

Un modelo predictivo es un conjunto de procesos ejercidos a través de técnicas computacionales de análisis de datos que ayudan a inferir la probabilidad de que ocurran determinadas situaciones previas a su consecución y, a su vez, detectar oportunidades. El software libre se ha convertido en uno de los movimientos tecnológicos de mayor auge en el siglo XXI. Para su desarrollo, ha sido necesario contar con un grupo de herramientas que hagan óptima su utilización y sean fáciles de aprender. Python es un lenguaje de programación que cumple con lo planteado y se perfila como una opción recomendada para el desarrollo de software libre. Por ello, este libro tiene como propósito presentar un modelo predictivo construido mediante Python, con el fin de identificar qué clientes recomendarían los productos y servicios que ofrece la empresa Distecom. En el estudio se empleó un enfoque mixto y se utilizó la encuesta a los colaboradores y clientes mediante un cuestionario estructurado. Se encuestó a seis empleados y una muestra de 122 clientes activos de una población total de 1006 clientes. También se ejecutó la debida tabulación y presentación mediante gráficos con los respectivos datos cuantitativos. La herramienta de entorno de desarrollo utilizada fue Python, con la cual se configuró las funcionalidades del proyecto. Asimismo, se destacan los beneficios de la propuesta como herramienta de gestión para la empresa Distecom.



William Rafael Raymondi Lomas. Magíster en Sistemas de Información, mención en Inteligencia de Negocios. Ingeniero en Networking y Telecomunicaciones. Doctorando en Ciencias de la Educación, Universidad Santander, México. Docente de la Facultad de Ciencias Matemáticas y Físicas, de la Universidad de Guayaquil, Ecuador. Autor de varios artículos científicos e investigaciones en el área de las telecomunicaciones.
william.raymondll@ug.edu.ec <http://orcid.org/0000-0003-1641-6193>



Sergio Israel Peña Guano. Magíster en Sistemas de Información, mención en Inteligencia de Negocios. Ingeniero en Networking y Telecomunicaciones. Doctorando en Ciencias de la Educación, Universidad Santander, México. Docente de la Facultad de Ciencias Matemáticas y Físicas, de la Universidad de Guayaquil, Ecuador. Autor de varios artículos científicos e investigaciones en el área de las telecomunicaciones.
sergio.penagua@ug.edu.ec <http://orcid.org/0000-0003-4021-1892>



9 780311 000470



Modelos predictivos para identificar probabilidades de recomendación

Modelos predictivos para identificar probabilidades de recomendación

William Rafael Raymondi Lomas
Sergio Israel Peña Guano



Modelos predictivos para identificar probabilidades de recomendación

Diseño: Ing. Erik Marino Santos Pérez.

Traducción: Prof. Dr. C. Ernan Santiesteban Naranjo.

Corrección de estilo: Prof. Dra. C. Leydis Iglesias Triana.

Diagramación: Prof. Dr. C. Ernan Santiesteban Naranjo.

Director de Colección Tecnología: Prof. Dr. C. Wilber Ortiz Aguilar.

Jefe de edición: Prof. Dra. C. Kenia María Velázquez Avila.

Dirección general: Prof. Dr. C. Ernan Santiesteban Naranjo.

© William Rafael Raymondi Lomas

Sergio Israel Peña Guano

Sobre la presente edición:

Primera edición

Esta obra ha sido evaluada por pares académicos a doble ciegos

Lectores/Pares académicos/Revisores: 0012 & 0070

Editorial Tecnocientífica Americana

Domicilio legal: calle 613sw 15th, en Amarillo, Texas. **ZIP:** 79104, EEUU

Teléfono: 7867769991

Fecha de publicación: 27 junio de 2023

Código BIC: KNTX

Código EAN: 9780311000470

Código UPC: 978031100047

ISBN: 978-0-3110-0047-0

La Editorial Tecnocientífica Americana se encuentra indizada en, referenciada en o tiene convenios con, entre otras, las siguientes bases de datos:





Contenido

Capítulo 1. Un acercamiento a la empresa Distecom	3
1.1. Pregunta problema.....	5
1.2. Idea a defender	6
1.3. Objetivos de la investigación	6
1.4. Justificación.....	6
Capítulo 2. Marco teórico	9
2.1. Conceptos tecnológicos.....	9
2.1.1. Modelo predictivo	9
2.1.2. Importancia del modelo predictivo	12
2.1.3. Tipos de modelos predictivos.....	14
2.1.4. Big Data.....	16
2.1.5. Beneficios.....	17
2.1.6. Inteligencia de negocios	18
2.1.7. Principales características de la inteligencia de negocio	19
2.1.8. Phyton	20
2.1.9. Flujos de trabajo naranja	21
2.1.10. Ventas por medios de referidos	31
2.1.11. Maneras de lograr clientes referidos.....	31
2.1.12. Tipos de referidos.....	32
2.1.13. Pasos para obtener referidos.....	33
2.2. Metodología	35
2.2.1. CRISP-DM	35
2.2.2. Kimball.....	38
2.2.3. Hefesto	41
2.2.4. Cuestionario	44
2.2.5. Encuesta	46
Capítulo 3. Marco metodológico	48
3.1. Criterios de éxito del negocio.....	48
3.2. Definición de requerimientos del proyecto	49



3.2.1. Inventario de recursos	49
3.2.2. Requerimientos funcionales y no funcionales.....	49
3.2.3. Roles y responsabilidades	51
3.2.4. Riesgos y contingencias	52
3.2.5. Plan del proyecto.....	53
Capítulo 4. Análisis de los resultados	55
4.1. Esquema de validación.....	56
4.2. Descripción de resultados.....	58
Capítulo 5. Presentación de soluciones tecnológicas, aplicaciones técnicas	73
5.1. Análisis de la situación actual	73
5.1.1. Organigrama de la institución	74
5.1.2. Objetivos de la institución.....	75
5.2. Descripción de las fuentes de datos.....	76
5.2.1. Exploración de los datos	76
5.3. Desarrollo del prototipo del Cuadro de Mando en Power BI.....	90
Conclusiones	98
Recomendaciones	101
Bibliografía	103



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación
William Rafael Raymondi Lomas
Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023





Capítulo 1. Un acercamiento a la empresa Distecom

Distecom es una empresa fundada el 14 de marzo del 2001. Está ubicada en las calles José Mascote 1441 y C. Ballén, de la ciudad de Guayaquil, provincia del Guayas, parroquia 9 de octubre. Esta empresa tiene como actividad principal la venta, mantenimiento y reparación de equipos informáticos; también presta servicios adicionales como instalación de cableado estructurado y eléctrico, instalación de sistemas de seguridad, Data Center, entre otros servicios. La empresa tiene una cartera de clientes conformada por empresas públicas y privadas.

La visión de Distecom es ser reconocida a nivel nacional e internacional como una proveedora de productos y servicios integrales y de alta calidad. También, pretende posicionarse como generadora de valor y alto nivel de satisfacción para sus clientes, empleados y socios. En ese mismo contexto, la empresa tiene como misión incrementar la competitividad y productividad; para ello, proporciona e implementa tecnología innovadora a nivel empresarial como solución práctica adaptada para cubrir las necesidades existentes.

Es difícil predecir el futuro, especialmente, en un contexto de cambios e incertidumbre pospandémica, como ocurrió a finales del año 2021. Ante esto es importante realizar análisis prospectivo sobre el comportamiento de los clientes. Investigaciones recientes demuestran que más del 70% de empresas consideran importante contar con un modelo predictivo para identificar las diferentes causas y efectos que limitan la continuidad y posicionamiento de los negocios. Por ello, se busca disponer de un modelo actualizado en inteligencia de negocios.

Indudablemente, el modelado predictivo es un sistema que emplea datos y estadísticas para predecir resultados a partir de una base de datos. Esta puede utilizarse para predicciones de todo tipo como, por ejemplo, el identificar la probabilidad de recomendación de los clientes de una empresa. Por tanto, el análisis predictivo utiliza técnicas como el aprendizaje automático y la minería de datos, en suma, promete grandes beneficios, como conjuntos de datos suficientemente grandes y manejo de información compleja. Los modelos predictivos aportan a la solución de nuevos problemas, un adecuado manejo y organización de los datos, ofrece privacidad y seguridad de la información.

En la denominada era del cliente, las empresas invierten fuertemente en la mejora de la experiencia del cliente debido al incremento competitivo que existe en los actuales momentos; además de la necesidad de diferenciarse, con el objetivo de cambiar la asistencia de un buen servicio por la entrega de una experiencia perfecta para el cliente. De tal manera, las empresas reúnen esfuerzos por invertir en tecnologías que ofrezcan una comprensión integral conforme las expectativas de los clientes.

Por lo antes dicho, se propone generar una aplicación de modelado predictivo con un software libre, herramienta capaz de analizar el comportamiento de los usuarios para identificar la probabilidad de recomendación de los clientes de la empresa Distecom. Es un software libre de aprendizaje automático para Data Mining y análisis predictivo, una herramienta adecuada que permite visualizar patrones mediante gráficos y tablas. Esto es un componente de minería de datos que permite una programación visual, rápida y versátil para un análisis exploratorio de la información. De ahí que, la aplicación es amigable e

intuitiva; permite procesar, filtrar información, modelar datos, evaluar modelos y técnicas de exploración (Haro, 2017).

Otro aspecto relevante es que los modelos actuales cubren otros elementos no focalizados en el cliente, tales como el cumplimiento del presupuesto anual de ventas. Por este motivo, se requiere desarrollar una alternativa enfocada a adquirir nuevos referidos, recuperar a los clientes potenciales que tienen tendencia a desertar, y todo esto mediante el análisis de grandes cantidades de datos.

La empresa Distecom requiere tener la capacidad de analizar la satisfacción de los clientes; ya que son ellos los sujetos de estudio. La prueba de aplicación del modelo predictivo permitirá conocer un punto central único para todos los tipos de comentarios recibidos, incluidos otros canales como son los medios de comunicación sociales. Con estos datos, la empresa podrá identificar con mayor rapidez y eficacia los vacíos existentes entre los hallazgos obtenidos y las acciones que necesitan ser implementadas para acercarse al usuario. De este modo, se dispondrá de elementos para la construcción de la lealtad del cliente y, más que eso, transformar ese cliente en un gran embajador de la marca.

1.1. Pregunta problema

¿Cómo contribuirá la aplicación de modelos predictivos a identificar la probabilidad de recomendación de los clientes de Distecom?



1.2. Idea a defender

La aplicación de un modelo predictivo permitirá identificar la probabilidad de recomendación de clientes en la empresa Distecom, mediante entrevistas, formularios de Google y censo, datos que serán utilizados en el aplicativo Phyton.

1.3. Objetivos de la investigación

Objetivo general

Diseñar un modelo de análisis predictivo que permita identificar la probabilidad de recomendación de los clientes Distecom para contribuir a una toma de decisiones gerencial basada en información actualizada.

Objetivos específicos

- Determinar los factores que influyen en la recomendación de los clientes de la empresa Distecom.
- Identificar el comportamiento de los clientes para desarrollar un modelo predictivo con el propósito de mantener la rentabilidad del negocio.
- Validar el modelo de análisis predictivo de la empresa Distecom que permita pronosticar la referencia de los clientes, utilizando Phyton.

1.4. Justificación

Esta investigación se realiza con el propósito de aportar al conocimiento existente sobre los beneficios que brindan los modelos predictivos; los que constituyen un sistema que emplea



datos y estadísticas para predecir resultados a partir de una base de datos, utiliza técnicas estadísticas informatizadas, y su utilidad se debe a que proporciona información precisa a la hora de tomar previsiones.

Para lograr los objetivos de estudio, se acude al empleo de técnicas de investigación con el objetivo principal de entender y analizar información real y específica que sirva de apoyo a las técnicas cuantitativas válidas y propias de la ciencia.

Con los resultados obtenidos se tendrá en cuenta la posibilidad de proponer cambios que aporten soluciones a la problemática planteada para de esta manera lograr un incremento de ventas y un aumento de la rentabilidad de la empresa.



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación

William Rafael Raymondi Lomas

Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023





Capítulo 2. Marco teórico

2.1. Conceptos tecnológicos

2.1.1. Modelo predictivo

Para Chang et al. (2020) estos modelos usan, principalmente, un enfoque de patrones de cambio y evaluación mediante datos de percepción remota de análisis estadístico y conocimiento experto. Estos modelos han encontrado que en la mayoría existe un elevado nivel de incertidumbre; sin embargo, análisis más complejos pueden generar un enfoque útil para estimar la validez de una predicción.

Evidentemente, los modelos predictivos posibilitan el validar prospecciones realizadas de gran importancia; al momento de su elaboración, los modelos no tienen puntos de comparación y es probable que la simulación generada sea una subestimación o una sobreestimación derivada de la falta de comprensión de los procesos de cambio; sin embargo, análisis más complejos pueden generar un enfoque útil para estimar la validez de una predicción.

Por lo consiguiente, el autor alega que la función se basa, inicialmente, en entrenar un sistema usando datos propios de la empresa, el cual debe contener información relevante sobre el comportamiento de los clientes en este caso de Distecom. Posteriormente, se estima lo pronosticado, en él se debe usar patrones de servicios y productos actuales que oferta la compañía, de este modo, se predice o estima indicadores que se desconocen.

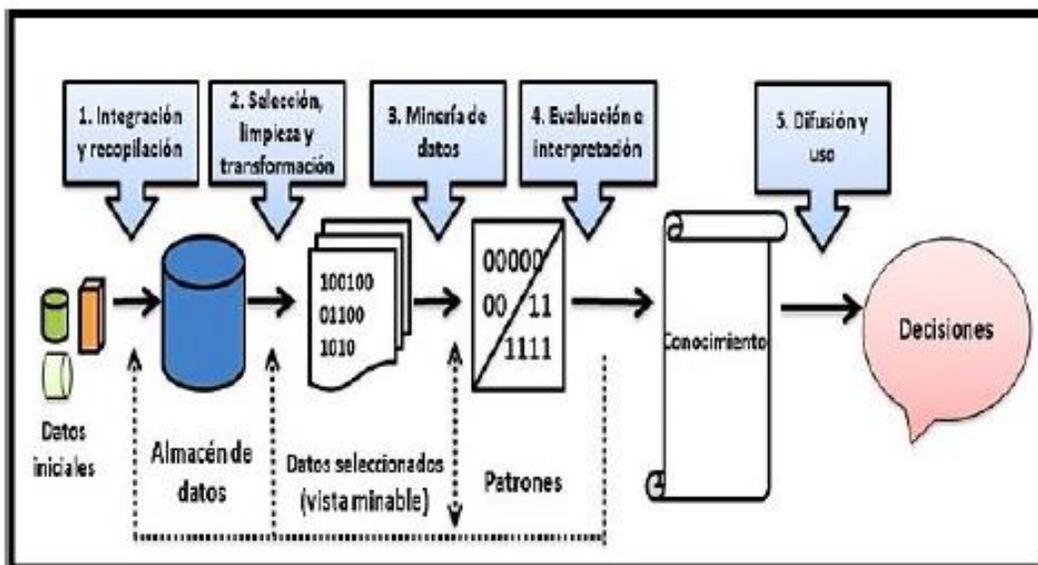


Figura 1 Ejemplo de modelo predictivo
Fuente: Maya et al. (2019)

Partiendo de la idea de Carvajal (2020), un modelo predictivo es especialmente útil en situaciones en que las empresas necesitan tomar decisiones rápidas basadas en grandes volúmenes de datos, sustentados en estadísticas inferenciales, que se utilizan para predecir situaciones.

En cierto sentido, el análisis predictivo es el conjunto de técnicas de inteligencia empresarial que elabora pronósticos a partir de la experiencia de la organización. En muchos casos, el modelo se elige sobre la base de la teoría de la detección para tratar de predecir la probabilidad de un resultado dada una cantidad establecida de datos de entradas. Estos modelos analizan los datos del pasado para determinar la probabilidad de ocurrencia de un evento determinado.

Por otro lado, Bedoya et al (2019) plantea que es un conjunto de procesos ejercidos a través de métodos computacionales de análisis de datos que favorecen al inferir la probabilidad de que ocurran situaciones explícitas previas a su obtención y, a su vez, divisar oportunidades de negocio.

Dentro de este contexto, los modelos predictivos pueden aplicarse a múltiples sectores porque centran sus esfuerzos en desarrollar esta rama del Big Data hasta el máximo nivel, al ser los encargados de tratar los datos y convertirlos en información de valor para la empresa, con el fin de predecir el comportamiento de los usuarios y evaluar el perfil de acierto que se puede obtener con las categorías de acciones en el mercado.

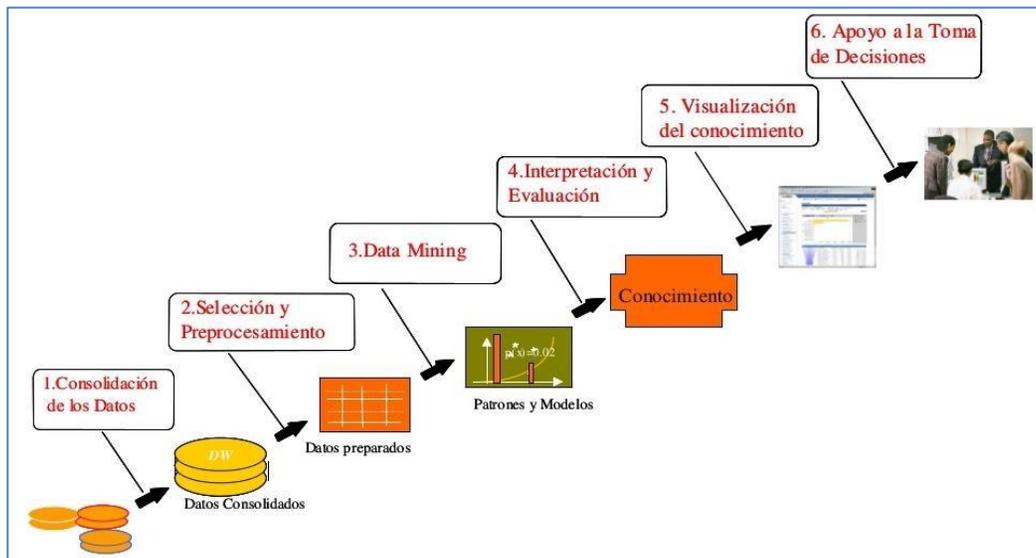


Figura 2 Proceso de los modelos predictivos

Es conveniente destacar que Ruiz (2018) afirma que un modelo predictivo permite valorar la relación que existe entre conjuntos de variables para posibilitar la toma de decisiones confiables. Efectivamente, el análisis predictivo construye un modelo estadístico que

requiere tener datos, atributos o comportamientos conocidos para predecir información. A través de ellos, se puede estimar valores futuros de variables de interés, que predigan las probabilidades de que un sujeto con ciertas características y bajo situaciones particulares reaccione de una forma determinada.

