamiento de información, seguido de la elaboración de un análisis descriptivo que explique la realidad y un entrenamiento de este mediante algoritmos y, finalmente, elaborar predicciones con el modelo final. Asimismo, IBM (s.f.), sostiene que esta metodología se fundamenta en tres pilares del análisis de datos: la clasificación (división de información); la regresión (búsqueda de interrelaciones) y la segmentación (agrupación por similitud).

Hurwitz y Kirsch (2018) proponen tres clasificaciones para los modelos de ML, los que se describen a continuación.

- **Aprendizaje supervisado:** consiste en el reconocimiento de patrones que son tomados como base para elaborar el modelo predictivo. Se suele usar cuando hay un conocimiento previo de la información que se trabajará y el algoritmo que se suele utilizar es la regresión (lineal, simple o

múltiple, logística, entre otras) que se entrena con información del pasado. Una aplicación de este modelo es la predicción de la próxima compra de un cliente en una empresa.

- **Aprendizaje no supervisado:** es el tipo de ML utilizado cuando se cuenta con grandes volúmenes de data que aún no se comprende. En este caso, el proceso es capaz de entender la información sin necesidad de la intervención humana y es muy útil en análisis dentro del sector salud o para segmentar clientes de una empresa.
- **Aprendizaje por reforzamiento:** se define como el aprendizaje por ensayo y error. A diferencia de los aprendizajes descritos anteriormente, este ML aprende a través de la retroalimentación que obtiene como resultado de una decisión que toma. Los ejemplos más comunes de este aprendizaje son los modelos para juegos como ajedrez.
- **Redes neuronales y aprendizaje profundo:** se basan en las redes neuronales que posee el cerebro humano y suelen usarse cuando se posee data no estructurada, es decir, no organizada. Los modelos de redes neuronales cuentan con capas de entrada y salida de datos, y de modificación de estos en función a nodos interconectados. El aprendizaje profundo se puede definir como una versión más compleja de las redes neuronales que integra tanto el aprendizaje supervisado, como el no supervisado.

Por último, se describirán los algoritmos de ML más utilizados según Hurwitz y Kirsch (2018).

- **Bayesiano:** el algoritmo Bayesiano permite codificar entendimientos o conocimientos previos de cómo debería verse el modelo. Son particularmente usados cuando no se cuenta con gran cantidad de información para el entrenamiento del modelo.
- **Clustering (agrupamiento):** son algoritmos de fácil ejecución que permiten encontrar elementos dentro del conjunto de datos que están relacionados. Los elementos que pertenecen al mismo *cluster* poseen más cosas en común que otros de diferentes *clusters*.
- **Árboles de decisión:** se estructuran a modo de ramificación para mostrar los resultados que se pueden generar en diferentes situaciones. Para este algoritmo se requieren de nodos que contienen las probabilidades de que ocurra un evento u otro y se suele usar para campañas de marketing.
- **Redes neuronales y aprendizaje profundo:** busca representar la forma en que un cerebro funciona a través de nodos interconectados. Las redes neuronales cuentan con una capa oculta y si

el algoritmo cuenta con más de una capa se vuelven algoritmos de aprendizaje profundo. Suele usarse para reconocimiento de imágenes, voz, data no estructurada en general.

- **Regresión lineal:** estos algoritmos son usados a menudo para análisis estadísticos y miden la correlación entre las variables del conjunto de datos. Asimismo, son usados para predecir valores futuros a partir de data histórica.

1.4.2 Data Mining

Data Mining (DM) o minería de datos es el proceso de descubrimiento de patrones y conocimientos a partir de grandes cantidades de datos. (Han, Kamber, & Pei, 2012). En ese sentido, esta herramienta ha sido constantemente utilizada en diferentes campos de investigación y laborales, así como en herramientas aplicativa en manufactura, banca, contabilidad y turismo, entre otros (Plotnikova, Dumas y Milani, 2020).

Para algunos autores, de manera análoga a ML, DM posee el enfoque descriptivo y el predictivo (Han et al, 2012; Hand, Mannila, & Padharic, 2001). El primero busca describir y explicar la situación actual, mientras que el segundo tiene como objetivo predecir algún patrón a través del modelo creado en la fase descriptiva. La diferencia entre ambos radica en que los modelos de DM no son capaces de aprender a menos que un usuario entrene la data.

En consecuencia, es razonable confundir los límites de estudio que DM presenta en comparación con ML, si acaso existen. En el presente trabajo no se busca plantear dichas fronteras; sin embargo, es preciso indicar la forma en que se usará cada herramienta.

Para el desarrollo de la tesis se entenderá a DM desde la perspectiva de la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*). Esta perspectiva, según Chapman et al. (2000) define a los proyectos de DM como un flujo continuo (véase la figura 10) que en esencia consta de seis etapas las cuales se describen a continuación.

- **Entendimiento del negocio:** esta fase se enfoca en comprender los objetivos y requerimientos del proyecto desde una perspectiva de negocio para convertirlo en un problema de DM. Cabe resaltar que los encargados de realizar esta función deben poseer conocimientos mixtos tanto del negocio como de DM.

- **Entendimiento de la información:** el siguiente paso es comprender de manera general la información con la que se cuenta. Esto se realiza mediante análisis exploratorios iniciales que pueden mostrar indicios importantes sobre los datos.
- **Preparación de la información:** preparar la data comprende las tareas de selección de información (que usualmente está en tablas), atributos, registros; limpieza de datos y transformación de los mismos si se requiere. El objetivo es obtener el conjunto de datos listo para los análisis.
- **Modelamiento:** la siguiente etapa consiste en seleccionar y aplicar diferentes modelos con el ajuste de parámetros que sean necesarios. Cabe mencionar que la naturaleza de algunas herramientas de modelaje requiere una presentación especial de los datos, por lo tanto, se puede requerir volver a la preparación de información.
- **Evaluación de resultados:** es una etapa crucial que se realiza cuando el modelo ya está construido. El mejor modelo será el que obtenga mejores resultados de acuerdo con la necesidad del negocio y no necesariamente el más sofisticado. En este paso, se evalúa si el modelo funciona bien o no.
- **Despliegue de información:** una vez que se obtenga el modelo final que cubra todas las necesidades del negocio (o al menos tanto como sea posible), se debe presentar la información de valor que se obtuvo a las partes interesadas del proyecto.