En todo caso, estos modelos predictivos, para poder llevar a cabo su misión, requieren de predictores y de la observación de los conjuntos de datos; aunque este no es el reto más complicado. El verdadero desafío para los modelos predictivos es encontrar adecuados subconjuntos de predictores o variables explicativas; es decir, hallar aquellos que aportan mayor utilidad, se ajustan mejor a los datos. El principio es que a mayor número de predictores y mayor profundidad en la investigación, aumentará la complejidad del análisis y la calidad de los resultados (Ruiz, 2018).

2.1.2. Importancia del modelo predictivo

Para Works (2020) los modelos predictivos se utilizan para crear un modelo matemático que capture las tendencias importantes. Se utiliza con los datos actuales para predecir lo que pasará a continuación o para sugerir acciones a llevar a cabo, con el fin de obtener resultados óptimos. A continuación, se detallan.

- Aumento del Big Data - Los datos de sistemas empresariales de una compañía podrían incluir datos de transacciones, resultados de ventas, quejas de clientes e información sobre marketing. A menudo se habla del análisis predictivo en el contexto del Big Data. Por ejemplo, los datos de ingeniería proceden de sensores, instrumentos y sistemas conectados

del mundo real. Cada vez más, los negocios toman decisiones basadas en los datos procedentes de esta valiosa mina de información.

- Aumento de la competitividad - Con el aumento en la competitividad, los negocios buscan una ventaja a la hora de proporcionar productos y servicios a mercados saturados. Los modelos predictivos basados en datos pueden ayudar a las empresas a resolver problemas de toda la vida de formas innovadoras. Los desarrolladores de productos pueden agregar capacidades predictivas a las soluciones existentes para aumentar el valor de cara al cliente.

- Tecnologías de vanguardia para Big Data y aprendizaje automático - Con el propósito de extraer valor del Big Data, los negocios aplican algoritmos a grandes conjuntos de datos mediante herramientas como Hadoop y Spark. Las fuentes pueden ser bases de datos transaccionales, archivos de registro de equipos, imágenes, vídeo, audio, datos de sensores u otros tipos de datos. Hay disponible una amplia gama de algoritmos de aprendizaje automático, incluidos regresión lineal y no lineal, redes neuronales, máquinas de vectores de soporte, árboles de decisión, etc. (Works, 2021).

Para Distecom, es valioso contar con un modelo predictivo que permita identificar la posibilidad en que los clientes refieran la empresa como una de las mejores opciones al momento de adquirir un producto o servicio tecnológico. De este modo, se podrá detectar posibles errores y dificultades existentes.

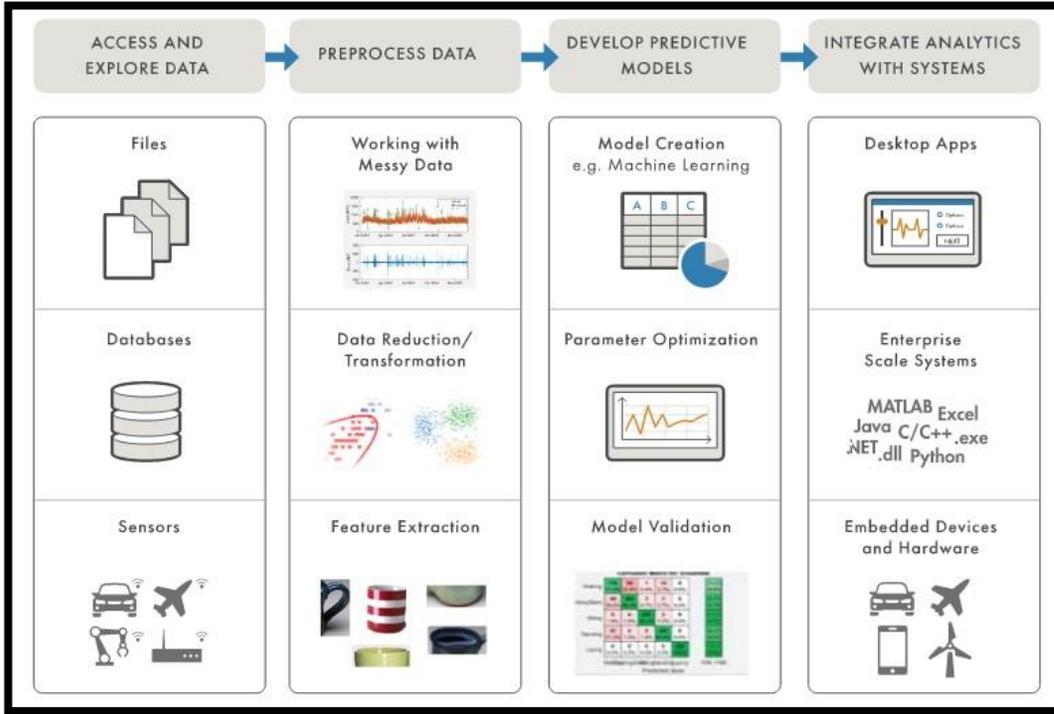


Figura 3. Importancia de los modelos predictivos

Fuente: Works (2021)

2.1.3. Tipos de modelos predictivos

Spain (2020) menciona que hay dos tipos de modelos predictivos, los modelos de clasificación y los de regresión.

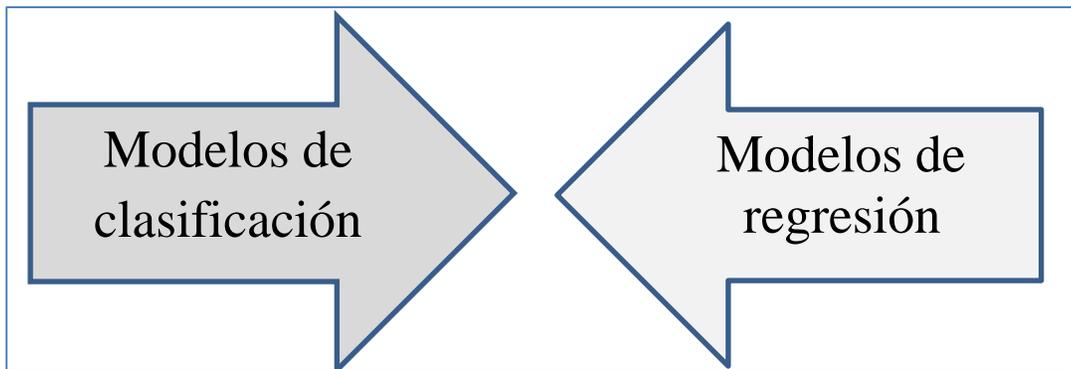


Figura 4. Tipos de los modelos predictivos

Fuente: Keyrus (2021)

Los modelos de clasificación permiten predecir la pertenencia a una clase. Por ejemplo, si tratamos de clasificar entre nuestros clientes quiénes son más propensos al abandono. Los resultados del modelo son binarios, o un sí o un no (en forma de 0 y 1) con su grado de probabilidad; o sea, nos pueden decir que un cliente nos abandonará con el 89% de probabilidad.

En conclusión, un modelo es una representación ideal de un sistema que tiene como objetivo analizar y predecir el comportamiento futuro. Ante ello, este tipo de modelo admite conocer la satisfacción del cliente con los productos y/o servicios que ofrece la empresa Distecom.

- Los modelos de regresión en cambio nos permiten predecir un valor. Por ejemplo, cuál es el beneficio estimado que obtendremos de un determinado cliente (o segmento) en los próximos meses o nos ayudan a estimar el *forecast* de ventas (Keyrus Spain, 2020).

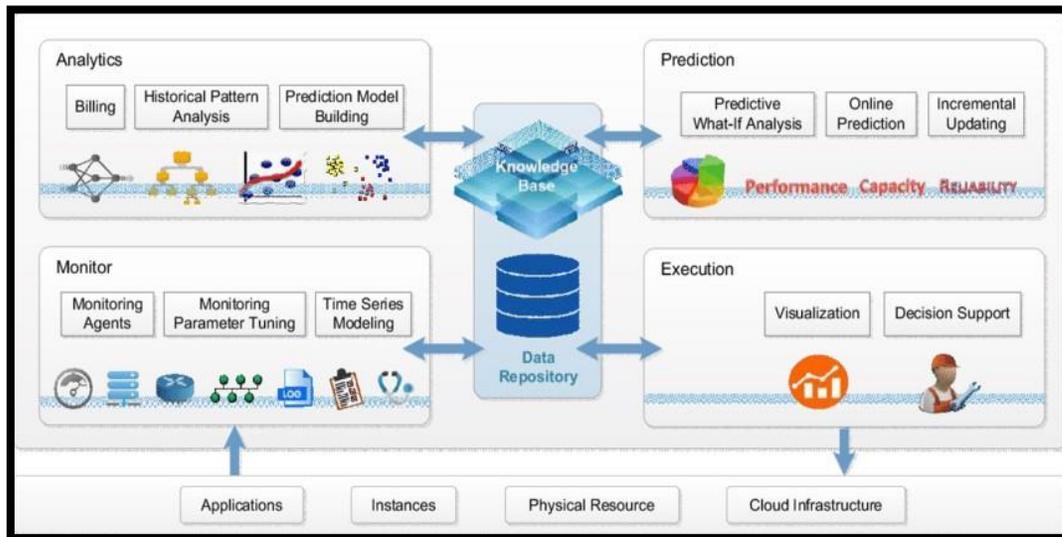


Figura 5. Tipo de modelos predictivos (de regresión)
 Fuente: Keyrus (2020)

Por tanto, esta técnica estadística brinda la oportunidad de extraer datos importantes de los objetos de estudio. En concordancia, interviene una variable dependiente y dos o más variables que en lo posterior se relacionan, a su vez, facilitan un conjunto de variables explicativas, tiene como objetivo final determinar el impacto sobre las variables.

2.1.4. Big Data

Para Altiuz (2020) existen modelos predictivos para diferentes industrias, sucede para distintas áreas en las empresas. Este autor se refiere a un gran conjunto de datos o combinación de estos, estructurados o no estructurados, que crece aceleradamente día a día. Esto hace que su análisis manual sea prácticamente imposible, las compañías ya no confían en la intuición y buscan que sus decisiones estratégicas, operacionales y comerciales sean precisas y personalizadas para cada consumidor a través del estudio de su comportamiento.

Según lo indagado en el transcurso de la investigación, se establece que es un término que representa un gran volumen de datos. En este punto, lo significativo no es tener un mayor número de contactos, lo que interesa es saber, qué hará la empresa Distecom con los datos adquiridos. Asimismo, soportan la toma de decisiones y las acciones estratégicas que se llevarán en el negocio.

2.1.5. Beneficios

Lo más importante que destaca el autor Altiuz (2020) son los beneficios. A continuación, se mencionan.

- “Mejorar la satisfacción de los clientes o *customer experience* (CX) por medio de análisis de reclamos buscando encontrar similitudes entre las sensaciones y opiniones de los clientes.
- El análisis de Big Data hace posible contar con información relevante sobre la industria y los mercados, se logra que las empresas pronostiquen ventas e identifiquen de qué manera pueden reducir sus costos.
- En términos de operaciones y producción, es muy útil realizar análisis predictivos de Big Data, ya que gracias a la información histórica las empresas pueden estudiar la demanda y optimizar sus procesos de logística y despachos” (p. 53).



Figura 6. Big Data

Fuente: <https://es.mailjet.com/blog/news/big-data/>

Según el citado autor, el Big Data discurre en dimensiones tales como la variabilidad y la veracidad; la primera atiende a crecientes velocidades y variedades de datos, considerando que los flujos de datos son impredecibles, cambiantes, mientras que el segundo obedece a la calidad de datos.

2.1.6. Inteligencia de negocios

El conocimiento empresarial nace de la proyección y evaluación que plantea la compañía. Ante ello, se estipula como objetivo la creación de plan de negocios avalado por un documento formal en el cual se detallan los objetivos a alcanzar, las herramientas y métodos a aplicar. Asimismo, abarca la percepción de su propia práctica y capacidad de interpretar información almacenada en un período determinado.

Por lo consiguiente, el conocimiento se deriva de información la cual se deriva de los datos. Vale mencionar, que existe una relación directa entre datos, información y conocimiento. Si la información se transforma en conocimiento, entonces ha ocurrido la intervención de un ente inteligente, Castallanos (2006) interpreta que “la visión cosmogónica del conocimiento varía de acuerdo con el observador y su propia apreciación de la realidad en la que se desenvuelve la empresa”; no obstante, Castallanos (2006) cuando cita a Pérez plantea: “En un sentido muy real, el científico no descubre la realidad, sino que la inventa, y por medio de observaciones y experimentos que también contiene fuertes elementos subjetivo; establece y perfecciona un grupo de modelos o teorías posibles ...”

A partir de la gestión del conocimiento, surge el concepto de inteligencia de negocios (*Business Intelligence*, inteligencia empresarial o inteligencia de negocios) que hace

referencia al conjunto de estrategias, acciones y herramientas orientadas a la dirección y creación de conocimiento mediante el análisis de datos existentes de la empresa Distecom S.A.

En las organizaciones se presenta una situación alrededor del manejo de la información, que cada vez es más numerosa y difícil de categorizar. Es por ello, que el presente proyecto aborda la problemática de lograr referencias por parte de los clientes hacia la empresa Distecom S.A., aquellas que mediante el uso de estrategias basadas en el conocimiento desarrollan acciones de inteligencia de negocios o, en otras palabras, que sean capaces de aprender a partir de sus miembros y sostener este conocimiento para la generación y transferencia de valor Berg de Valdivia (2007). Asimismo, se argumenta que coexisten tres elementos que intervienen en la conceptualización de negocios; estos son los sistemas de información, el mecanismo de innovación y la toma de decisiones de ellos.

Siguiendo la perspectiva del valor empresarial, se presenta la necesidad de determinar qué impacto tiene el aplicar los modelos predictivos para determinar e identificar la posibilidad de recomendación de los clientes. De este modo, se determinan acciones que se perfilan a fortalecer y optimizar la gestión de ventas y tiempo.

2.1.7. Principales características de la inteligencia de negocio

En efecto, existen varias características de la inteligencia de negocios, según lo menciona la Universidad Internacional de Valencia (2018), podemos elaborar diferentes aspectos, entre los que se destacan las promociones, las planificaciones, el control financiero, la



optimización de los costes o incluso las tendencias de las ventas de productos en el mercado. A continuación, se detallarán.

- Es una herramienta fundamental para la planificación empresarial; entre otras cosas, porque cuenta con unas características específicas que hacen que los negocios puedan obtener una información muy valiosa para mejorar su rendimiento.
- Las empresas pueden tomar las decisiones adecuadas teniendo en cuenta información veraz y fiable.
- Existe distinción entre los conocimientos técnicos que se pueda tener y las herramientas que se utilizarán para obtener la información deseada.
- El *Business Intelligence* ofrecerá la oportunidad de poder identificar todos los productos y clientes de un mercado específico; así como las ventas que se están produciendo o aquellos sectores en los que no se está teniendo buenos resultados.

2.1.8. Phyton

Es un lenguaje de programación denominado multiparadigma, el cual contiene visualización de datos y aprendizaje automático de código abierto. Crea flujos de trabajo de análisis de datos de forma visual, con una caja de herramientas amplia y diversa.

2.1.9. Flujos de trabajo naranja

En este flujo de trabajo se exploran los arcos de la historia, particularmente, la de Little Match Seller. Primero se selecciona la historia del corpus de cuentos de Andersen. Luego se crea una tabla, donde cada oración del cuento es una fila separada. Posteriormente, se utiliza el análisis de sentimientos para calcular el sentimiento de cada oración, luego se observan los arcos emocionales a través de la historia.

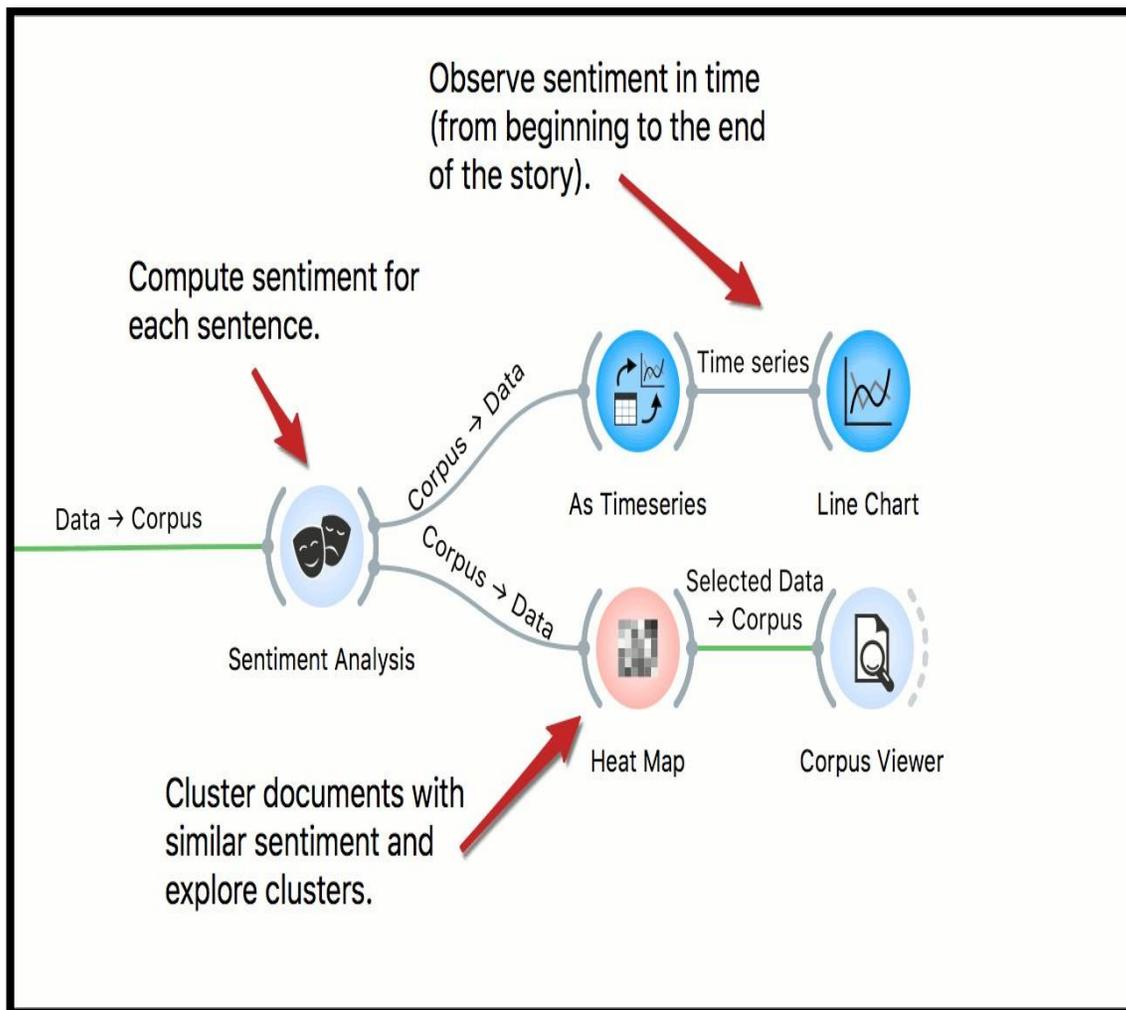


Figura 7. Arcos de la historia

Análisis de datos de Twitter - Los tweets son una valiosa fuente de información para científicos sociales, gerentes de marketing, lingüistas, economistas, etc. En este flujo de trabajo se recuperan datos de Twitter, se los procesa previamente y se descubren temas latentes con el modelado de temas. Se observan los temas en un mapa de calor.

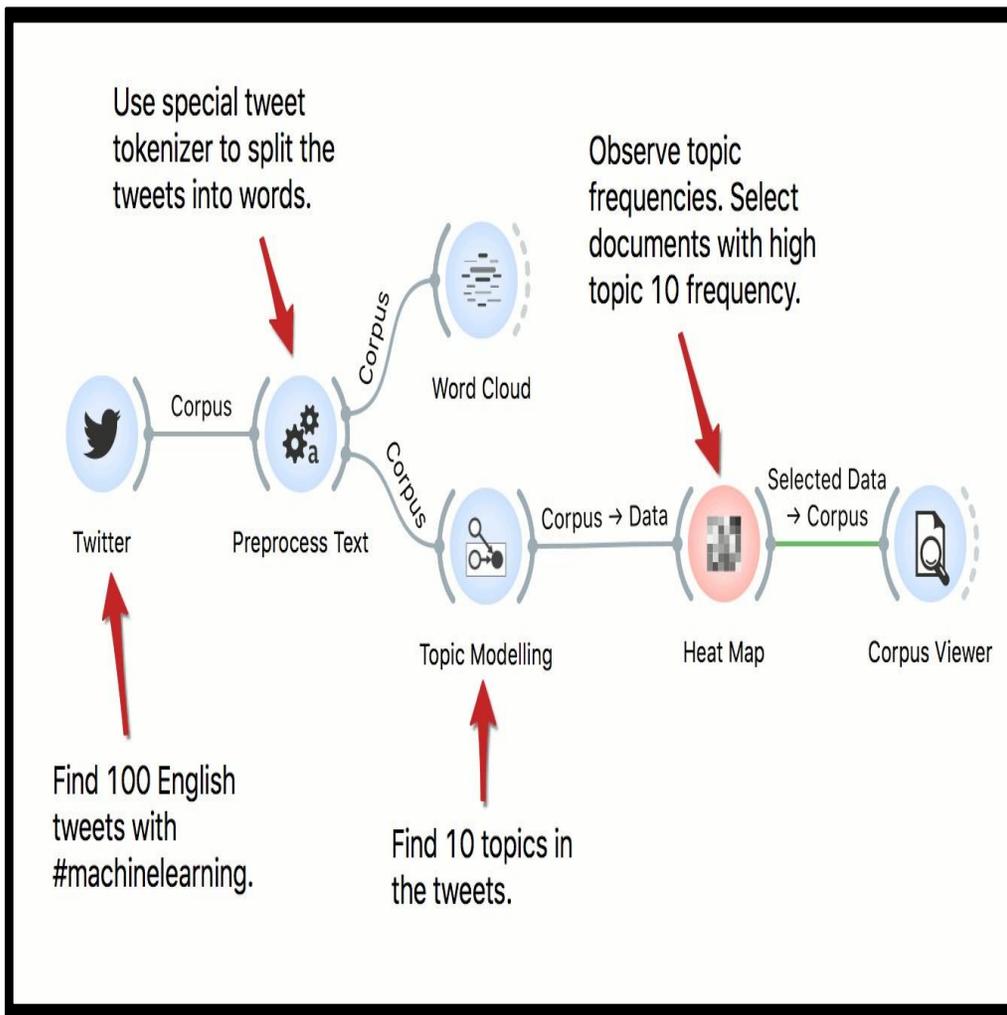


Figura 8. Análisis de datos de Twitter

Clasificación de texto. Se puede utilizar modelos predictivos para clasificar documentos por autoría, tipo, sentimiento, etc. En este flujo de trabajo, se clasifican los documentos por su índice *Aarne-ThompsonUther*, que es el tema definitorio del cuento. Se utilizan dos aprendices simples, regresión logística y bayes ingenuo, los cuales se pueden inspeccionar en el nomograma.

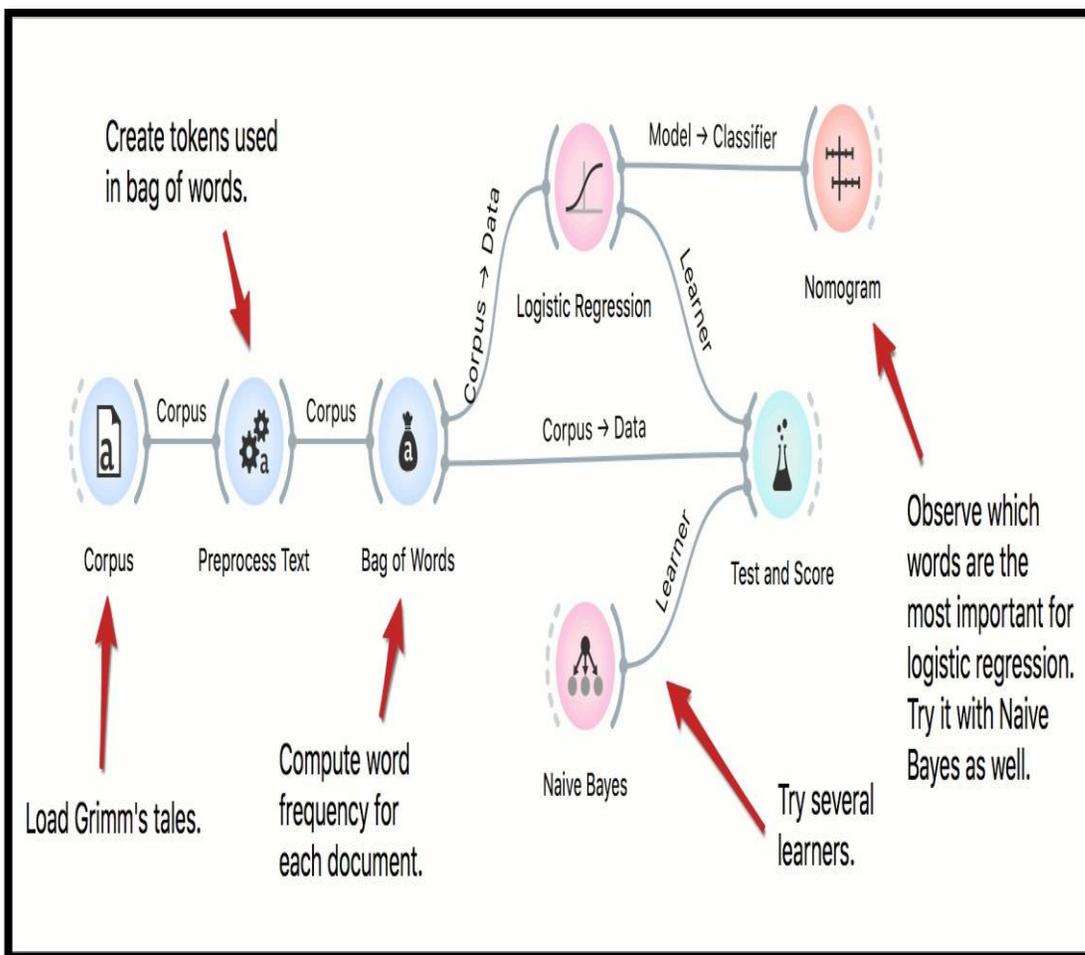


Figura 9. Clasificación de texto

Agrupación de texto - El flujo de trabajo agrupa el corpus de cuentos de Grimm. Se inicia con el procesamiento previo de los datos y construcción de la matriz de la bolsa de palabras. Luego, se calculan las distancias de coseno entre documentos y se utiliza la agrupación jerárquica que muestra el damerograma. Se observan qué tan bien se corresponde el tipo de cuento con el clúster en el MDS.

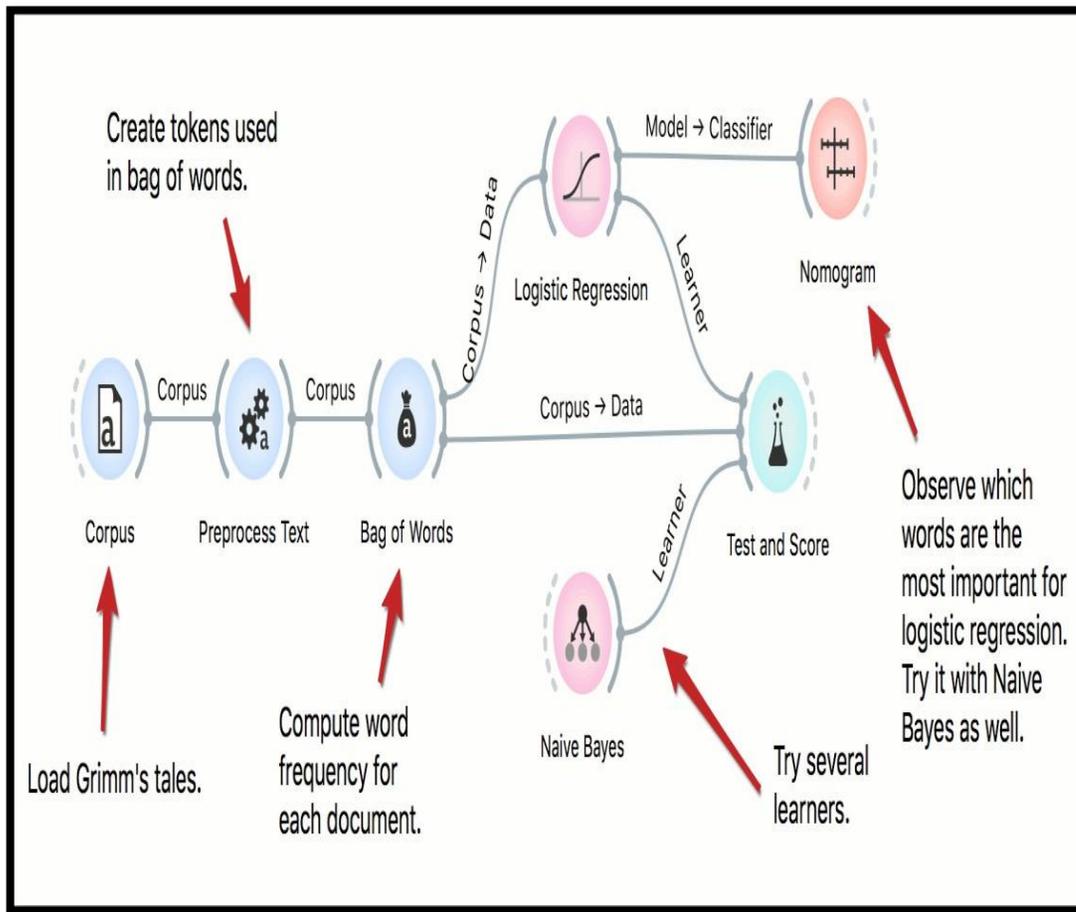


Figura 10. Agrupación de texto

Preprocesamiento de texto - La minería de texto requiere un preprocesamiento cuidadoso. Aquí hay un flujo de trabajo que utiliza un preprocesamiento simple para crear tokens a partir de documentos. Primero, aplica minúsculas, luego divide el texto en palabras, y finalmente, se eliminan las palabras vacías frecuentes.

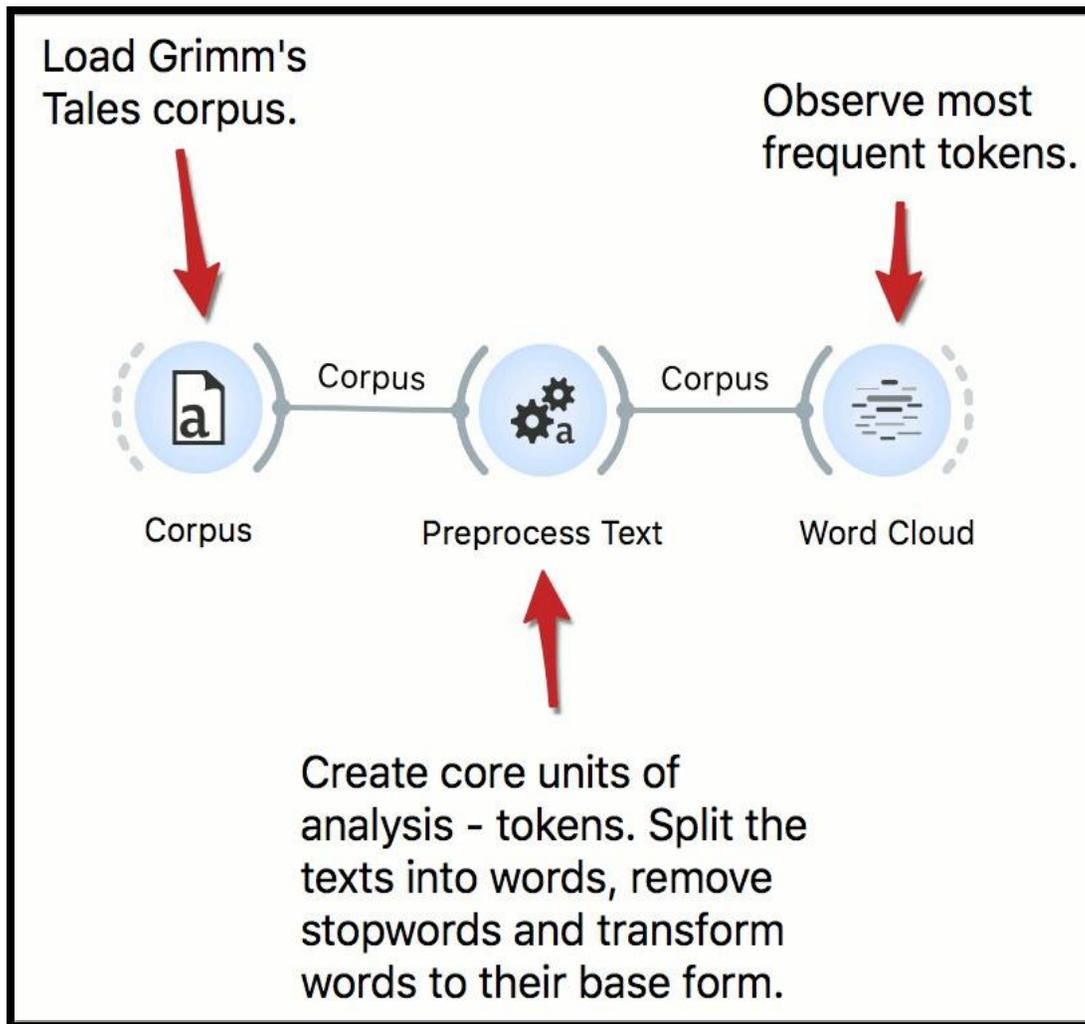


Figura 11. Preprocesamiento de texto

Clasificación errónea - la regresión logística puede arrojar, en primer instancia, datos de manera errada. Ante ello, se puede clasificar en matriz de confusión y resaltarlos en el gráfico de dispersión.