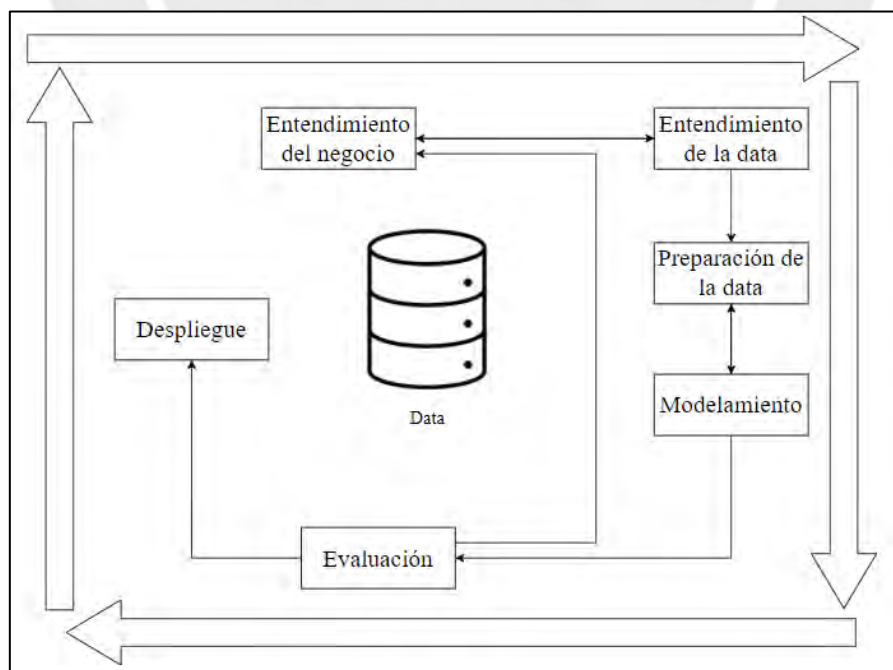


Figura 10. Flujo de proyectos DM
Fuente: Adaptado de Chapman, et al. (2000)

Cabe señalar que, las herramientas para el modelamiento de la data son algoritmos que, en la mayoría de los casos, son iguales a los descritos en la sección de ML.

En la tesis, cuyo marco teórico es este trabajo de investigación, se usará la herramienta CRISP-DM como herramienta central y, luego de evaluar la data resultante, se usará ML si se requiere una mejora con aprendizaje no supervisado.

1.4.3 *Clustering* (agrupamiento)

Como se explicó en el acápite 1.4.1, el *clustering* es un algoritmo de ML recurrentemente utilizado y de fácil ejecución (Hurwitz y Kirsch, 2018). Charu y Chandan (2014) complementan esta información al explicar que su amplio uso se debe, también, a que la implementación de dicho algoritmo, respecto a otros, es más fácil en variados campos de aplicación. Además, los autores proponen una clasificación para los distintos métodos de *clustering*, los cuales se describen a continuación.

- **Métodos de partición (*partitioning methods*):** son los métodos que buscan descubrir las agrupaciones internas en la data mediante la optimización de una función objetivo y, de forma iterativa, mejorar la calidad de los grupos definidos. Según Charu y Chandan (2014), el algoritmo más utilizado de este método es *K-means clustering*.

Este algoritmo inicia con la definición de k centroides desde los cuales se formarán los primeros *clusters*. Luego, se ejecuta un proceso iterativo que consiste en recalcular los centroides mediante la minimización de la suma de errores cuadrados. Las iteraciones se repiten hasta que se cumple la condición de convergencia o la variación en las coordenadas de los centroides sea insignificante (menor a 1%) entre iteraciones (Charu & Chandan, 2014). Un resumen del algoritmo se presenta en la tabla 4.

De acuerdo con Charu y Chandan (2014), el desempeño de este algoritmo es impactado en gran medida por dos factores: los centroides iniciales y los K *clusters* que se definen. En ese sentido, los autores sostienen que los centroides iniciales suelen definirse mediante números aleatorios generados durante el estudio; mientras que la estimación del número de *clusters* se puede realizar mediante métodos como el gap estadístico o el coeficiente de Silhouette.

Tabla 4

Algoritmo K - means clustering

Algoritmo
1. Seleccionar los k puntos como centroides iniciales.
2. Repetir. 2.1. Formar k <i>clusters</i> al asignar cada punto a su centroide más cercano. 2.2. Recalcular el centroide de cada <i>cluster</i> .
3. Hasta que el criterio de convergencia es alcanzado.

Nota. Adaptado de Charu y Chandan (2014)

- **Métodos de jerarquía (*hierarchical methods*):** estos métodos fueron elaborados como una alternativa de *clustering* que no requiere una definición previa del número de *clusters* (Charu & Chandan, 2014). James, Witten, Hastie y Tibshirani (2013) resaltan una ventaja adicional de estos métodos sobre los de partición: la capacidad de ofrecer un gráfico basado en árboles llamado *dendogram* (véase figura 11).

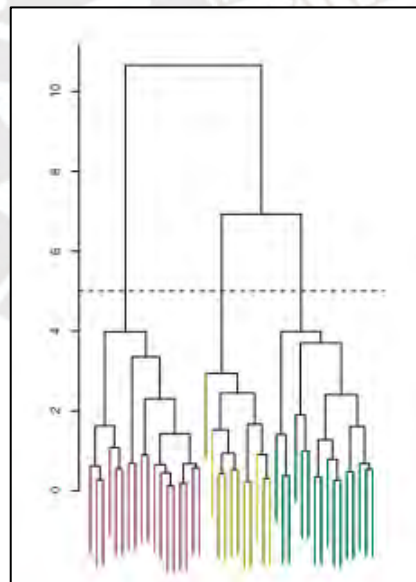


Figura 11. Dendogram de tres clusters
Fuente: Tomado de James et al. (2013)

Charu y Chandan (2014) clasifican los métodos jerárquicos en aglomerantes y divisivos. Los primeros efectúan una jerarquización de abajo hacia arriba (*bottom-up*), ya que forma *clusters*

individuales que luego combinará de acuerdo con sus similitudes. El algoritmo utilizado para este método se ilustra en la tabla 5.

Tabla 5

Agglomerative hierarchical algorithm

Algoritmo
<ol style="list-style-type: none"> 1. Calcular la matriz de disimilitud de todos los puntos. 2. Repetir <ol style="list-style-type: none"> 2.1. Combinar grupos de la forma $C_{(a \cup b)} = C_a \cup C_b$ y establecer la cardinalidad del clúster en $N_{(a \cup b)} = N_a + N_b$. 2.2. Agregar una nueva fila y columna a la matriz con las distancias entre el nuevo <i>cluster</i> $C_{(a \cup b)}$ y el resto. 3. Hasta que solo quede un <i>cluster</i> máximo.

Nota. Tomado y adaptado de Charu y Chandan (2014)

Los segundos, en cambio, efectúan una jerarquización de arriba hacia abajo (*top-down*), dado que comienza con un *cluster* macro que luego divide en *clusters* más específicos. El algoritmo que se utiliza para desarrollar este método se presenta en la tabla 6.

Tabla 6

Divisive hierarchical algortihm

Algoritmo
<ol style="list-style-type: none"> 1. Comenzar con el nodo raíz que consta de todos los puntos dados. 2. Repetir <ol style="list-style-type: none"> 2.1. Dividir el nodo padre en dos partes: C_1 y C_2 usando el bisector <i>K-means</i> para maximizar la distancia de Ward. 2.2. Construir el <i>dendogram</i> y de los <i>clusters</i> obtenidos elegir el que tenga mayor error cuadrático. 3. Hasta que se obtienen <i>Singleton leaves</i>.

Nota. Adaptado de Charu y Chandan (2014)

1.5 Softwares de aplicación

Según De Leeuw (2011), la vasta cantidad de data que hoy se captura y almacena ha provocado la necesidad imperiosa de utilizar *softwares* estadísticos con mayor capacidad de procesamiento que en el pasado. Frente a esta demanda, en la actualidad existen muchos *softwares* que se utilizan en diferentes campos y con objetivos distintos. Al respecto, Muenchen (2019) realizó un análisis de la popularidad que poseen los *softwares* con base en diferentes criterios. En primer lugar, Muenchen estudió cuáles eran los *softwares* más demandados para el campo laboral de ciencias de datos y los más requeridos fueron Python, SQL y Java. El detalle de los resultados se ilustra en la figura 12.

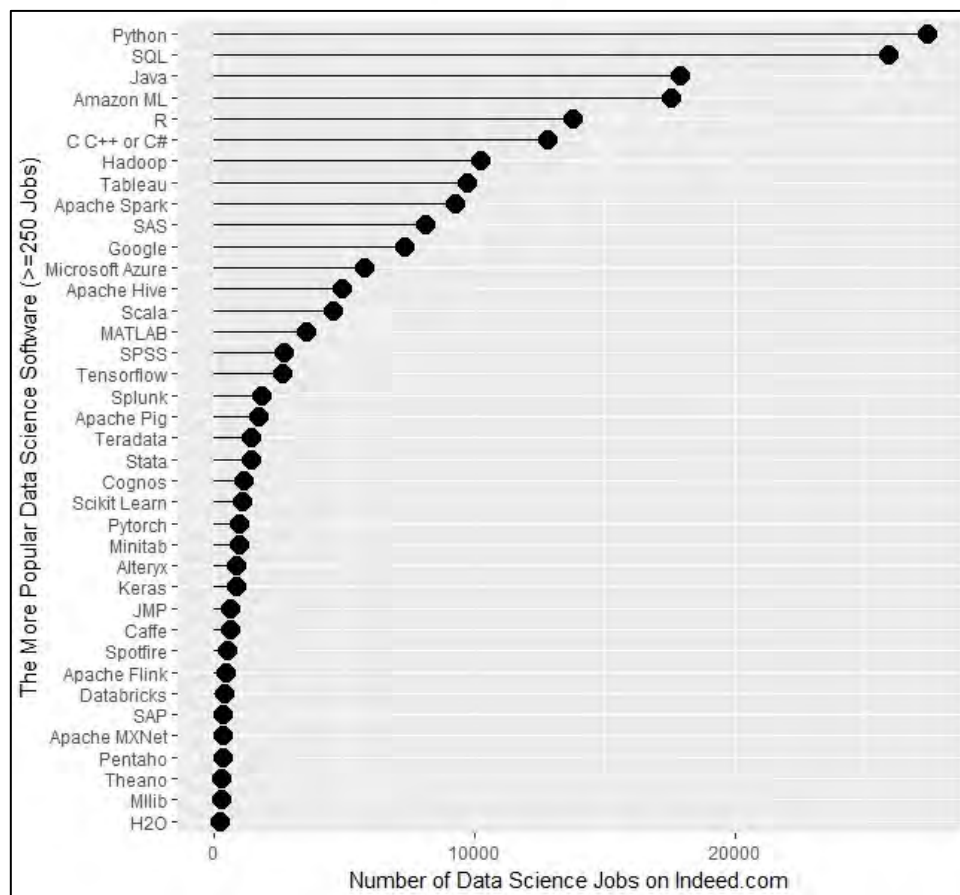


Figura 12. Softwares más requeridos según la oferta laboral
Fuente: Tomado de Muenchen (2019)

En segundo lugar, con el criterio del número de artículos académicos en el que los *softwares* fueron usados, aquellos que fueron más populares son SPSS Statistics, R y SAS. El detalle se presenta en la figura 13.

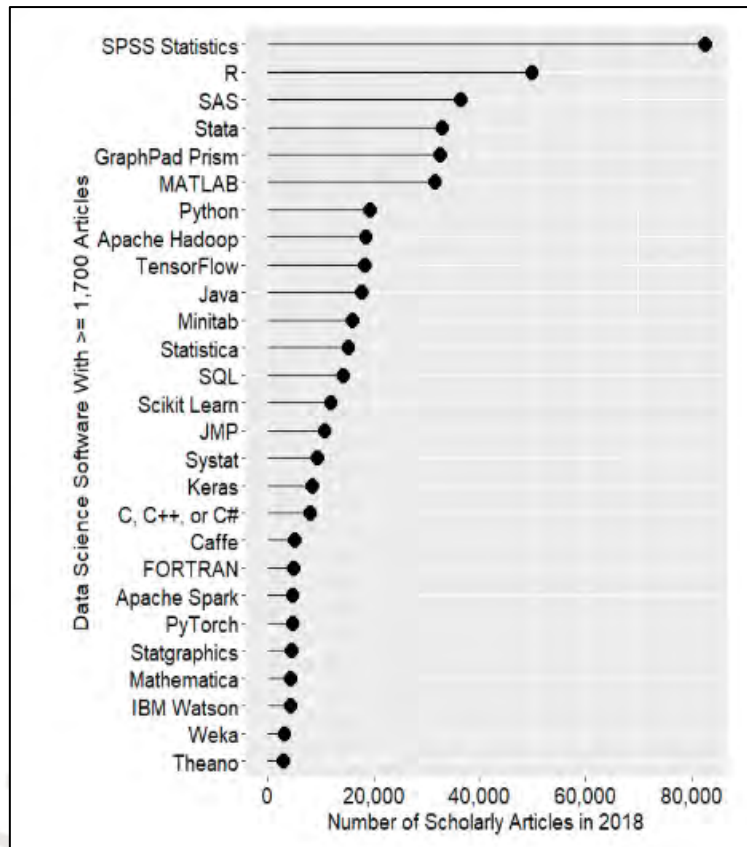


Figura 13. Softwares más populares según número de estudios académicos
Fuente: Tomado de Muenchen (2019)

Si bien, las herramientas de *machine learning* se pueden aplicar en diferentes *softwares*, la tesis contemplará el uso exclusivo de R, Python y SQL. Los dos primeros porque son los más usados en la actualidad y de código abierto, es decir, gratuitos. En cambio, el tercero es requerido para la extracción de información de las bases de datos de la empresa.