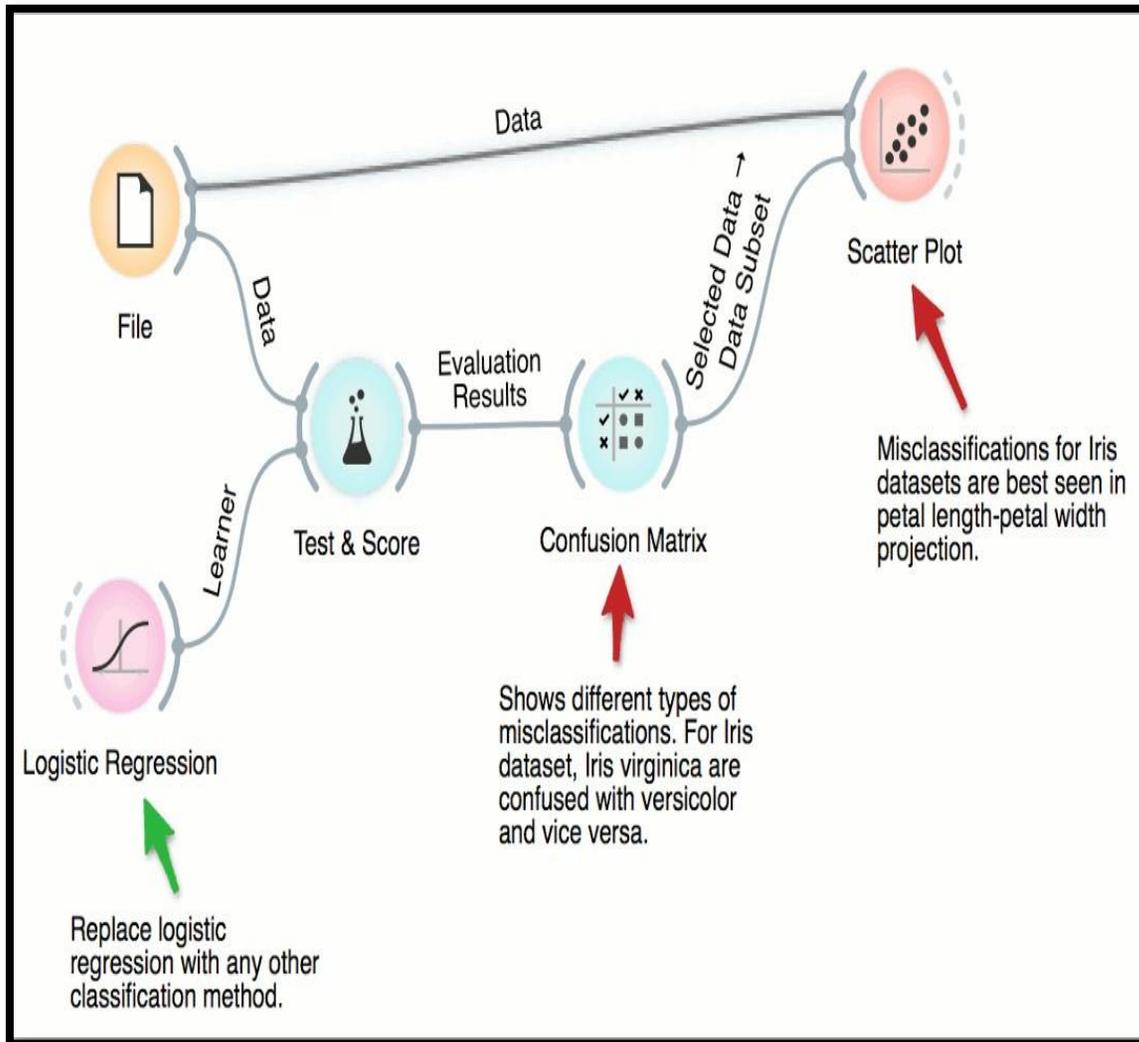


Figura 12. Clasificación errónea



Validación cruzada – técnica empleada para evaluar un flujo de trabajo, se analiza estadísticamente y garantiza la calificación de los resultados obtenidos, widget central se encarga de las pruebas y puntuación, recibe los datos de los estudiantes, se realiza la validación cruzada y puntúa la precisión predictiva; posteriormente, se generan las puntuaciones para un examen más detallado.

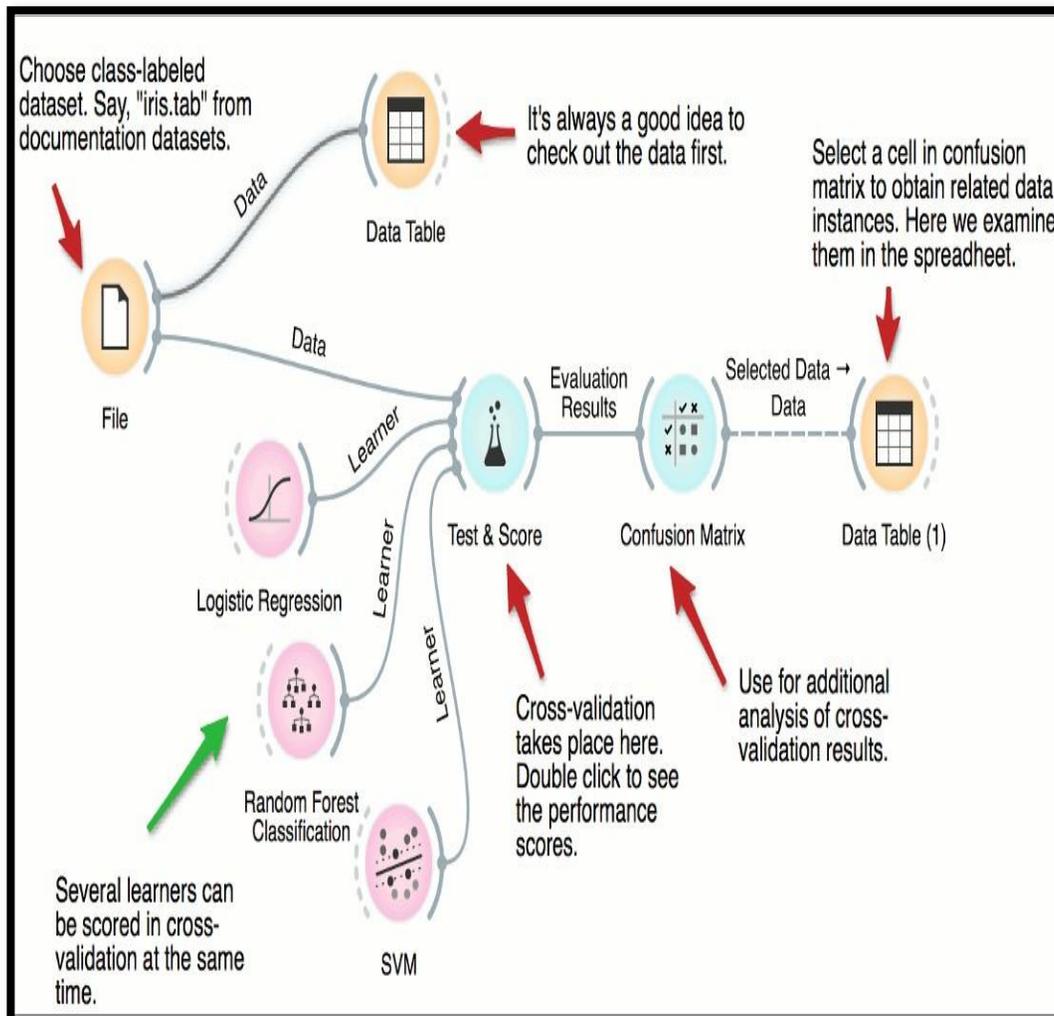


Figura 12 Validación cruzada

Clasificación de funciones - Para problemas supervisados, donde las instancias de datos se anotan con etiquetas de clase, es importante saber cuáles son las características más informativas. El widget de clasificación proporciona una tabla de características y sus puntajes de informativa, y admite la selección manual de características. En el flujo de trabajo, se lo usa para encontrar las dos mejores características (del 79 inicial del conjunto de datos seleccionado en marrón) y mostrar su diagrama de dispersión.

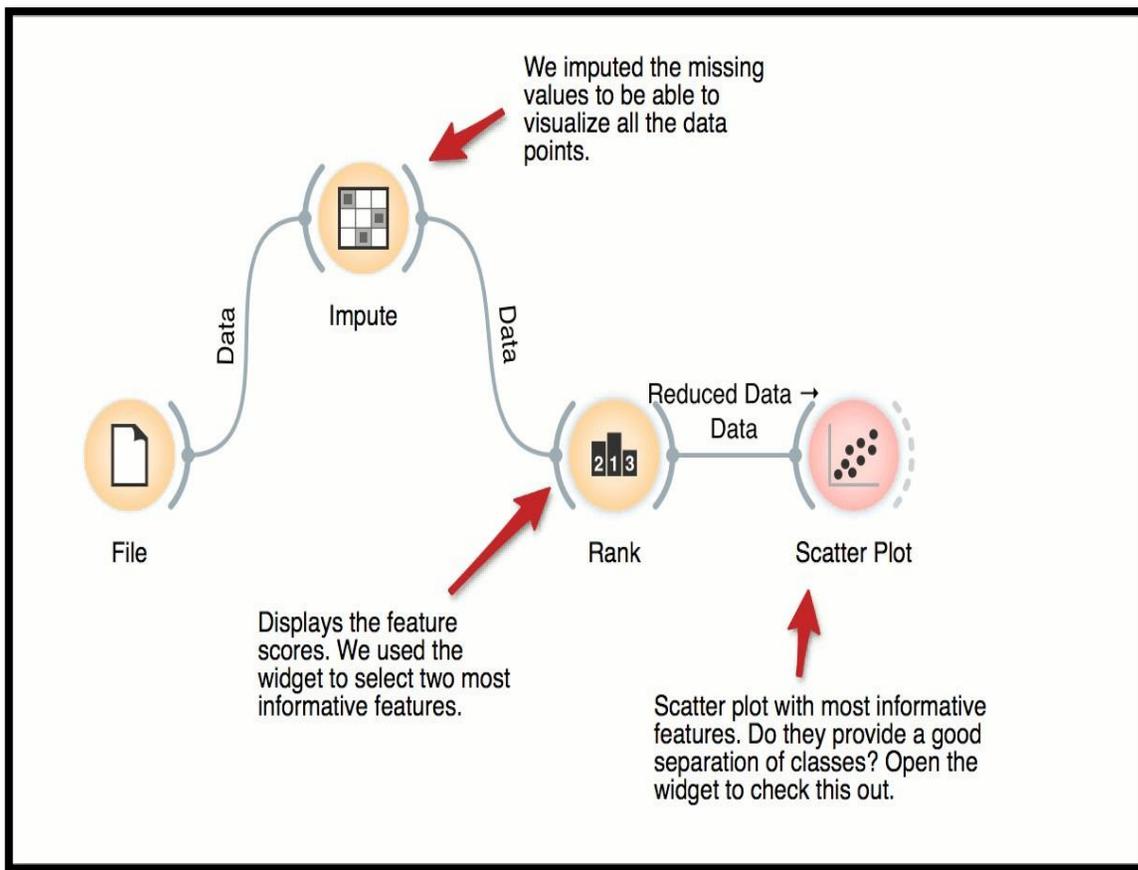


Figura 13. Clasificación de funciones

Inspección de racimos – Se utiliza el conjunto de datos del zoológico en combinación con la agrupación jerárquica para descubrir grupos de animales. ¡Ahora que se tienen los grupos, se busca descubrir qué es importante para cada grupo! Pase los grupos a Box Plot y use 'Ordenar por relevancia' para descubrir qué define un grupo.

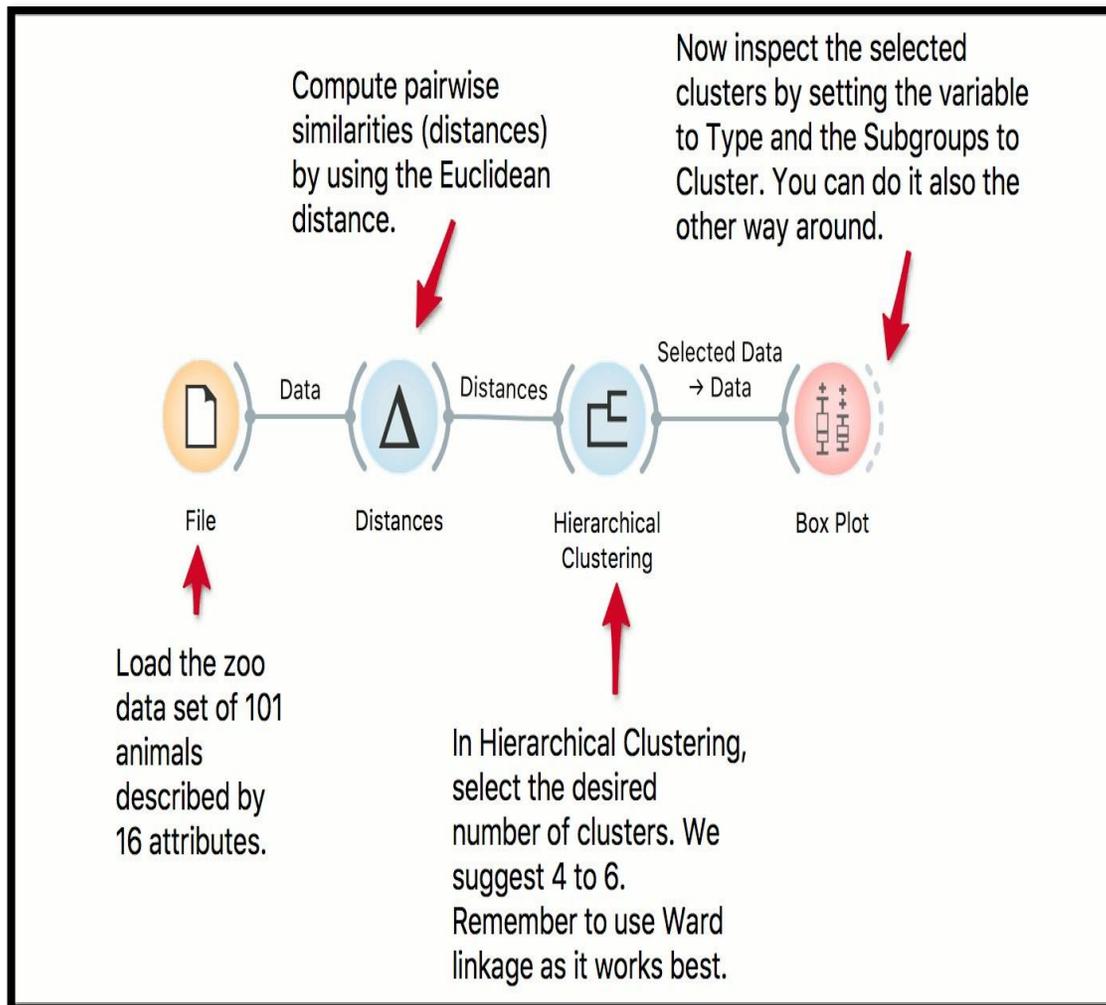


Figura 14. Inspección de racimos



Agrupación jerárquica - El flujo de trabajo agrupa los elementos de datos en un conjunto de datos de iris examinando primero las distancias entre las instancias de datos. La matriz de distancia se pasa a la agrupación jerárquica, que genera el dendrograma. Seleccione diferentes partes del dendrograma para analizar más a fondo los datos correspondientes.

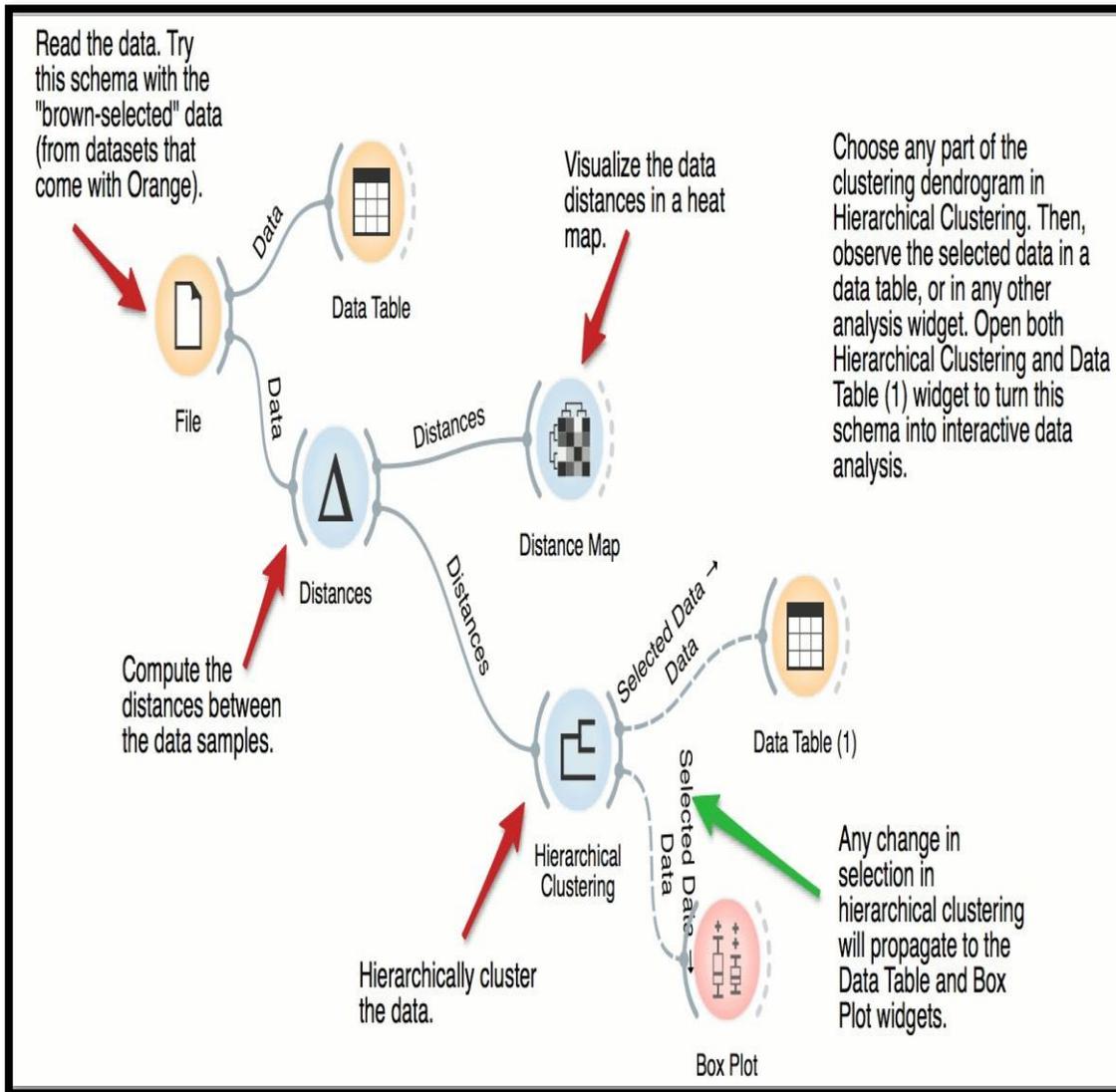


Figura 15. Inspección de racimos

2.1.10. Ventas por medios de referidos

Como lo menciona Larc (2016), todo cliente referido es la evaluación de tu desempeño en el área de ventas o comercial. Este punto mide desde la presentación del producto hasta la fase de posventa. Siendo justamente aquí donde se gana la fidelidad del cliente, y este a su vez, refiere nuestros servicios a otras personas convirtiéndose así en referidos; También se mide qué tan bueno son los servicios y productos que se ofrecen, y la construcción de relaciones de negocios con los clientes. A veces de manera errónea, se piensan que los referidos se obtienen cuando algún cliente brinda el nombre de algún comprador potencial siendo esto falso e incluso se olvidan de evaluar la posventa.

2.1.11. Maneras de lograr clientes referidos

Para Gálvez (2020) existen cinco puntos indispensables para lograr que el cliente que ya ha visitado un negocio con anterioridad pueda referirlo a sus allegados o personas más cercanas.

- Conviértete en amigo de tus clientes. Reúnete con ellos, en eventos de redes de negocios, en el club deportivo, en una comida informal, que no necesariamente implique juntas de negocios clientes que son tus amigos, de manera natural quieren ponerte en contacto con sus amigos.
- Provee un buen servicio, asegúrate de que tú y tu compañía proveen la más alta calidad en tu industria. La credibilidad resultante reduce en gran medida el riesgo de hacer negocios contigo.

- Se proactivo con sus necesidades, anticipa lo que les gustaría a tus clientes y haz todo lo que sea posible para tenerlo listo antes de que tu cliente lo requiera.
- Sorprende a tu cliente, encuentra algo adicional a lo que la espera que puedas hacer por él, por ejemplo, envíale información actualizada sobre algo que sea de su interés y que pueda serle útil.
- Dales referidos a tus clientes. Muéstrale en dónde puede hacer nuevos negocios, es muy importante que esto lo hagas realmente de corazón porque de lo contrario tu cliente lo notará. Si quieres ofrecer referidos a cambio de algo, terminarás pareciendo manipulador (p. 69)

La satisfacción de los clientes con los productos y/o servicios que ofrece la empresa es el primer paso y uno de los más importantes, para que el cliente pueda referir; no obstante, participan diferentes aspectos que se deben asociar, por ejemplo, la atención al cliente y la posventa que es una fase trascendental debido a que muchos se conforman con vender una sola vez y no dan el seguimiento debido. Esta acción permite evaluar o controlar la satisfacción del cliente.

2.1.12. Tipos de referidos

En palabras de Casermeiro (2014), no todos los referidos son iguales, algunos son más valiosos que otros. Ordenados del menos valioso al más valioso, aquí están los tres tipos de referidos.

- Solicitados: es cuando pides y recibes un referido de un cliente actual. En este tipo de referidos solicitados, el truco está en “preguntar sin preguntar”; es decir, que le digas a tu cliente que piense en alguien que también pudiera hacer negocios contigo, pero antes de que te dé el nombre.
- Proactivos: sin preguntar te llama para darte el nombre de un cliente potencial. Este es mejor que un referido solicitado porque indica que has construido una relación de confianza y libre de riesgos con tu cliente.
- No-solicitados: un prospecto te llama porque uno de tus clientes lo ha recomendado; esta es la mejor clase de referidos porque indica que tu cliente te tiene tanta confianza, que ahora es un vendedor más para tu compañía.

Existen diferentes leads que deben ser diferenciados al momento de almacenar los datos, error que muchas empresas cometen es guardar los diferentes referidos en un mismo archivo. Estos pueden ser registrados como renovación, *outdoor*, entre otros.

2.1.13. Pasos para obtener referidos

Carrillo (2020) en el artículo científico estipula que se deben seguir los siguientes pasos para la obtención de referidos.

- Identifica a clientes con altos contactos - Hay personas que son más conectadas que otras; es decir, describe a este tipo de personas como conectores.
- Determina incentivos - Hacer una lista de incentivos que podrías dar a tus clientes, amigos o proveedores por recomendar o vender, es por lo que se puede crear links de afiliados (si

vendes productos de forma digital). Puedes utilizar plataformas con WooCommerce en WordPress o Shopify. No obstante, no todos los incentivos deben ser monetarios, ni relacionados a descuentos.

- Crea un sistema de recordatorios - Envía correos o WhatsApp recordando a tus clientes los beneficios de recomendarte, también puedes pedir a tus clientes nombres de referidos para que tú llames directamente a la persona.

- Hazlo fácil de compartir - Es fácil convencer a un amigo de que vaya a una clase gratis de prueba, facilita la decisión de obtener una membresía posterior y pague un año entero.

- Haz que tu producto o servicio tenga algo que valga la pena comentar - Si el producto o servicio es excelente, será muy difícil hacerlo significativamente mejor en el corto plazo.

La experiencia que recibe tu cliente, por otro lado, siempre se puede mejorar (p. 57)

2.1.14. Beneficios de la venta por referidos

Laseca (2016) argumenta los siguientes beneficios de ventas por referidos.

- Es más fácil conseguir reuniones con prospectos referidos - Cuando vienes indicado o referido por una persona conocida por el prospecto has recibido un crédito o préstamo del activo más valioso en ventas: la confianza.

- El porcentaje de operaciones ganadas es superior en venta por referidos - Las oportunidades de venta generadas a través de referidos duplican la ratio de operaciones ganadas con prospectos adquiridos por otras fuentes.

- Los clientes referidos son menos sensibles a precio - Cuando un prospecto reclama precio está transmitiendo de manera implícita que el valor percibido por tu solución es inferior al dinero que le cuesta.
- Las fuentes de referidos se fidelizan a medida que “trabajan” para ti. - Cuando realizamos una “inversión” personal en algo o alguien sentimos una necesidad de ser consistentes o coherentes con esa decisión para justificarnos ante nosotros mismos que hemos hecho lo correcto (p. 39).

2.2. Metodología

2.2.1. CRISP-DM

Según Villena (2016), el CRISP-DM “... es el proceso de análisis de datos como un proyecto profesional, ya que, establece un contexto que influye en la elaboración de los modelos”.

Este modelo cubre las fases de un proyecto, sus tareas respectivas y las relaciones entre estas tareas. Sin embargo, considerando el nivel de descripción, no es posible identificar todas las relaciones; es decir, no será posible identificar las relaciones que podrían existir entre cualquier tarea según los objetivos, el contexto, y el interés del usuario sobre los datos (p. 67).

Por otro lado, Espinosa (2020) argumenta que este método “es el más empleado para el proceso de proyectos de minería de datos”. La metodología CRISP-DM consta de seis etapas.

- Conocimiento del problema o negocio
- Comprensión de datos
- Preparación de datos
- Modelado
- Evaluación del modelo
- Implementación del modelo (p. 32).

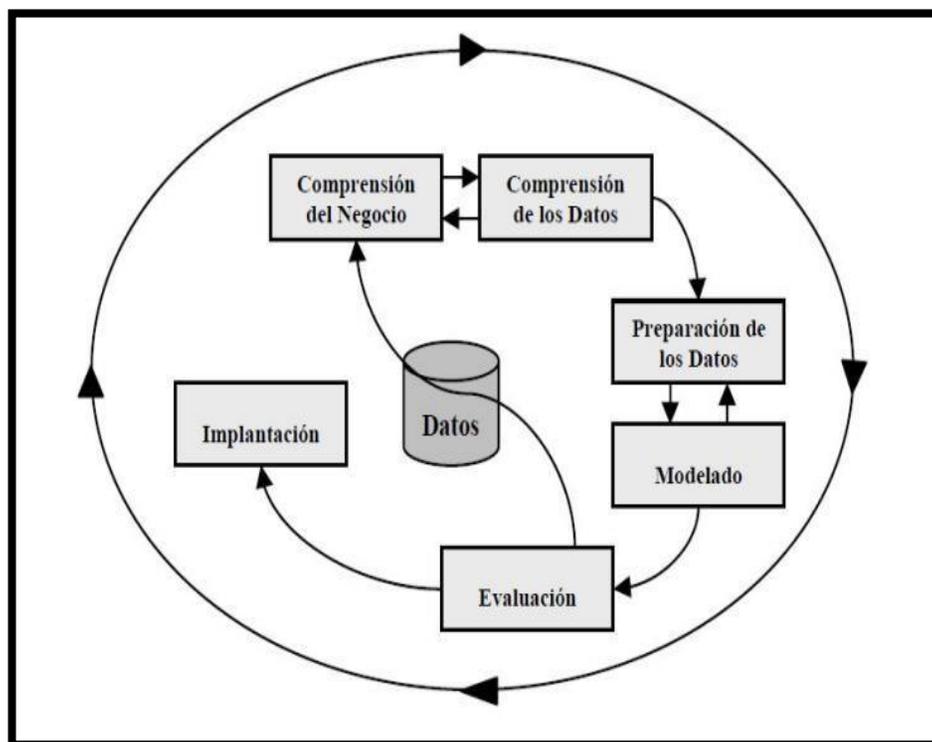


Figura 17. Fases de la CRISP-DM
Fuente: Rodríguez (2016)

Conocimiento del problema o negocio: dentro de esta etapa es importante tener en cuenta los problemas que surgen o giran en torno al negocio; ya que, si no se tiene completo conocimiento del problema, de nada servirán las fases siguientes.

- Identificación del problema
- Establecer los objetivos
- Análisis de la situación actual

Comprensión de datos: son importantes las fases se detallan a continuación.

- Recolección de datos
- Descripción de datos
- Exploración de datos

Preparación de datos: es indispensable esta etapa; ya que seleccionan los datos más relevantes que se ajustan a la problemática con el propósito de transformarlos de acuerdo con los resultados obtenidos en la fase anterior, a su vez, utilizarlos en la siguiente etapa, por lo que se describe en las siguientes actividades.

- Limpieza de datos
- Creación de indicadores
- Transformación de datos

Modelado: esta fase recoge los datos obtenidos en las fases anteriores con el fin de diseñar el modelo de minería de datos. Para ello, se deben realizar las siguientes actividades.

- Selección de técnica de modelado
- Selección de datos de prueba
- Obtención de modelo

Evaluación del modelo: la fase de evaluación determina la efectividad y calidad del modelo sobre la base de un análisis métrico estadístico; asimismo, comparando los resultados previos con resultados actuales y con el apoyo de un experto dentro del dominio de la problemática.

Implementación del modelo: esta es la fase final del modelo CRISP-DM donde se proponen acciones concretas y se observa el conocimiento adquirido a través del modelo. Por lo que, es importante documentar los resultados de forma clara y concreta para que el usuario final pueda asegurar los datos que están debidamente documentados con el propósito de hacer una revisión y análisis de proyecto final.

2.2.2. Kimball

Para Túllume (2016) el Kimball "... se enfoca principalmente en el diseño de la Base de Datos que almacenará la información para la toma de decisiones". La principal ventaja de esta metodología es el almacenamiento de datos; ya que está formado por pequeños datamarts, las cuales constituyen la estructura del modelo en diversas dimensiones. Estos se encuentran diseñados para la consulta y generación datos, por lo que el datawarehouse debe

ser completo para que pueda ser exportado directamente por las herramientas de reporting

(p. 49).

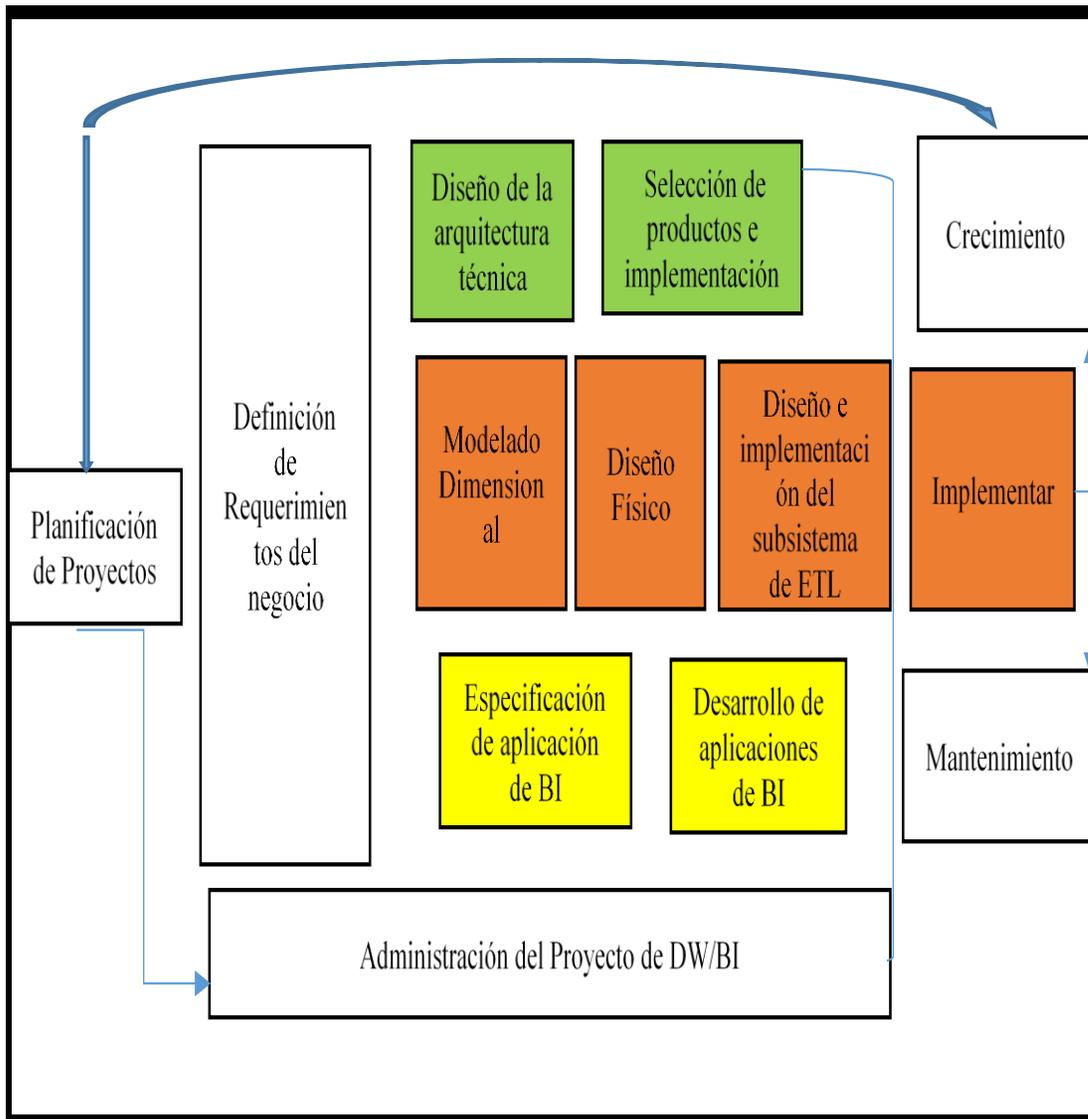


Figura 18. Fases de Kimball
Elaboración: propia

Por otra parte, Calzada (2015) expresa que esta metodología se adapta adecuadamente a las circunstancias, por tanto, la empresa debe prepararse y estipular estrategias para contrarrestar los efectos que puedan perjudicarlo.

Permite certificar los datos que vamos a utilizar, y rechazar aquellos datos que no cumplan con los parámetros que no cumplan lo que se requiera e incluso crear procesos para su corrección y posterior inclusión en el DW, filtrarlos y agregarlos como consideremos oportuno. Se aboga por una aproximación de abajo hacia arriba primando el contacto con el usuario final (p. 67).

Por otra parte, Martínez (2020) se basa en lo que se denomina ciclo de vida dimensional del negocio. Este ciclo de vida del proyecto está basado en cuatro principios básicos.

- Centrarse en el negocio: es indispensable analizar e identificar los diversos requerimientos que la empresa necesite solventar y su valor asociado para desarrollar relaciones sólidas con el negocio, agudizando su análisis y la competencia consultiva de los implementadores.
- Construir una infraestructura de información adecuada: el diseño de una base de información única es significativa para el alto rendimiento de la empresa, donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
- Realizar entregas en incrementos significativos: crear el almacén de datos (DW) en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos.

- Ofrecer la solución completa: proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios, esto significa tener un almacén de datos sólidos, bien diseñados, con calidad probada, y accesible (p. 57).

2.2.3. Hefesto

Según Chagcha (2016), Hefesto es una metodología propia, cuya propuesta está fundamentada en una muy amplia investigación, en comparación con metodologías existentes y experiencias propias en procesos de confección de almacenes de datos.

La idea principal es comprender cada paso que se realizará para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo, ni por qué. Lo que se busca es entregar una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios (p. 84).

Por otra parte, Silva et al (2019) expresa que posee las características listadas, a continuación.

- En cada una de las fases se puede distinguir fácilmente los objetivos que se persiguen, así como los resultados esperados; adicionalmente, son de fácil comprensión.
- La estructura del Data Warehouse es de fácil y rápida adaptación, esto se debe a que fue construida sobre la base de los requerimientos de los usuarios.
- La resistencia que presentan los usuarios finales al cambio se ve reducido gracias a que en cada etapa se los considera para determinar el comportamiento y las funciones que se incorporan al diseño del DW.