1.5.1 R

Es un entorno y un lenguaje de programación orientado al análisis estadístico. *The R Foundation* (s.f.) detalla que, dentro de su entorno, R facilita el uso de diferentes tipos de gráficas y elementos estadísticos (regresiones lineales o no lineales, pruebas estadísticas, análisis de distribuciones, entre otros). De acuerdo con Brownlee (2019), R cuenta con beneficios claves en la realización de trabajos académicos los cuales se describen a continuación.

- **Ser de fuente abierta:** R se puede descargar de la página oficial de *The R Foundation* sin costo alguno y con todas las funciones básicas habilitadas para su uso inmediato.

- **Poseer un gran número de librerías:** una de las mayores ventajas de R respecto a otros códigos es que cuenta con una vasta colección de paquetes que se pueden anexar a las funciones básicas del software para realizar análisis específicos.
- **R es un *software* maduro:** los inicios de R se remontan al lenguaje estadístico S, del que obtuvo aprendizajes para realizar una versión mejorada.

Por otro lado, Brownlee (2019) también menciona las principales debilidades del *software*, las cuales se explican a continuación.

- **Escasa o nula estandarización entre paquetes:** dado que cada paquete dentro de R establece sus propios parámetros incluso para operaciones iguales, se hace necesario revisar la documentación de cada paquete.
- **Documentación poco práctica:** si bien existe mucha documentación que explican los paquetes, esta rara vez logra ayudar a los lectores para las casuísticas que experimentan y deben buscar casos de uso similares en la web.
- **Poca escalabilidad:** R no está programado para funcionar a través de más de una computadora, ni capacitado para trabajar con grandes cantidades de datos que vayan más allá de la memoria del dispositivo desde el que se ejecuta.

1.5.2 Python

Python es un entorno y un lenguaje de programación paradigma (i.e., soporta la programación orientada a objetos, la imperativa y la funcional) que permite trabajar con sistemas integrados rápida y eficientemente (Python™, s.f.). Esta herramienta es muy versátil y sus aplicaciones llegan a los campos de la ciencia, de desarrollo web, empresarial, entre otros. La gestión de este *software* está a cargo de *Python Software Foundation*, el cual lo facilita de manera gratuita.

De acuerdo con BBVA API_Market (2016), cada vez más empresas están añadiendo programadores de Python a sus equipos de trabajo. En esa misma línea, a continuación, se detallan los mayores beneficios de usar este *software* en comparación con R.

- **Librerías diversificadas:** Python, de manera similar a R, posee muchas librerías. No obstante, a diferencia de R, estas no solo se utilizan para análisis estadísticos, sino que son aplicables a otros campos de investigación o laborales.

- **Mayor velocidad de operación:** Python posee una mayor velocidad de ejecución que R.
- **Capacidad de ejecutar modelos con gran cantidad de data:** debido a la naturaleza de Python, es posible realizar proyectos de ML en su plataforma; mientras que en R se podría complicar.

Asimismo, se menciona que una debilidad de Python (al igual que en R) es que su documentación no es buena y que los programadores suelen quejarse por este motivo.

En el desarrollo de la tesis, debido a la naturaleza de las herramientas de mejora que se plantean utilizar, se usará el lenguaje R. Este *software* y sus paquetes estadísticos permitirán ejecutar los diferentes análisis que se requieren.

1.5.3 SQL

SQL (*Structured Query Language*) es un lenguaje estándar que permite el acceso, consulta y gestión de bases de datos. Para los objetivos del ML es muy importante ser capaz manejar grandes cantidades de información y preparar dicha data para hacer uso de las herramientas.

La estructura de SQL se basa en las bases de datos relacionales las cuales tienen como principio contar con un identificador clave que permitirá relacionar características o variables de diferentes bases para agrupar información según se requiera. En la tesis, se utilizará grandes volúmenes de información que están coleccionadas en diferentes bases de datos, por lo que será imprescindible utilizar este lenguaje.

Capítulo 2. Estado del arte

En este capítulo se presentan casos de estudio de aplicaciones del CRM en las decisiones estratégicas que las empresas pueden tomar y su relación con las herramientas de DM y ML dentro del sector *retail*. En particular, se ilustra el uso de la técnica de *clustering* para segmentar clientes, así como los resultados obtenidos por los autores citados.

De acuerdo con Bai y Qin (2016), los clientes y sus necesidades han cambiado considerablemente durante la última década. En la actualidad, los consumidores están más expuestos a diferentes ofertas y sus necesidades cambian rápidamente. Por ello, las empresas han replanteado sus prioridades y progresivamente han colocado al cliente en el centro de sus operaciones y se han orientado a generar y recibir valor de este (Payne & Frow, 2005).

De esta forma, en las empresas de diferentes rubros, las áreas de CRM se han convertido en unidades estratégicas relevantes y que mejoran el desempeño de las compañías desde la perspectiva de los clientes (Varajão & Cruz-Cunha, 2016). Según Payne y Frow (2005), el CRM es un enfoque que busca mejorar las relaciones con los clientes y para ello se puede valer de diferentes estrategias, herramientas analíticas y de la información que las empresas capturan.

En particular, en el sector *retail* las estrategias que más impacto generan dentro de las empresas son las de adquisición, retención y fidelización de clientes. En este trabajo de investigación se desarrollará el estado del arte de dichas estrategias con especial profundidad en las dos últimas.

Farquhar (2003) refiere que si una empresa busca generar relaciones de valor a largo plazo con sus clientes es imprescindible que primero los retenga. En este sentido, Ramageri y Desai (2013) sostienen que el DM puede ser utilizado como una herramienta que permita desarrollar la retención de clientes. De acuerdo con estos autores, las empresas cuentan con información del comportamiento del cliente a partir de la cual pueden determinar las necesidades que ellos buscan satisfacer a través de la compra de los productos que ofrece la compañía. A partir de este análisis es posible predecir lo que cada cliente buscará para poder ofrecer productos que retengan a clientes antiguos y atraigan a los nuevos.

Ahora bien, DM es una herramienta aplicable en las estrategias de adquisición y retención de CRM para el sector *retail*; no obstante, ML y *Marketing Analytics* también pueden ser funcionales. Así, Torrejón (2018) investigó sobre el diseño de estrategias de retención para el servicio posventa de una empresa automotriz mediante el uso de *Marketing Analytics* y ML. Él comenzó su investigación con la elaboración de un diagnóstico de la empresa en donde, a través de un mapeo de procesos del servicio, e identificó siete fallas en el servicio posventa. Torrejón, priorizó dichas fallas mediante un proceso de

análisis jerárquico y determinó las tres fallas prioritarias entre las que destaca la ausencia de diferenciación de clientes que permita generar distintas acciones para cada segmento.

Para solucionar dicha problemática, el autor propuso identificar las variables que explicaban la satisfacción del cliente mediante una regresión logística que es un algoritmo de ML (Géron, 2017). Los resultados obtenidos permitieron entender las determinantes de la satisfacción, las cuales fueron realizar un mantenimiento conforme a lo acordado y un buen lavado del auto. Al implementar mejoras en estas actividades dentro del servicio se estimó que la satisfacción aumentó en 7.5%.