- Los modelos conceptuales y lógicos que se implementan son de sencilla comprensión y análisis.
- El tipo de ciclo de vida que contenga a la metodología marcan independencia el uno del otro.
- Las herramientas que se utilicen para la construcción del DW son independientes de la metodología
- La metodología es independiente de las estructuras físicas y su distribución, que contengan el DW.
- Los resultados obtenidos al finalizar una fase se convierten en un nuevo punto inicial para el siguiente paso.
- Se aplica tanto para el Data Warehouse como para los Data Mart. (p. 405)

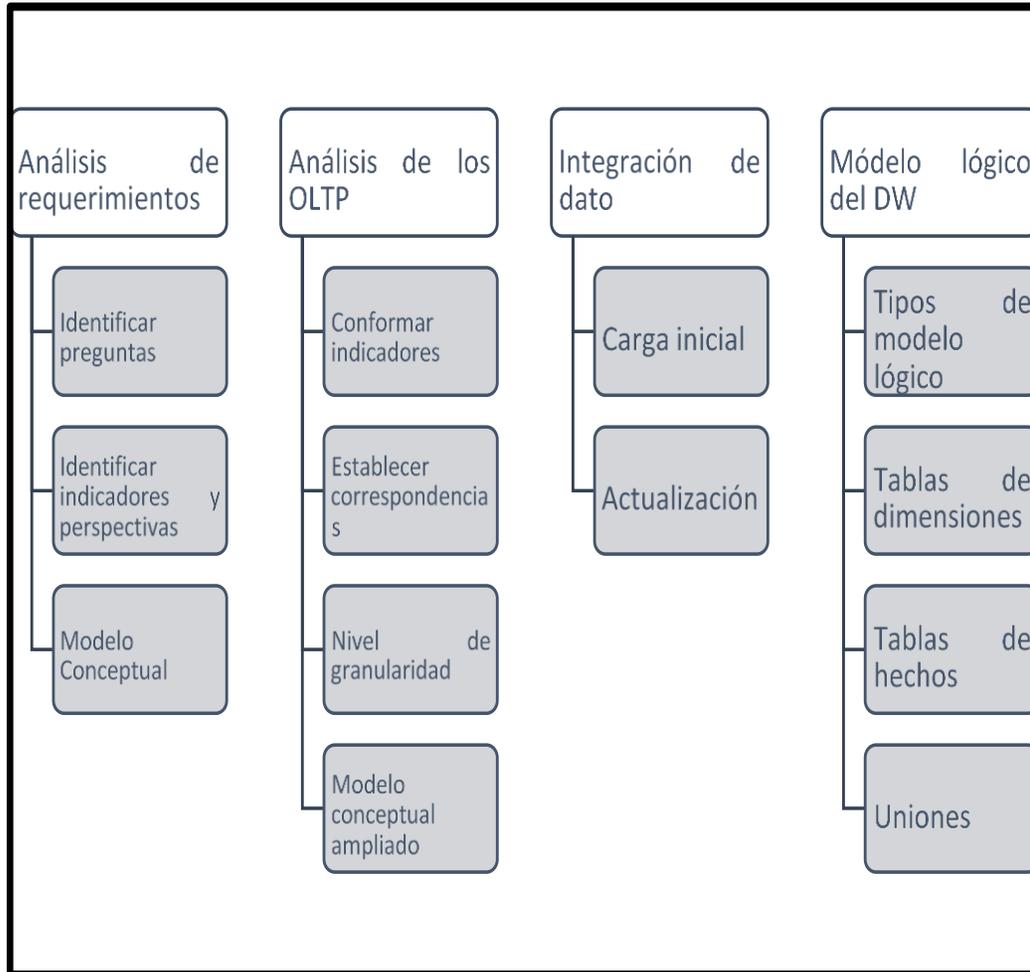


Figura 16. Fases de la metodología Hefestos
 Elaboración propia

Para Cabrera (2017), la metodología de Hefesto contempla cuatro pasos para la construcción de un Datawarehouse o Datamart: el análisis de requerimientos, el análisis de los OLTP, el modelo lógico del DW y la integración de datos. Hefesto “... es considerado como metodología por su simpleza y su didáctica para realizar este tipo de sistemas, a más que se puede estructurar esta metodología con un método ágil de desarrollo” (p. 56).

2.2.4. Cuestionario

Para López & Fachelli (2015) el cuestionario “... es el reflejo de toda la problemática planteada en la investigación, debe traducir en constatación empírica los conceptos que están en el esquema generador que es el modelo de análisis”.

Es un instrumento rígido que busca recoger la información de los entrevistados a partir de la formulación de unas mismas preguntas intentando garantizar una misma situación psicológica estandarizada en la formulación de las preguntas y asegurar después la comparabilidad de las respuestas (p. 17).

Según lo citado expresa, que antes de emplear un cuestionario debe estar previamente estructurado, esto con el fin de compilar información relevante sobre la problemática, hecho que asegura buscar una posible solución. Además, cuenta con preguntas previamente confeccionadas, siendo las mismas preguntas de opción de respuestas múltiples, cerradas y abiertas, pues son más fácil de analizar cada respuesta, y en lo posterior, ser debidamente interpretadas de manera gráfica, debido a su carácter cuantitativo y simplificado.

Asimismo, López y Fachelli (2015) consideran que una pregunta formulada adecuadamente, sin sesgo, es la que no ejerce una influencia en el sentido de la respuesta, no induce a una respuesta inexacta en relación con la información que se busca.

- La formulación de las preguntas abiertas y cerradas dependen de la agrupación de la investigación.



- Un cuestionario no debe basarse en un sinnúmero de preguntas, su objetivo más bien es compilar información relevante a través de preguntas específicas y compendiadas; es decir, concretas y directas.
- Por otro lado, se sugiere que las preguntas sean cerradas, limitadas y categorizadas, pues el objetivo se fundamenta en la estructuración de una problemática previamente indagada, siendo su propósito centrarse los objetivos, para así simplificar la investigación para finalmente buscar o encontrar una solución.

Para Bernal (2010) permite estandarizar y uniformar el proceso de recopilación de datos. Un diseño inadecuado recoge información incompleta, datos imprecisos y, por supuesto, genera información poco confiable

Es un conjunto de preguntas diseñadas para generar los datos necesarios, con el propósito de alcanzar los objetivos del proyecto de investigación. Se trata de un plan formal para recabar información de la unidad de análisis objeto de estudio y centro del problema de investigación (p. 194).

De igual modo, Bernal (2010) menciona que dentro del contexto del cuestionario existen las preguntas abiertas que permiten al encuestado contestar en sus propias palabras, el investigador no limita las opciones de respuesta.

Ofrecen diversas ventajas para el investigador, ya que permite que las personas entrevistadas indiquen sus reacciones generales ante un determinado aspecto o rasgo. Esto conlleva ciertas desventajas: se dificulta el proceso de edición y codificación, así como la interpretación de los patrones de datos y las frecuencias de las respuestas (p. 194).

2.2.5. Encuesta

Según López & Fachelli (2015), la encuesta “... es una técnica considerada para la recolección de datos a través de la interrogación de los sujetos, cuya finalidad es la de obtener información de manera sistemática”.

Es una de las técnicas de investigación que permite compilar datos y se realiza a través de un cuestionario, instrumento de recogida de los datos y la forma protocolaria de realizar las preguntas (cuadro de registro) que se administra a la población o una muestra extensa de ella mediante una entrevista donde es característico el anonimato del sujeto (pp. 8-10).

En otro orden de ideas, Bernal (2010) se fundamenta en un cuestionario o conjunto de preguntas que se preparan con el propósito de obtener información de las personas. “Es una de las técnicas de recolección de información más usadas, a pesar de que cada vez pierde mayor credibilidad por el sesgo de las personas encuestadas” (p. 191).

En una investigación, esta técnica es importante, tiene un enfoque descriptivo que persigue una compilación de datos a través del instrumento cuestionario que consta de cinco preguntas perfectamente elaboradas tanto para el personal como para clientes activos de la empresa Distecom; acción que debe fundamentarse sin modificar su entono ni la problemática existente. Por tanto, los resultados deben presentarse en forma de tabla y gráfico.



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación

William Rafael Raymondi Lomas

Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023





Capítulo 3. Marco metodológico

3.1. Criterios de éxito del negocio

El éxito empresarial está en constante cambio. Para la empresa Distecom, un modelo a seguir es el liderazgo soporte que se consolida en el desarrollo oportuno de soluciones para enfrentar los desafíos presentados. Además, cuenta con una mentalidad visionaria y recursos necesarios para afrontar acontecimientos; enfoque empresarial que se enmarca en la capacidad de innovar, la calidad de las ventas, la eficiencia de la atención al cliente, la formación del personal para trabajar en un entorno competitivo, la comercialización e integración de la comunidad tecnológica indicadores, que aprueban patrones de satisfacción del cliente, mecanismos que dan paso para contar con la recomendación de los clientes y la toma de decisiones basadas en reportes semestrales de los resultados para lograr el cumplimiento de los objetivos trazados.

3.2. Definición de requerimientos del proyecto

3.2.1. Inventario de recursos

Para la presente investigación la empresa Distecom cuenta con los siguientes recursos detallados en la tabla 1.

Tabla 1. Inventario de recursos

1	Servidor HP
2	Procesadores instalados
1	Fuentes de poder 750w
2	20gb en RAM DDR3
1	Base de datos propia de la empresa
2	TB DE discos

Fuente: elaboración propia

3.2.2. Requerimientos funcionales y no funcionales

A continuación, se detallan los requerimientos funcionales y no funcionales de la presente investigación (tabla 2 y 3).

Tabla 2. Historia de usuario 1

Promedio de atención al cliente

Yo como:	Director del Departamento de Atención al cliente
Quiero:	Un dashboard con datos estadísticos de los cliente atendidos dentro de la empresa Distecom
Para:	Llevar un control de rotación de clientes para corroborar si existe la posibilidad de recomendación.

Validación:	Prioridad: 1
<ul style="list-style-type: none">• Cada usuario debe tener una cuenta de usuario y contraseña para poder evaluar la atención recibida.• La actualización de los datos debe ser automatizada.• La granularidad de los datos será de días, semanas y meses.	Estimación: Cada 12 h

Fuente: elaboración propia

Tabla 3. Historia de usuario 2

Atención de clientes en el área de servicios

Yo como:	Director del Departamento de Atención al cliente
Quiero:	Reportes estadísticos de los clientes atendidos en el área de servicios dentro de la empresa Distecom
Para:	Determinar la satisfacción del cliente del servicio brindado por parte del personal de la empresa antes mencionada.
Validación:	Prioridad: 1
<ul style="list-style-type: none">• Todo servicio debe ser categorizado adecuadamente.• La actualización de los datos debe ser automatizada.• La granularidad de los datos será de días, semanas y meses.	Estimación: Cada 12 h

Fuente: elaboración propia

3.2.3. Roles y responsabilidades

Tabla 4. Roles y responsabilidades

Roles	Responsabilidades
Gerente de administración y finanzas	Es el responsable de proponer el plan anual de adquisiciones de bienes, controlar los procesos de abastecimiento de bienes y servicios, siendo en este caso un vínculo directo con proveedores y clientes, a fin de conocer cuál es el estado de la distribución.
Gerente comercial	Se encarga de planificar, dirigir y evaluar las operaciones de los establecimientos comerciales dedicados a la venta al por menor o de ciertos departamentos dentro de esos comercios: manejar al personal y asignarles tareas
Especialista de tecnología	Es el encargado de administrar la base de datos.
Coordinador de gestión de calidad y desarrollo organizacional	Se encarga de los procesos dentro de la compañía, analizar y diseñar planes de mejora que resuelvan los errores detectados
Gerente de operaciones	Su función es planificar, implementar y supervisar el desarrollo óptimo y la ejecución de todas las actividades y procesos diarios
Jefe de atención al usuario y reclamaciones	Encargado de gestionar, tramitar e impulsar la atención ágil y oportuna a los clientes.

Fuente: elaboración propia

3.2.4. Riesgos y contingencias

Tabla 5. Riesgos, probabilidad e impacto

Riesgos	Probabilidades	Impacto
Falta de presupuesto	Bajo	Medio
Inadecuada atención al cliente	Medio	Alto
Complejidad de los servicios	Medio	Medio
Datos incompletos	Bajo	Bajo

Fuente: elaboración propia

Tabla 6. Riesgos y contingencia

Riesgos	Contingencia
Falta de presupuesto	Se requiere un control adecuado de los gastos e ingresos, con el propósito optimizar los recursos que se pretende utilizar en el lapso del proyecto.
Inadecuada atención al cliente	Estimar un adecuado cronograma de servicio con el propósito de corregir la atención del cliente.
Datos incompletos	Implementar una adecuada recolección de datos con el objetivo de adquirir datos de por lo menos dos meses.

Fuente: elaboración propia



3.2.5. Plan del proyecto

Tabla 7. Cronograma de actividades

PLAN OPERATIVO 2021									
Cuadro de mando: Phyton									
¿QUÉ?	¿CÓMO?	PERIODO			¿CON QUÉ?		CRONOGRAMA		
SUBPROYECTO	ACTIVIDADES	Días	DESDE	HASTA	PRESUPUESTO	HUMANOS	E	F	M
Inicio	Levantamiento de la información	3 días	02/01/2021	04/01/2021	\$300,00	Ingeniero en sistema	■		
	Identificar los indicadores	3 días	05/01/2021	07/01/2021			■		
Planificación	Definir plan de proyecto	9 días	08/01/2021	16/01/2021	\$250,00			■	
	Revisión de infraestructura	10 días	17/01/2021	26/01/2020				■	
Desarrollo	Diseño lógico y físico del modelo dimensional	6 días	27/01/2021	02/02/2021	\$2.300,00			■	
	Proceso de extracción	6 días	03/02/2021	09/02/2021				■	
	Transformacion de datos	6 días	10/02/2021	16/02/2021					■
	Carga de datos	6 días	17/02/2021	23/02/2021					■
	Prueba de integridad de datos	6 días	24/02/2021	02/03/2021					■
	Diseño del cuadro de mando en Phyton	6 días	03/03/2021	09/03/2021					■
Fin del proyecto		10 días	10/03/2021	19/03/2021	\$3.800,00				■

Fuente: elaboración propia



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación

William Rafael Raymondi Lomas

Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023



Capítulo 4. Análisis de los resultados

A continuación, se presenta y describe un cuadro comparativo con el propósito de presentar aspectos de un antes y después.

Tabla 7. Cuadro comparativo

	Antes	Ahora
Verificar reportes en tiempo real	Se visualizaban los reportes con 48 horas de retraso	Se visualizan datos en tiempo real, sin importar si hacían reportes semanales, si es fin de semanas o días feriados. Permiten vigilar reportes diarios.
Presentación gráfica de los indicadores	La representación gráfica era poco descifrable.	Brindan información detallada estadísticamente expresado. Es parte de la solución a la problemática planteada
Proceso manual de carga de datos	La carga de los datos al ser manual, lo realizaban los vendedores	El proceso automatizado aporta información prácticamente inmediata. En menos de 24 horas se contará con la visualización de datos relevantes.
Integración de datos	Se monopolizaba diferentes fuentes de datos	Toda la información se carga en la base de datos final, con su debida categorización.

Fuente: elaboración propia

Por otra parte, se ejecutará un análisis de los datos compilados a través de la encuesta realizada al personal de la empresa Distecom, así como a sus clientes activos.

4.1. Esquema de validación

En este punto, la validación de los resultados describe la solución de la propuesta planteada en el presente proyecto, a su vez, detalla al personal involucrado dentro de la siguiente tabla.

Tabla 8. Cargo y tiempo de los servidores

Roles	Años en la institución
Gerente de administración y finanzas	8 meses
Gerente comercial	2 años y 3 meses
Especialista de tecnología	5 años
Coordinador de gestión de calidad y desarrollo organizacional	7 años
Gerente de operaciones	9 años
Jefe de atención al usuario y reclamaciones	9 años

Fuente: elaboración propia

Se determinó una población de 1000 personas con la finalidad de obtener una opinión directa de los clientes de la empresa Distecom, por tanto, se aplicó la siguiente fórmula.



$$N*(Z)^2*p*q$$

$$n= \frac{Z^2*p*q}{e}$$

(e)

N 1000

<i>Z</i>	1,96
<i>e</i>	0,05
<i>p</i>	0,9
<i>q</i>	0,1

Al momento de reemplazar valores se obtiene una muestra de 122 personas, a las cuales se le realizó la encuesta que se considera como la técnica más utilizada para la recolección de información. Consta de seis preguntas abiertas estructuradas, con respuesta de escala, con el fin de conocer la opinión del personal de la empresa, y cinco preguntas dirigidas a los clientes de la empresa antes mencionada, como se visualiza en la siguiente tabla.

Tabla 9. Preguntas de la encuesta

Nro.	Pregunta
1	Desde su perspectiva, ¿qué tan receptiva es la empresa para responder o dar solución a las preguntas o inquietudes sobre los productos? Marque 5 si considera excelente, 4 muy bueno, 3 bueno, 2 regular y 1 malo.
2	Desde su punto de vista profesional, ¿considera que los representantes de atención al cliente de Distecom ofrecen solución ágil y oportuna al cliente?
3	Desde su punto de vista como colaborador de la empresa Distecom, ¿cuál de las siguientes palabras usaría para describir los productos y servicios que ofrece la empresa? Escoja según su criterio personal.
4	En una escala del 1 al 10, ¿qué posibilidad existe que usted como empleado recomiende los servicios de la empresa Distecom, a un amigo, familiar o colega?
5	Responda según su criterio personal, ¿cree usted que la empresa Distecom debe actualizar la base de datos de manera automatizada?

- 6 Por favor ordene (1 al 3) de acuerdo con su interés, ¿considera trascendental adaptar un modelo predictivo que permita identificar la posibilidad de que los clientes de Distecom recomienden los productos y servicios que ofrecen la empresa?
- 7 Desde su punto de vista como cliente, ¿considera estar satisfecho con la forma en que nuestro representante de atención al cliente atendió su requerimiento?
- 8 ¿Considera que nuestro representante de servicio al cliente le ofreció solución ágil y oportuna?
- 9 En una escala del 1 al 10, ¿qué posibilidad hay de que recomiende esta empresa a un amigo, familiar o colega?
- 10 Desde su punto de vista como usuario, ¿cuál de las siguientes palabras usaría para describir los productos y servicios que ofrece la empresa Distecom? Escoja según su criterio personal.
- 11 ¿Qué posibilidades hay de que usted como cliente vuelva a comprar alguno de nuestros productos?

Fuente: elaboración propia

4.2. Descripción de resultados

Pregunta 1. Desde su perspectiva, ¿qué tan receptiva es la empresa para responder o dar solución a las preguntas o inquietudes sobre los productos? Marque 5 si considera excelente, 4 muy bueno, 3 bueno, 2 regular y 1 malo.

Tabla 10. Requerimiento de los clientes

Ítem	Detalle	Frecuencia	Porcentaje
1	1	0	0%
	2	0	0%
	3	2	33%
	4	3	50%



	5	1	17%
Total	6		100%

Fuente: elaboración propia

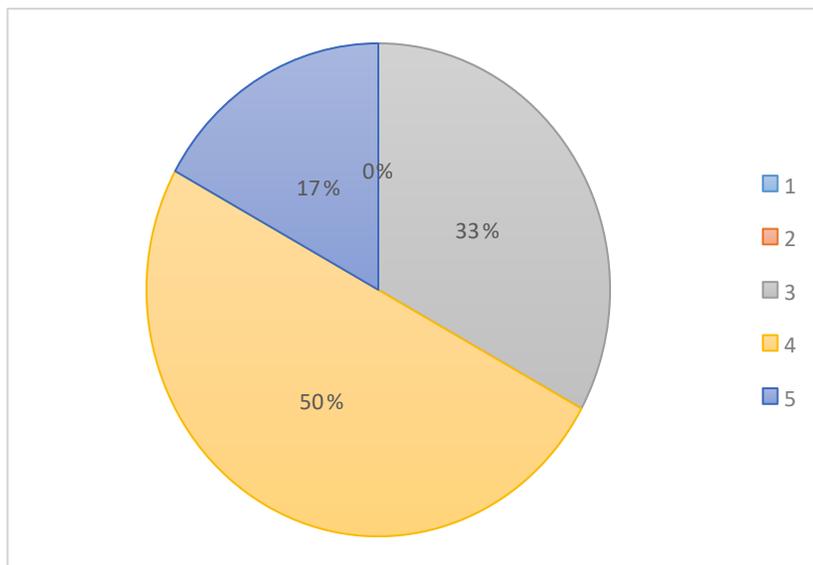


Figura 17. Satisfacción de atención al cliente interno
 Fuente: elaboración propia

De acuerdo con los datos obtenidos mediante la encuesta realizada a los colaboradores de la empresa Distecom (tabla 12, figura 21) se pudo evidenciar que un que el 50% indica que la empresa atiende los requerimientos del cliente; sin embargo, existen algunas falencias dentro del esquema de atención al cliente, pues es de vital importancia formar una base de datos e identificar las necesidades de cada cliente.