Otra técnica utilizada para gestionar clientes y retenerlos en la empresa es el *clustering*. Zahrotun (2017) realizó la implementación de esta técnica de DM dentro de una tienda *e-commerce* con el objetivo de mejorar las relaciones con los clientes en el largo plazo. El autor aplicó la metodología KDD (por sus siglas en inglés *Knowledge Discovery in Database*), para entender la data con la que contaba, así como los métodos de *Fuzzy C-Means* y *Fuzzy FRM* [*Frequency, Recency and Monetary value*]. El primer método permite ejecutar el *clustering per se*; mientras que el segundo define las dimensiones que caracterizarán a los *clusters*, las cuales son frecuencia, *recency* (tiempo que ha transcurrido desde la última acción del cliente), y valor monetario.

Al implementar los métodos, Zahrotun (2017) obtuvo tres escenarios (de uno, dos y tres *clusters* respectivamente). Dichos resultados fueron sometidos a la prueba de pureza que calificó al escenario de tres *clusters* como el que poseía el mejor valor de pureza y, por lo tanto, fue el seleccionado. De esta manera, Zahrotun caracterizó a cada *cluster* a partir de las dimensiones señaladas previamente y cada segmento puede ser gestionado de forma distinta.

Mientras que la retención de un cliente se puede resumir en los actos de recompras seguidas; la lealtad de este hacia una compañía es más compleja de definir. Por un lado, algunos autores realizaron investigaciones orientadas a inferir una escala según la cual se puede medir el nivel de fidelización de un cliente; por otro lado, otros investigadores estudiaron la fidelidad de un cliente y su relación con otras variables inherentes al servicio.

Sobre el primer punto, Bobâlcă, Gătej y Cioanu (2012) realizaron un estudio cuyo estado del arte evidenciaba que, hasta antes de 1970, se solía definir la lealtad únicamente a través del acto de recompra. Según Kuusik (2007), esta tendencia continuó así hasta que en 1969 se definió la lealtad a partir de la actitud y el comportamiento y este fue el modelo de medición de lealtad más aceptado por la comunidad de marketing. Sin embargo, Oliver (1997), Olsen (2002); y Harris y Goode (2004) estudiaron y desarrollaron un modelo que explica que la lealtad se puede medir en cuatro factores diferentes: cognitivo (evaluación de los atributos de la compañía), afectivo (grado de aprecio a la

empresa), conativo (tendencia conductual a la recompra) y el actuar (hablar bien de la empresa y recomendarla).

Bajo el concepto de una lealtad dada por cuatro factores, Bobâlcă et al. (2012) propusieron la hipótesis de que la lealtad es un proceso compuesto de dichos factores. Para contrastar esta hipótesis, realizaron una investigación con 676 estudiantes de Alexandru Ioan Cuza University of Iasi mediante un análisis cuantitativo (encuesta) y cualitativo (entrevista). Los resultados obtenidos mostraron evidencia estadística de que la dimensión afectiva, la conativa y el accionar sí forman parte de la lealtad; no obstante, el componente cognitivo no pasó la prueba de fiabilidad, ya que obtuvo un alfa de Cronbach menor a 0.7.

Sobre el segundo punto, Ibrahim et al. (2020) investigaron el efecto de aplicar un sistema de calidad en el canal *e-commerce* en la lealtad de los clientes mediante el uso de CRM. Los autores comenzaron su estudio mencionando que en una coyuntura donde la tecnología avanza muy rápido, la competencia se agudiza, por ello es necesario que las empresas generen ventajas competitivas. Una forma de alcanzar dicho objetivo es enfocarse en la gestión de los clientes; en particular, en la satisfacción y lealtad de los clientes a través del CRM.

Los autores realizaron una encuesta a estudiantes de la Universidad de Sriwijaya para obtener la información que requerían. Dicho cuestionario buscó capturar la percepción de los encuestados sobre el servicio de calidad (variable independiente) y su fidelidad (variable dependiente) a los servicios de compras *online*. Con la data capturada de 50 estudiantes, utilizaron una regresión lineal y los resultados obtenidos fueron que la calidad del servicio influye positivamente en la fidelidad, de modo que un aumento de 1% en dicho parámetro incrementará en 0.204 el valor de la fidelidad.

Asimismo, dado que obtuvieron un p-value de 0.003, afirmaron que sí existe una significancia estadística, por lo tanto, hay una influencia significativa del servicio de calidad en la fidelidad de los encuestados. Finalmente, el coeficiente de determinación (R^2) que obtuvieron fue de 0.172. Esto significa que el 17.2% de la variable fidelidad es explicada por la variable calidad del servicio y el 82.8% restante es influenciado por otras variables no examinadas.

El estudio de Bowen y Chen (2001) complementa lo expuesto previamente. Los autores trabajaron en una investigación para determinar la relación entre la fidelidad del cliente y la satisfacción que este obtiene por el servicio o producto recibido. En el estado del arte, Bowen y Chen resaltan los beneficios que traen los clientes fidelizados a una empresa. Por ejemplo, el aumento de ingresos apalancado por la reducción en gastos de marketing (Reichheld & Sasser, 1990), la promoción que los clientes leales hacen al negocio (Raman, 1999) y la variedad de productos comprados por estos clientes dentro de los hoteles (Bowen y Shoemaker, 1998). En este marco, el enfoque del estudio se orienta a hallar y

desarrollar métodos que ayuden a los hoteles a incrementar la lealtad de sus clientes; tomando como base el supuesto de que la lealtad se mide mediante la acción de recompra en la empresa y las opiniones positivas que el cliente tenga sobre esta. Bowen y Chen utilizaron un análisis cualitativo (entrevista) y una encuesta que tuvo un total de 546 respuestas. Cabe mencionar que la escala usada en la encuesta fue de 1: muy insatisfecho hasta 7: muy satisfecho.

Los resultados que obtuvieron apuntan a que sí existe una relación entre la satisfacción y la fidelidad del cliente de tipo no lineal. Esto se sustenta en que cuando la satisfacción alcanza el nivel de 5, el porcentaje de clientes que indicaron que volvería fue de 10.4%; cuando la satisfacción fue de 6, dicho porcentaje subió hasta 24.6%; y, por último, cuando la satisfacción alcanzó al máximo de 7, el indicador de lealtad incrementó hasta llegar al 65%.

Finalmente, otros autores estudiaron las técnicas que pueden explicar la fidelización de un cliente a partir de información que se obtiene por el comportamiento del mismo cliente. Estas investigaciones utilizan la información transaccional de los clientes que suele ser profusa en los giros retail, banca, seguros y *e-commerce* en general. Seyedhosseini, Maleki y Gholamian (2010) señalan que el análisis *FRM* puede definir las dimensiones que explique la lealtad de un cliente. Dicho análisis se complementa con el *CLV (Customer Lifetime Value)* el cual cuantifica la fidelización del cliente a partir de la frecuencia, *recency* y valor monetario de este (Liu & Shih, 2005). Por otro lado, Bhagawan, Bishnu y Dhan (2018), aplicaron un *clustering* jerárquica para segmentar a los clientes de acuerdo con tres dimensiones: soporte, confianza y elevación. Los autores definieron dichas variables a partir de la información transaccional de un supermercado y las utilizan para explicar la fidelidad de los clientes. Como resultado de la investigación, los autores obtuvieron cinco *clusters* y definieron estrategias comerciales para cada segmento con el objetivo de mejorar la gestión de su cartera de clientes.

Capítulo 3. Conclusiones

A partir del estudio del marco teórico y del estado del arte se concluye lo siguiente.

- En el mundo, el sector servicios aporta un 61% al PBI mundial: mientras que en Perú este aporte asciende a un 59%. Por otro lado, este sector genera empleo para un 40% de la PEA. Por lo tanto, es relevante tanto en la economía mundial como en la peruana.
- Los servicios poseen una naturaleza de procesos, ya que cuentan con una gestión de recursos de diversas índoles, un inicio, un fin y un objetivo que es satisfacer al cliente. Estos se pueden estudiar mediante herramientas que se usan también para procesos de manufactura como la diagramación del flujo, matriz QFD, entre otros.
- En la actualidad, los clientes del sector *retail* están expuestos a abundantes y variadas ofertas, por lo tanto, para una empresa es más complicado retener y fidelizar a su clientela; así como adquirir nuevos clientes. En este marco, las compañías están posicionando al cliente en el centro de sus operaciones, y el área de CRM es muy relevante porque su objetivo es generar relaciones de valor con los clientes.
- El uso de data, a través del CRM, ha generado beneficios económicos a las empresas, ya que genera valor tanto para el cliente como para la organización. Las empresas adquieren valor al entender el comportamiento de sus clientes mediante el estudio de sus compras. Los clientes, por su parte, obtienen más valor cuando las empresas logran ofrecerle productos o servicios más relacionados con sus necesidades.
- El enfoque CRM es compatible con las herramientas de DM y ML. En ese sentido, las estrategias principales de CRM: adquisición, retención, y fidelización, pueden ser reforzadas positivamente con el uso de estas herramientas.
- En el contexto peruano, hay precedentes de aplicaciones de las herramientas analíticas descritas en el sector *retail*, como el caso de la tesis de posgrado de Torrejón que aplicó *clustering* (herramienta de ML) en una empresa *retail* de venta de autos. Dicho estudio propuso una mejora en la gestión de retención de clientes, que es una estrategia de CRM, y obtuvo un incremento en la tasa de retención de 7.5%.

Referencias

- Alukal, G. (2007). Lean *Kaizen* in the 21st Centruy. *Quality Progress*, 40(8), 69-70.
- Bagozzi, R.P. (1974). Marketing as an organized behavioral system of exchange. *Journal of Marketing*, 38(4), 77-81. doi: 10.1177/002224297403800414
- Bai, F. & Qin, Y. (2016). The implementation of relationship marketing and CRM: how to become a customer-focused organization. *Journal of Business & Economic Policy*, 3(2), 112-124.
- Banco Mundial (2019). Services, value added (% of GCP). Recuperado de: <https://datos.bancomundial.org/indicador/NV.SRV.TOTL.ZS>
- BBVA API_Market, 2016. Pros and cons of Python and R for data science. Recuperado de: <https://bbvaopen4u.com/en/actualidad/pros-and-cons-python-and-r-data-science>
- BBVA Research. (2018). *Perú: situación retail moderno 2018*. Lima, Perú. Recuperado del sitio Internet de BBVA Research: Peru | Retail Outlook 2018 <https://www.bbvaresearch.com/en/publicaciones/peru-outlook-for-modern-retailing-2018/>
- Berman, B., Evans, J.R., & Chatterjee, P. (2018). *Retail management: a strategic approach* (13a. ed.). Londres, Inglaterra: Pearson Education.
- Berry, L.L. (1983). Relationship marketing. *Emerging perspectives on services marketing*, 66(3), 33-47.
- Bhagawan, R., Bishnu, P.G., & Dhan, P.G. (2018). Enhancement of Supermarket Business and Market Plan by Using Hierarchical Clustering and Association Mining Technique. *2018 International Conference on Networking and Network Applications (NaNA), 2018*, 384-389. doi: 10.1109/NANA.2018.8648716
- Bibault, J.E., Giraud, P., & Burgun, A. (2016). Big data and machine learning in radiation oncology: State of the art and future prospects. *Cancer Letters*, 382(1), 110-117. doi: 10.1016/j.canlet.2016.05.033
- Bitner, M.J., Ostrom, A.L., & Morgan, F.N. (2008). Service blueprinting: a practical technique for service innovation. *California Review Management*, 50(3), 66-94. doi: 10.2307/41166446
- BlackSip. (2019). *Reporte de industria: El e-commerce en Perú*. Lima, Perú. Recuperado del sitio Internet de Asociación de Emprendedores del Perú (ASEP): <https://asep.pe/wp-content/uploads/2019/08/Reporte-de-industria-del-eCommerce-Peru-2019-eBook.pdf>
- Bobâlcă, C., Gătej, C., & Cioanu, O. (2012). Developing a scale to measure customer loyalty. *Procedia Economics and Finance*, 3(2012), 623-638. doi: 10.1016/S2212-5671(12)00205-5
- Boulding, W., Staelin, R., Ehret, M., & Johnston, W.J. (2005). A CRM roadmap: What we know, potential pitfalls, and where to go. *Journal of Marketing*, 69, 155-166. doi: 10.1509/jmkg.2005.69.4.155
- Bowen, J.T., & Chen, S. (2001). The relationship between customer loyalty and customer satisfaction. *International Journal of Contemporary Hospitality Management*, 13(5), 213-217. doi: 10.1108/09596110110395893