Pregunta 2. Desde su punto de vista profesional, ¿considera que nuestros representantes de atención al cliente ofrecen solución ágil y oportuna al cliente?



Tabla 11. Solución ágil y oportuna

Ítem	Detalle	Frecuencia	Porcentaje
2	Si	4	67%
	No	2	33%
	Tal vez	0	0%
Total		6	100%

Fuente: elaboración propia

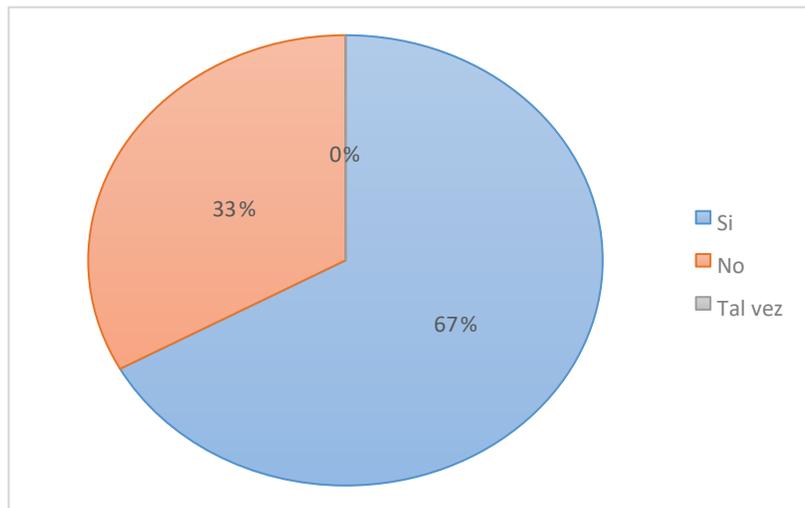


Figura 18. Solución ágil y oportuna

Fuente: elaboración propia

Según los datos recolectados a través de la encuesta (tabal 12, figura 21) efectuada al personal de la empresa antes mencionada, el 67% argumenta que los representantes de atención al cliente ofrecen soluciones ágiles y oportunas a los requerimientos, pero el 33% indica lo contrario.

Pregunta 3. Desde su punto de vista como colaborador de la empresa Distecom, ¿cuál de las siguientes palabras usaría para describir los productos y servicios que ofrece la empresa?

Escoja según su criterio personal.

Tabla 12. Productos y servicios que ofrece la empresa

Ítem	Detalle	Frecuencia	Porcentaje
3	Confianza	2	33%
	Alta calidad	3	50%
	Útil único	1	17%
	Demasiado caro	0	0%
	Poco práctico	0	0%
	No confiable	0	0%
	Mala Calidad	0	0%
Total		6	100%

Fuente: elaboración propia

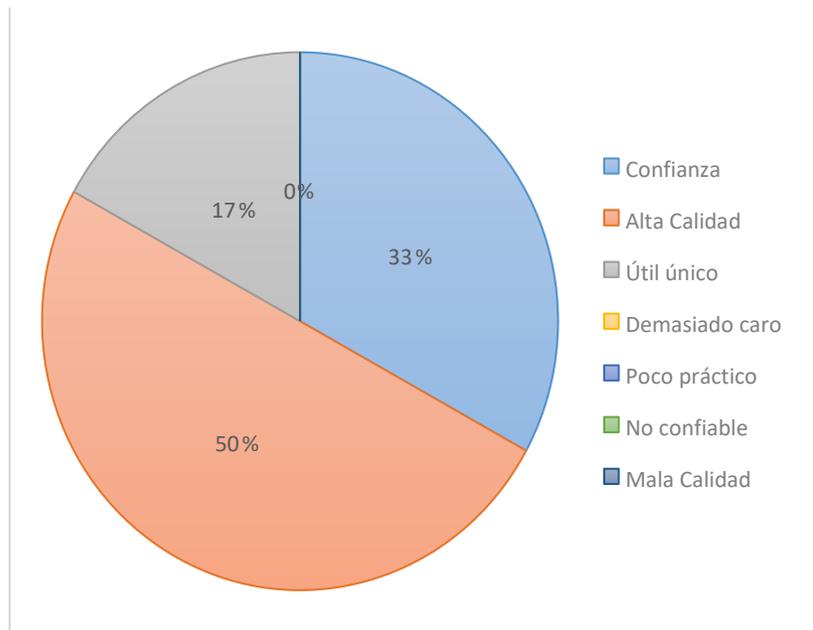


Figura 19. Productos y servicios que ofrece la empresa
 Fuente: elaboración propia

Según los datos derivados de la encuesta se estableció que la mayoría considera que los productos y servicios que ofrece la empresa Distecom (tabla 13, figura 22) son de alta

calidad, un 33% expresó que son de confianza y un 17% cree que son de utilidad única. Estos factores afirman la viabilidad y aceptación de los productos y servicios que ofrece la compañía.

Pregunta 4. En una escala del 1 al 10, ¿qué posibilidad existe que usted como empleado recomiende los servicios de la empresa Distecom, a un amigo, familiar o colega?

Tabla 13 Recomendación de servicios de la empresa

Ítem	Detalle	Frecuencia	Porcentaje
4	De 1-3	0	0%
	De 4-6	1	17%
	De 7-9	4	67%
	10	1	17%
Total		6	100%

Fuente: elaboración propia

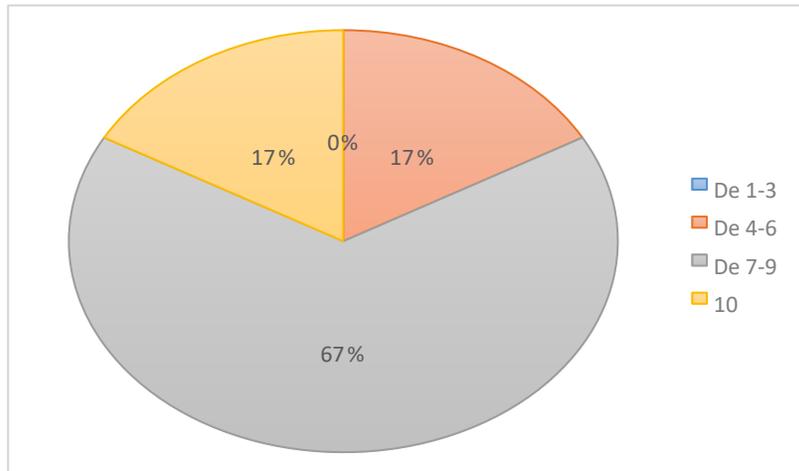


Figura 23. Recomendación de servicios de la empresa
 Fuente: elaboración propia

De acuerdo con los datos logrados se determinó que los encuestados (tabla 14, figura 23) determinaron que en una escala del 1 al 10, un 7-9 consideran que definitivamente la empresa Distecom merece ser referida a amigos, familiares y/o colegas, pues los productos y servicios que oferta la empresa cuentan con calidad, garantía, aceptación convenida.

Pregunta 5. Según su criterio personal, ¿cree usted que la empresa Distecom debe actualizar la base de datos de manera automatizada?

Tabla 14. Actualización de base de datos

Ítem	Detalle	Frecuencia	Porcentaje
5	Extremadamente probable	6	100%
	Muy probable	0	0%
	Poco probable	0	0%
	No tan probable	0	0%
	Nada probable	0	0%
Total		6	100%

Fuente: elaboración propia

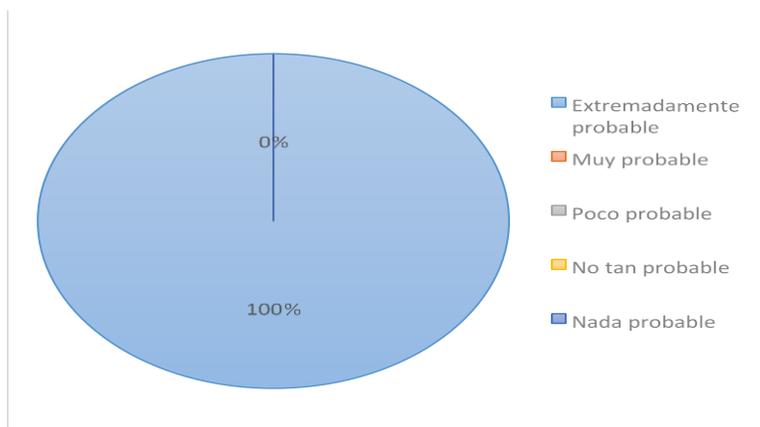


Figura 20. Actualización de base de datos
 Fuente: elaboración propia

De acuerdo con las encuestas procedentes de los colaboradores de la empresa se logró determinar que un 100% considera que es extremadamente probable que la empresa Distecom necesite actualizar su base de datos de manera automatizada (tabla 15, figura 24).

Pregunta 6. Por favor ordene (1 al 3) de acuerdo con su interés, ¿considera trascendental adaptar un modelo predictivo que permita identificar la posibilidad de que los clientes de Distecom recomienden los productos y servicios que ofrecen la empresa?

Tabla 15. Adaptación de modelo predictivo

Ítem	Detalle	Frecuencia	Porcentaje
6	1	0	0%
	2	0	0%
	3	6	100%
Total		6	100%

Fuente: elaboración propia

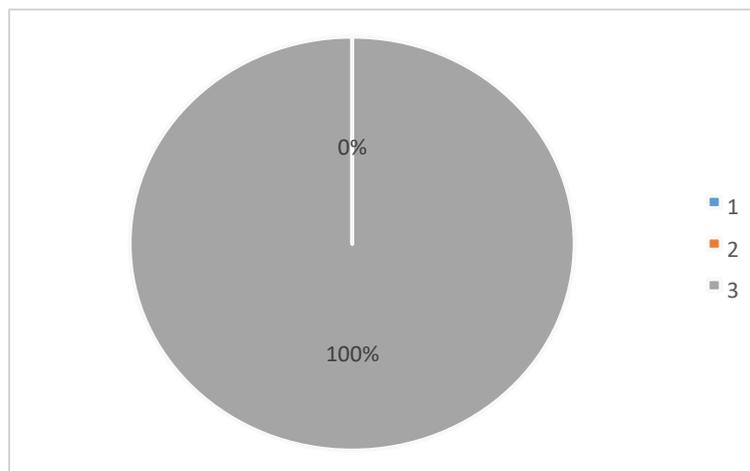


Figura 21. Adaptación de modelo predictivo
Fuente: elaboración propia

De acuerdo con las encuestas derivadas al personal de la empresa (tabla 16, figura 25) se observó que el 100% considera que es trascendental adaptar un modelo predictivo que le permita identificar la probabilidad de que los clientes recomienden tantos los productos como el servicio que la empresa Distecom ofrece.

Encuesta dirigida a los clientes

Pregunta 1. Desde su punto de vista como cliente, ¿considera estar satisfecho con la forma en que nuestro representante de atención al cliente atendió su requerimiento?

Tabla 16. Atención al cliente

Ítem	Detalle	Frecuencia	Porcentaje
7	Si	82	67%
	No	40	33%
	Tal vez	0	0%
Total		122	100%

Fuente: elaboración propia

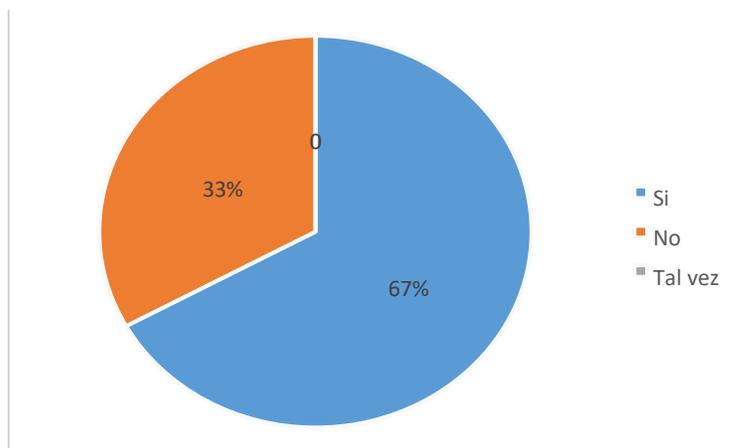


Figura 22. Atención al cliente

Fuente: elaboración propia



Según los datos derivados a través de la encuesta aplicada a los clientes activos de la empresa Distecom (tabla 17, figura 26) se pudo demostrar que un que el 67% indica que la empresa atiende los requerimientos del cliente; sin embargo, existen algunas falencias dentro del esquema de atención al cliente, pues que es de vital importancia en formar una base de datos e identificar las necesidades de cada cliente.

Pregunta 2. ¿Considera que nuestro representante de servicio al cliente le ofreció solución ágil y oportuna?

Tabla 17. Ofrece solución ágil y oportuna

Ítem	Detalle	Frecuencia	Porcentaje
8	Si	95	78%
	No	27	22%
	Tal vez	0	0%
Total		122	100%

Fuente: elaboración propia

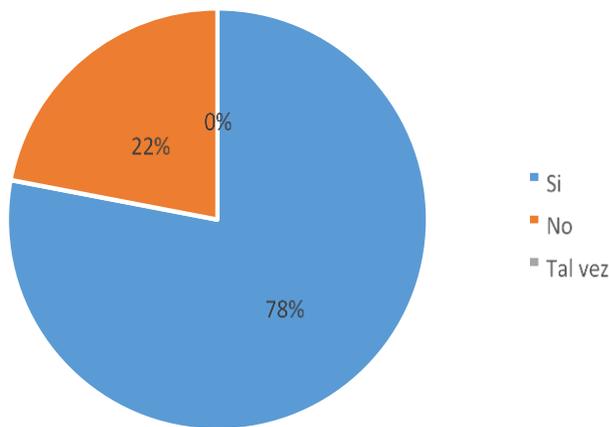


Figura 23. Ofrece solución ágil y oportuna

Fuente: elaboración propia



Mediante los datos recabados de la encuesta (tabla 18, figura 27) formalizada a los clientes reales de la empresa antes mencionada, el 78% refuta que los representantes de atención al cliente ofrecen soluciones ágiles y oportunas a los requerimientos; no obstante, el 22% indica lo contrario, ya que es importante el cumplimiento de las metas establecidas.

Pregunta 3. En una escala del 1 al 10, ¿qué posibilidad hay de que recomiende esta empresa a un amigo, familiar o colega?

Tabla 18. Recomendación de servicios de la empresa

Ítem	Detalle	Frecuencia	Porcentaje
9	De 1-3	0	0%
	De 4-6	0	0%
	De 7-9	34	28%
	10	88	72%
Total		122	100%

Fuente: elaboración propia

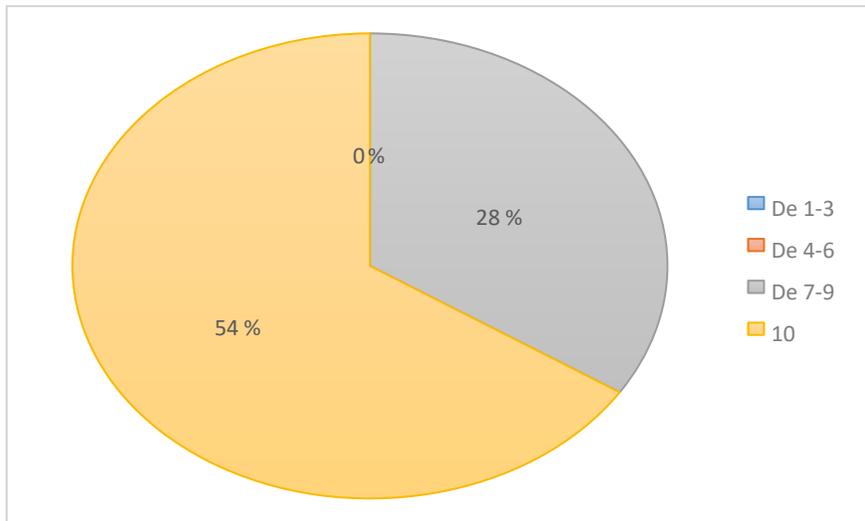


Figura 24. Recomendación de servicios de la empresa
 Fuente: elaboración propia



De acuerdo con los datos procedidos de las encuestas se estableció que la mayoría indica que sí recomendarían los servicios que ofrece la empresa Distecom (tabla 19, figura 28), ya que consideran que los productos y servicios de la empresa son de alta calidad. Por su parte, el 28% que expreso que sí recomendarían los servicios y productos de la empresa antes mencionada, sin embargo, indican que existen algunas faltas en la atención del cliente.

Pregunta 4. Desde su punto de vista como usuario, ¿cuál de las siguientes palabras usaría para describir los productos y servicios que ofrece la empresa Distecom? Escoja según su criterio personal.