- Bowen, J.T., & Shoemaker, S. (1998). Loyalty: a strategic commitment?. *The Cornell Hotel and Restaurant Administration Quarterly*, 39(1), 12-25. doi: 10.1016/S0010-8804(97)83878-5
- Brownlee, J. (2019). *What is R*. Recuperado de: <https://machinelearningmastery.com/what-is-r/>
- Brunelli, M. (2015). *Introduction to the analytic hierarchy process*. Aalto, Finlandia: SpringerBriefs in Operations Research. doi: 10.1007/978-3-319-12502-2.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0. Step-by-step data mining guide. USA: SPSS, Inc.
- Charu, A., & Chandan, R. (2014). *Data clustering. Algorithms and applications*. Florida, USA: CRC Press.
- Columbus, L. (2017). *53% Of companies are adopting big data analytics*. [noticia online]. Forbes. Recuperado de: <https://www.forbes.com/sites/louiscolombus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/#469a096739a1>
- Coughlan, A.T., Anderson, E., Stern, L.W., & El-Ansari A.I (2006). *Marketing channels (7a. ed.)*. Londres, Inglaterra: Prentice Hall.
- Cristiano J.J., Liker, J.K., & White III, C.C. (2000). Customer-driven product development through quality function deployment in the US and Japan. *Journal of Product Innovation Management: an international publication of the product development & management association*, 17(4), 286-308. doi: 10.1111/1540-5885.1740286
- Cruz, J.A., & Wishart, D.S. (2006). Applications of machine learning in cancer predictions and prognosis. *Cancer informatics*, 2(2006), 59-77. doi: 10.1177/117693510600200030
- Cui, G., Wong, M.L., & Lui, H.K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, 52(4), 597-612. doi: 10.1287/mnsc.1060.0514
- De Leeuw, J. (2011). Statistical Software: An Overview. En: *International Encyclopedia of Statistical Science* (ed. 2011), 1470-1473.
- Fang, S.H., Fei, Y.X., Xu, Z., & Tsao, Y. (2017). Learning transportation modes from smartphone sensors based on deep neural network. *IEEE Sensors Journal*, 17(18), 6111-6118. doi: 10.1109/JSEN.2017.2737825
- Farquhar, J.D. (2003). Retaining customers in traditional retail financial service: interviewing 'les responsables'. *International Review of Retail, Distribution & Customer Research*, 13(4), 393-405. doi: 10.1080/0959396032000129499
- Fitzsimmons, J.A., & Fitzsimmons, M.J. (2011). *Service management. Operations, strategy, information technology (7a. ed.)*. New York, USA: McGraw-Hill.
- Gallino, S., & Moreno, A. (2014). Integration of online and offline channels in retail: the impact of sharing reliable inventory availability information. *Management Science*, 60(6), 1434-1451. doi: 10.1287/mnsc.2014.1951
- Géron, A. (2017). *Hands-On machine learning with Scikit-Learn & TensorFlow*. California, USA: O'Reilly Media, Inc.

- Grönroos, C. (1990). *Service management and marketing: managing the moments of truth in service competition*. Maryland, USA: Lexington Books.
- Grönroos, C. (1994) Qua Vadis, marketing? Toward a relationship marketing paradigm. *Journal of Marketing Management*, 10(5), 347-360. doi: 10.1080/0267257X.1994.9964283
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining. Concepts and techniques*. (3a. ed.). Massachusetts, USA: Elsevier, Inc.
- Hand, D., Mannila, H., & Padharic, S. (2001). *Principles of Data Mining*. Massachusetts, USA: The MIT Press.
- Harris, L., & Goode, M. (2004). The Four Stages of Loyalty and the Pivotal Role of Trust: A Study of Online Service Dynamics. *Journal of Retailing*, 80(2), 139-158. doi: 10.1016/j.jretai.2004.04.002
- Hurwitz, J., & Kirsch, D. (2018). *Machine learning for dummies*. USA: John Wiley & Sons, Inc. [recurso electrónico]. Recuperado de: <https://www.ibm.com/downloads/cas/GB8ZMQZ3>
- IBM. (s.f.). *What is Machine Learning?*. Recuperado de: <https://www.ibm.com/topics/machine-learning>
- Ibrahim, A., Napian, S.D.R., Firanisa, A., Sartika, Sari, R.D., & Audya, M. (2020). The effect of e-commerce application service quality of customer loyalty using customer relationship management. *Sriwijaya International Conference of Information Technology and Its Applications (SICONIAN 2019)* 172, 680-687. doi: 10.2991/aisr.k.200424.103
- Instituto Nacional de Estadística e Informática. (2019). Perú: Informe Económico Trimestral. IV trimestre 2018. Recuperado de: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1649/libro.pdf
- International Organization for Standardization. (2015). *Quality Management Systems – Requirements* (ISO standard no. 9001:2015). Recuperado de: <https://www.iso.org/obp/ui/#iso:std:iso:9001:ed-5:v1:en>
- ISOTools Excellence (2016). *¿Cómo es un mapa de procesos basado en la norma ISO 9001 2015?*. Recuperado de: <https://www.nueva-iso-9001-2015.com/2016/05/como-es-un-mapa-procesos-basado-norma-iso-9001-2015/>
- Jahangiri, A., & Rakha, H.A. (2015). Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. *IEEE transactions on intelligent transportation systems*, 16(5), 2406-2417. doi: 10.1109/TITS.2015.2405759
- Jacka, J.M., & Keller, P.J. (2009). *Business process mapping: improving customer satisfaction*. (2a. ed.). New Jersey, USA: John Wiley & Sons, Inc.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. New York, USA: Springer. doi: 10.1007/978-1-4614-7138-7
- Jordan, M.I., & Mitchell, T.M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. doi: 10.1126/science.aaa8415

- Khan, A., Ehsan, N., Mirza, E., & Sheikh, Z. (2012). Integration between Customer Relationship Management (CRM) and Data Warehousing. *Procedia Technology* 1(2012), 239-249. doi: 10.1016/j.protcy.2012.02.050
- Kuusik, A. (2007). *Affecting customer loyalty: do different factors have various influences in different loyalty levels?* The University of Tartu Faculty of Economics and Business Administration Working Paper No. 58-2007. doi: 10.2139/ssrn.1025989
- Lamoreaux, N.R., Raff, D.M.G., & Temin, P. (2002). Beyond markets and hierarchies: toward a new synthesis of American business history. *American Historical Review*, 108, 404-469. doi: 10.3386/w9029
- Levitt, T. (1969). *The marketing mode: pathways to corporate growth*. New York, USA: McGraw-Hill.
- Liu, D.-R., & Shih, Y.-Y. (2005). Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management*, 42(3), 387-400. doi: 10.1016/j.im.2004.01.008
- Lovelock, C., Reynoso, J., D'Andrea, G., Huete, L.M., & Wirtz J. (2011). *Administración de servicios. Estrategias para la creación de valor en el nuevo paradigma de los negocios* (2a. ed.). Ciudad de México, México: Pearson Educación.
- Magar, V.M., & Shinde, V.B. (2014). Application of 7 quality control (7 QC) tools for continuous improvement of manufacturing processes. *International Journal of Engineering Research and General Science*, 2(4), 364-371. Recuperado de: <http://www.ijergs.org/files/documents/APPLICATION-45.pdf>
- Manene, L. (2011). Los diagramas de flujo: su definición, objetivo, ventajas, elaboración, fases, reglas y ejemplos de aplicaciones. Recuperado de: https://d1wqtxts1xzle7.cloudfront.net/60656037/Los_diagramas20190920-8696-u4r0qz.pdf?1568997372=&response-content-disposition=inline%3B+filename%3DEl_28_julio_2011_en_Estructura_Organizat.pdf&Expires=1595042698&Signature=OvrPySAD6RVAikYg1pEdUkDcGQr5rGlZE0pHksG2uyjtJ0-cZvHYbNMnw4zEuz-nvLpytY5LwSNSN5rWccUgqbiyyfK-jIVrgQuoqcikeSMgkm~BVU3ZuPsvjbQWUHpbFsf4v7IveWLVAg4zFJJk4Wj9hG5NdrcR7TD9sgRfX8wjzVjQMGeixJxExqf4DQTbwQhGPh9nOf8fb2GP4TrJZvpYjDsJK00c4flz28P-ulM7bAmdTpNn0vjpcFBL4CUJ-EZXrOJuQOxE17kjN9rhbGkArW~eevqkaq0D1mqQv9osXYhoy8oEiCroKJ60WxFdCSS9kWZpSux8FVptok6A__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- Ministerio de Planificación Nacional y Política Económica. (2009). Guía para la elaboración de diagramas de flujo. Recuperado de: <http://evalperu.org/sites/default/files/resources/file/3.%20MPNGE%20guia%20diagramas-flujo-2009.pdf>
- Ministerio del Trabajo y Promoción del Empleo. (2018). Informe anual del empleo en el Perú. Recuperado de: https://cdn.www.gob.pe/uploads/document/file/285846/IAE_2017__14-12-2018_.pdf
- Morgan, R.M., & Hunt, S.D. (1994). The commitment-trust theory of relationship marketing. *Journal of marketing*, 58(3), 20-38. doi: 10.2307/1252308
- Muenchen, R. A. (Febrero 2019). *The popularity of Data Science software*. [artículo online]. R4stats.com. Recuperado de: <http://r4stats.com/articles/popularity/>

- Murugaiah, U., Benjamin, S.J., Marathamuthu, M.S., & Muthaiyah, S. (2009). Scrap loss reduction using the 5-whys analysis. *International Journal of Quality & Reliability Management*, 27(5), 527-540. doi: 10.1108/02656711011043517
- Oliver, R. (1997). *Satisfaction: A behavioral perspective on the consumer* (1a. ed.). New York, USA: McGraw-Hill.
- Olsen, S. (2020). Comparative evaluation and the relationship between quality, satisfaction, and repurchase loyalty. *Journal of the Academy of Marketing Science* 30(3), 240-249. doi: 10.1177/0092070302303005
- Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing*, 69(4), 167-176. doi: 10.1509/jmkg.2005.69.4.155
- Payton, F.C., & Zahay, D. (2003). Understanding why marketing does not use the corporate data warehouse for CRM applications. *Journal of Database Marketing & Customer Strategy Management*, 10(4), 315-326. doi: 10.1057/palgrave.jdm.3240121
- Plotnikova, V., Dumas, M., & Milani, F.P. (2020). Systematic literature review: adaptations of data mining methodologies. *PeerJ Computer Science*, 6(2), 267-272. doi: 10.7717/peerj-cs.267
- Porter, M.E. (1985). *Ventaja competitiva: creación y sostenimiento de un desempeño superior*. México D.F., México: Grupo Editorial Patria.
- Project Management Institute. (2004). *A guide to the project management body of knowledge (PMBOK guide)*. (6a. ed.). Pennsylvania, USA: Project Management Institute.
- Python™ (s.f.). *Python is powerful... and fast; plays well with others; runs everywhere; is friendly & easy to learn; is open*. Recuperado de: <https://www.python.org/about/>
- Ramageri, B.M., & Desai, B.L. (2013). Role of data mining in retail sector. *International Journal on Computer Science and Engineering*, 5(1), 47-50. doi: 10.1.1.302.943
- Reichheld, F.F., & Schefter, P. (2000). E-loyalty: your secret weapon on the web. *Harvard Business Review*, 78, 105-113.
- Raman, P. (Agosto 1999). Way to create loyalty. *New Straits Times*.
- Ruan, X., Yin, Z., & Frangopol, D. M. (2015). Risk matrix integrating risk attitudes based on utility theory. *Risk Analysis*, 35(8), 1437-1447. doi: 10.1111/risa.12400
- Samaniego, J.F. (2019). *Andre Fernández, CEO de Hockerty: "La optimización continua nos ha llevado a donde estamos hoy"* [noticia online]. Hablemos de empresas. Recuperado de: <https://hablemosdeempresas.com/empresa/hockerty-sastreria-online/>
- Syedhosseini, S.M., Maleki, A., & Gholamian, M. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. *Expert Systems with Applications*, 37(2010), 5259-5264. doi: 10.1016/j.eswa.2009.12.070
- Schmenner, R.W. (2008). Manufacturing, service, and their integration: some history and theory. *International Journal of Operations & Production Management*, 29(5), 431-443. doi: 10.1108/01443570910953577
- The R Foundation. (s.f.). *What is R?* [página web]. Recuperado de: <https://www.r-project.org/about.html>

- Torrejón, D.S.R. (2018). *Diseño de Estrategias de Retención de Posventa en una Empresa Automotriz de Vehículos Livianos Aplicando Marketing Analytics* (Tesis de Maestría). Recuperado de: <http://repositorio.pucp.edu.pe/index/handle/123456789/145951>
- Valletta, J.J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124, 203-220. doi: 10.1016/j.anbehav.2016.12.005
- Varajão, J., & Cruz-Cunha, M.M. (2016). Main motivations for CRM adoption by large Portuguese companies - A principal component analysis. *Procedia Computer Science*, 100(2016), 1269-1279. doi: 10.1016/j.procs.2016.09.165
- Wang, G. (2019). Machine learning for inferring animal behavior from location and movement data. *Ecological informatics*, 49, 69-76. doi: 10.1016/j.ecoinf.2018.12.002
- Yacuzzi, E., & Martín, F. (2003). *QFD: conceptos, aplicaciones y nuevos desarrollos* (Serie Documentos de Trabajo No. 234). Recuperado de: <http://hdl.handle.net/10419/84469>
- Zahrotun, L. (2017). Implementation of data mining technique for customer relationship management (CRM) on online shop tokodiapers.com with fuzzy c-means clustering. *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 299-303. doi: 10.1109/ICITISEE.2017.8285515
- Zentes, J., Morschett, D., & Schramm-Klein, H. (2011). *Strategic retail management* (2a. ed.). Wiesbaden, Alemania: Gabler Verlag.
- Zeithaml, V. A., & Bitner, M. J. (2000). *Service marketing* (2a. ed.). New York, USA: McGraw-Hill.
- Zineldin, M. (2006). The royalty of loyalty: CRM, quality and retention. *Journal of Consumer Marketing*, 23(7), 430-437. doi: 10.1108/07363760610712975