Tabla 19. Productos y servicios que ofrece

Ítem	Detalle	Frecuencia	Porcentaje
10	Confianza	23	19%
	Alta Calidad	80	66%
	Útil único	19	15%
	Demasiado caro	0	0%
	Poco práctico	0	0%
	No confiable	0	0%
	Mala Calidad	0	0%
Total		122	100%

Fuente: elaboración propia

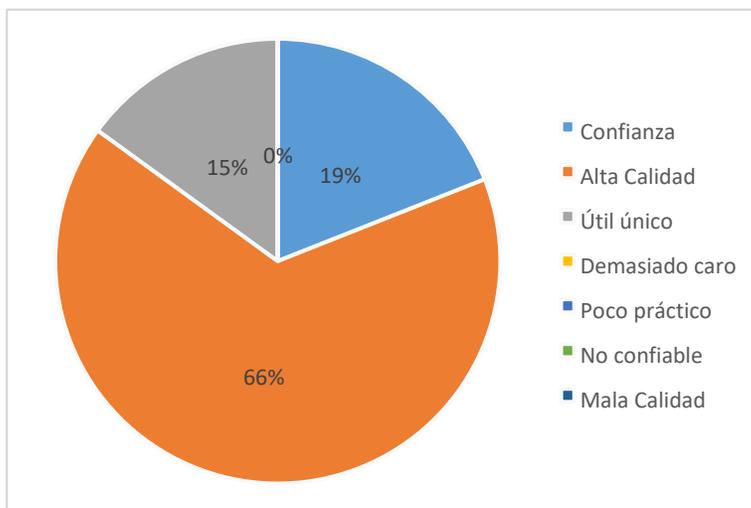


Figura 25. Productos y servicios que ofrece

Fuente: elaboración propia

Como se pudo determinar a través de los datos obtenidos a partir de la encuesta ejecutada a los clientes (tabla 20, figura 29) se instauró que la mayoría considera que los productos y servicios que ofrece la empresa Distecom son de alta calidad; un 19% expresó que son de confianza; y un 15% cree que son de utilidad única. Estos factores afirman la viabilidad y aceptación de los productos y servicios que ofrece la compañía.

Pregunta 5. ¿Qué posibilidades hay de que usted como cliente vuelva a comprar alguno de nuestros productos?

Tabla 20. Posibilidad de compra

Ítem	Detalle	Frecuencia	Porcentaje
11	extremadamente probable	122	100%
	Muy probable	0	0%
	poco probable		0%

No tan probable	0	0%
Nada probable	0	0%
<hr/>		
Total	122	100%

Fuente: elaboración propia

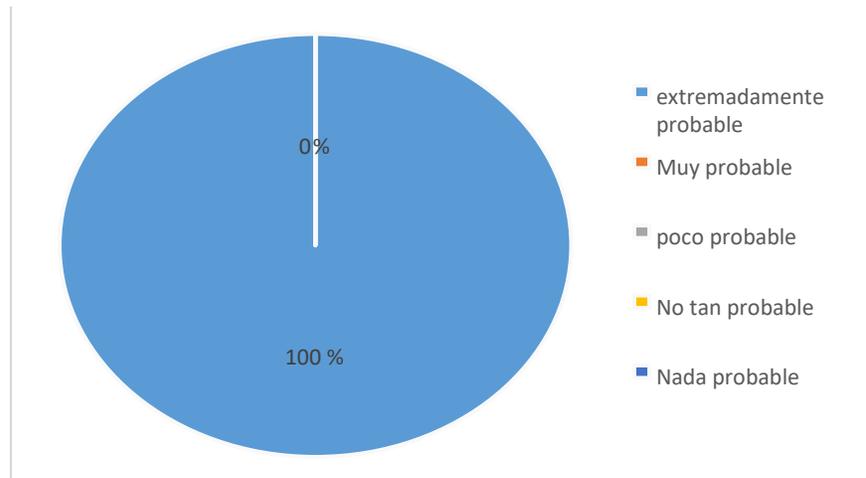


Figura 26. Posibilidad de compra

Fuente: elaboración propia

De unión con las búsquedas procedentes de los clientes de la empresa (tabla 21, figura 30) se logró determinar que un 100% consideran que es extremadamente probable que vuelvan a comprar algunos de los productos de la empresa Distecom, pues consideran que los servicios y productos son de alta calidad, confiables, y buen precio.

Análisis de los resultados

La encuesta aplicada a la muestra de clientes y colaboradores se diseñó indagando sobre variables tales como la actitud de la empresa para ofrecer soluciones y responder inquietudes sobre la oferta de bienes y servicios. También sobre el nivel del servicio al cliente e intenciones.



Sobre la actitud de la empresa para ofrecer soluciones y responder a inquietudes de los clientes, la puntuación fue desde el nivel medio hasta el máximo, siendo el cuarto nivel donde se concentró el 50% de las respuestas, esto se complementa con un 67% que indica que se ofrecen respuestas ágiles y oportunas y los conceptos con los que se relaciona la oferta de la empresa es confianza y alta calidad.

Producto de lo anterior, es que la mayoría indicó que en una probabilidad entre siete y diez recomendarían la oferta de la empresa ante terceros. Esto es una clara intención de referencias, como resultado de un cliente satisfecho.



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación
William Rafael Raymondi Lomas
Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023





Capítulo 5. Presentación de soluciones tecnológicas, aplicaciones técnicas

5.1. Análisis de la situación actual

El desarrollo de este libro está enfocado a trabajar con la empresa Distecom, la cual tiene la necesidad de identificar la posibilidad de que los clientes refieran la empresa. De ahí que, se requiere analizar la satisfacción del cliente, razón por la que se decidió ejecutar una encuesta por medio de campañas y sondeo puerta a puerta. El propósito es recabar datos referentes e importantes tales como atención al cliente, calidad del servicio, entre otros.

Actualmente, la empresa cuenta con un sistema totalmente manual que no permite visualizar datos estadísticos exactos que sirvan para evaluar la tendencia futura de las ventas, clientes, productos disponibles, entre otros, esto con el afán de captar clientes potenciales y reforzar la fidelización de los clientes actuales. Por ello, existe la necesidad de actualizar y generar información que brinde facilidad y reconocimiento de satisfacciones de las necesidades de los clientes con los productos y servicios que ofrece la empresa.

5.1.1. Organigrama de la institución



Figura 27. Organigrama de la empresa

5.1.2. Objetivos de la institución

Objetivos de calidad

- Coordinar y atender los requerimientos de los clientes de manera oportuna.
- Disponibilidad de datos actualizados.
- Cumplir el plan de entrenamiento
- Cumplir con el número de clientes en atención prioritaria
- Satisfacer las necesidades de los usuarios
- Conocer el porcentaje de recomendación de los clientes reales.

Objetivos estratégicos

Estos objetivos atienden de manera exclusiva las metas que propone la empresa en un tiempo determinado con el fin de lograr su visión empresarial. A continuación, se detallan los siguientes.

- Se estima alcanzar el 95% de referencia de los clientes, en un plazo de máximo un año.
- Reforzar la lealtad de los clientes activos.
- Atraer referidos nuevos.
- Administrar eficientemente los costos operacionales.
- Mejorar la eficiencia productiva.
- Ofrecer una amplia gama de productos.
- Incrementar los ingresos por servicios y ventas de productos.
- Elevar la variedad de productos en un plazo de cinco años.
- Fortalecer el trabajo en equipo.

5.2. Descripción de las fuentes de datos

Los datos obtenidos para este proyecto provienen de sondeos tales como encuestas personales o vía web, entre otras formas de conseguir data.

5.2.1. Exploración de los datos

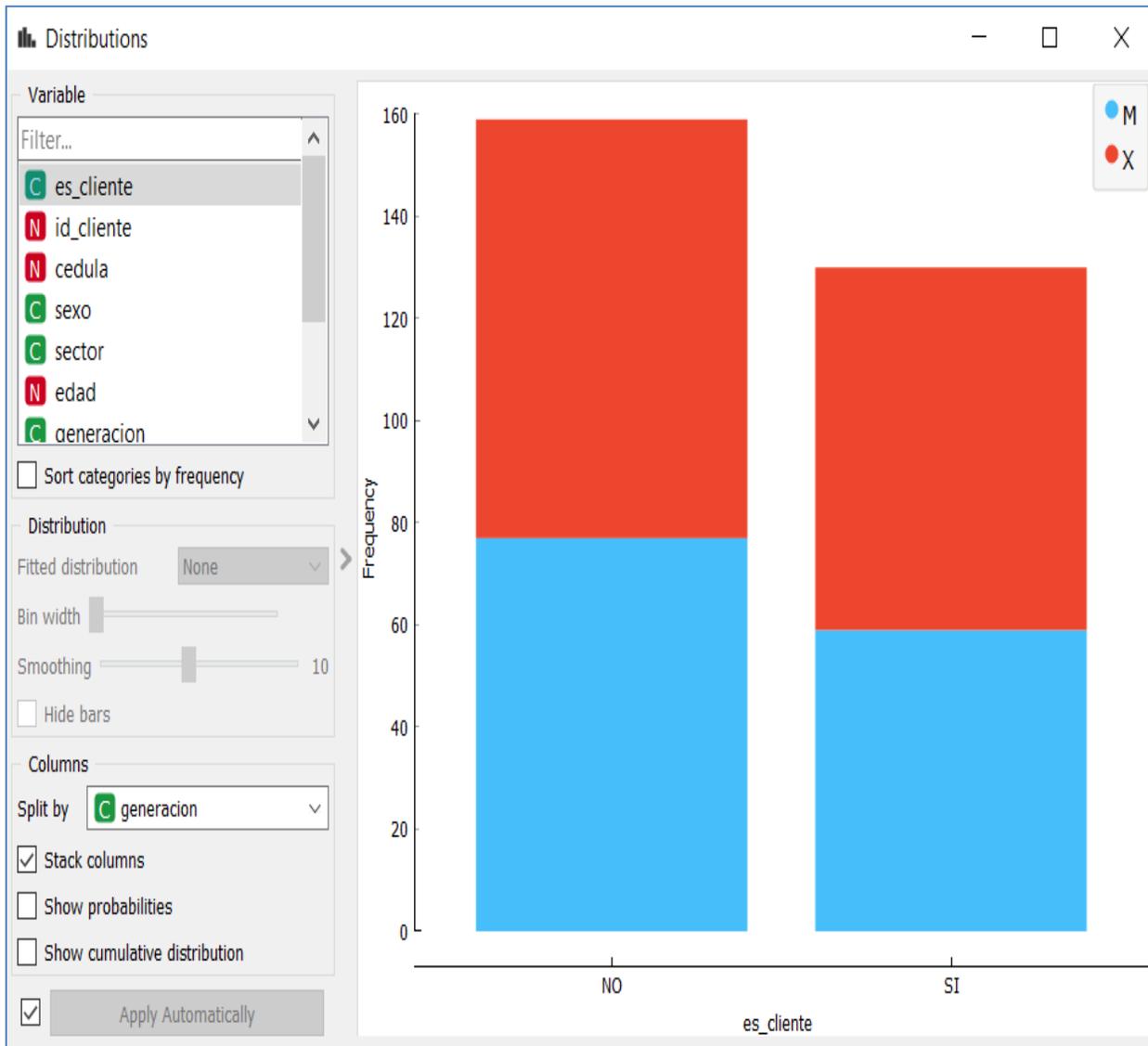


Figura 28. Generación
Fuente: elaboración propia



Distributions

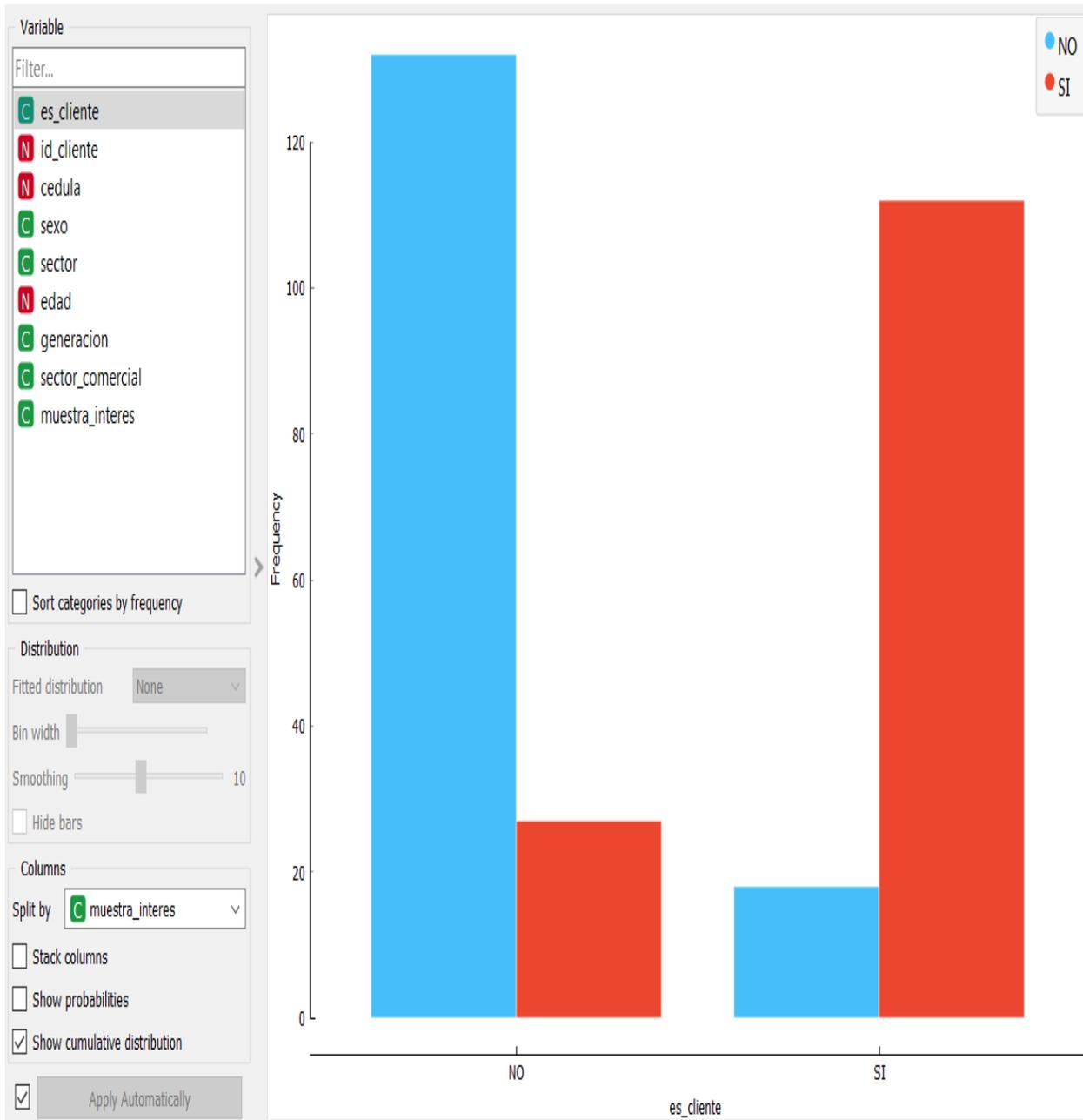


Figura 29. Muestra Interés

Fuente: elaboración propia



Distributions

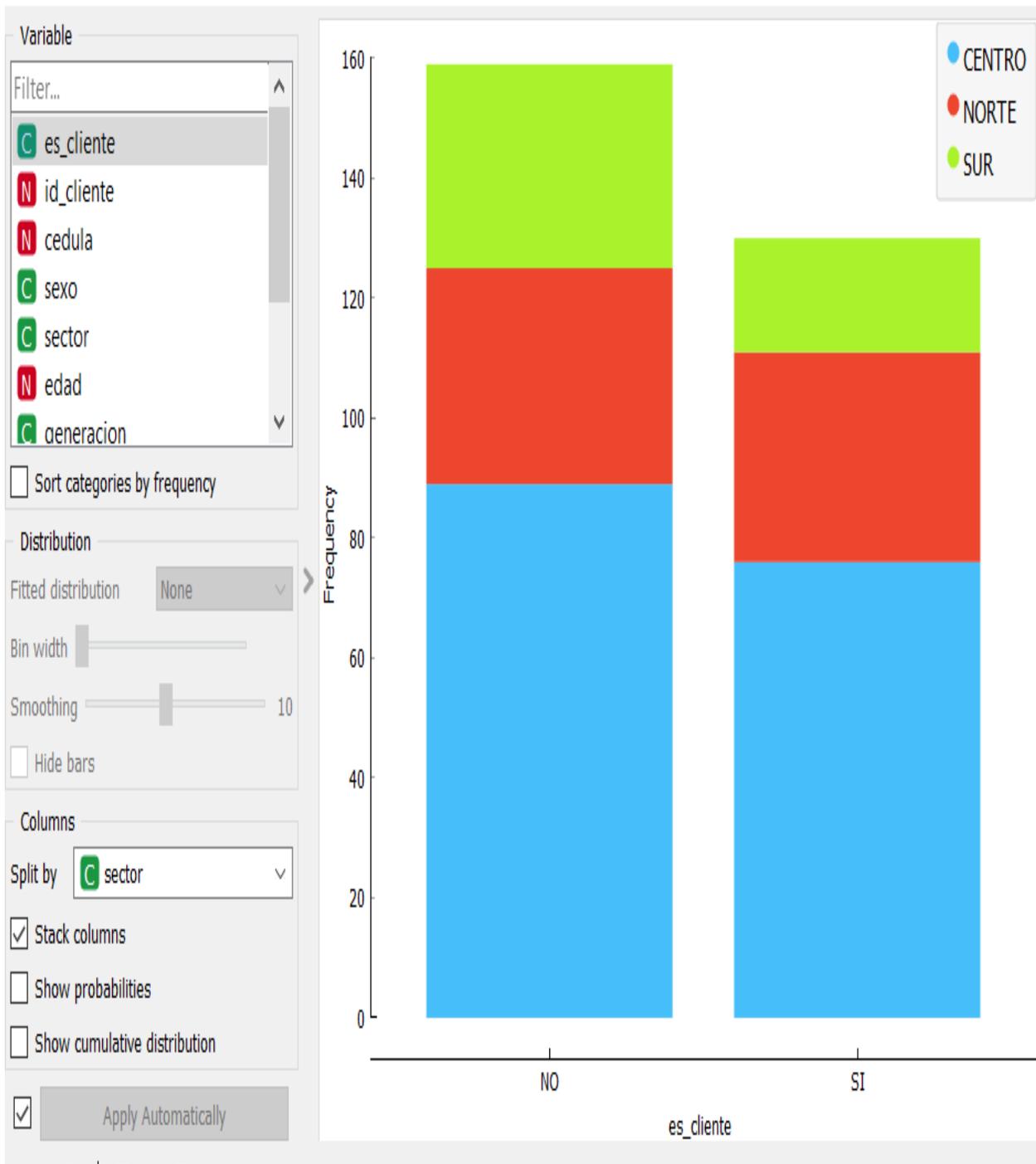


Figura 30. Sector
Fuente: elaboración propia

Distributions

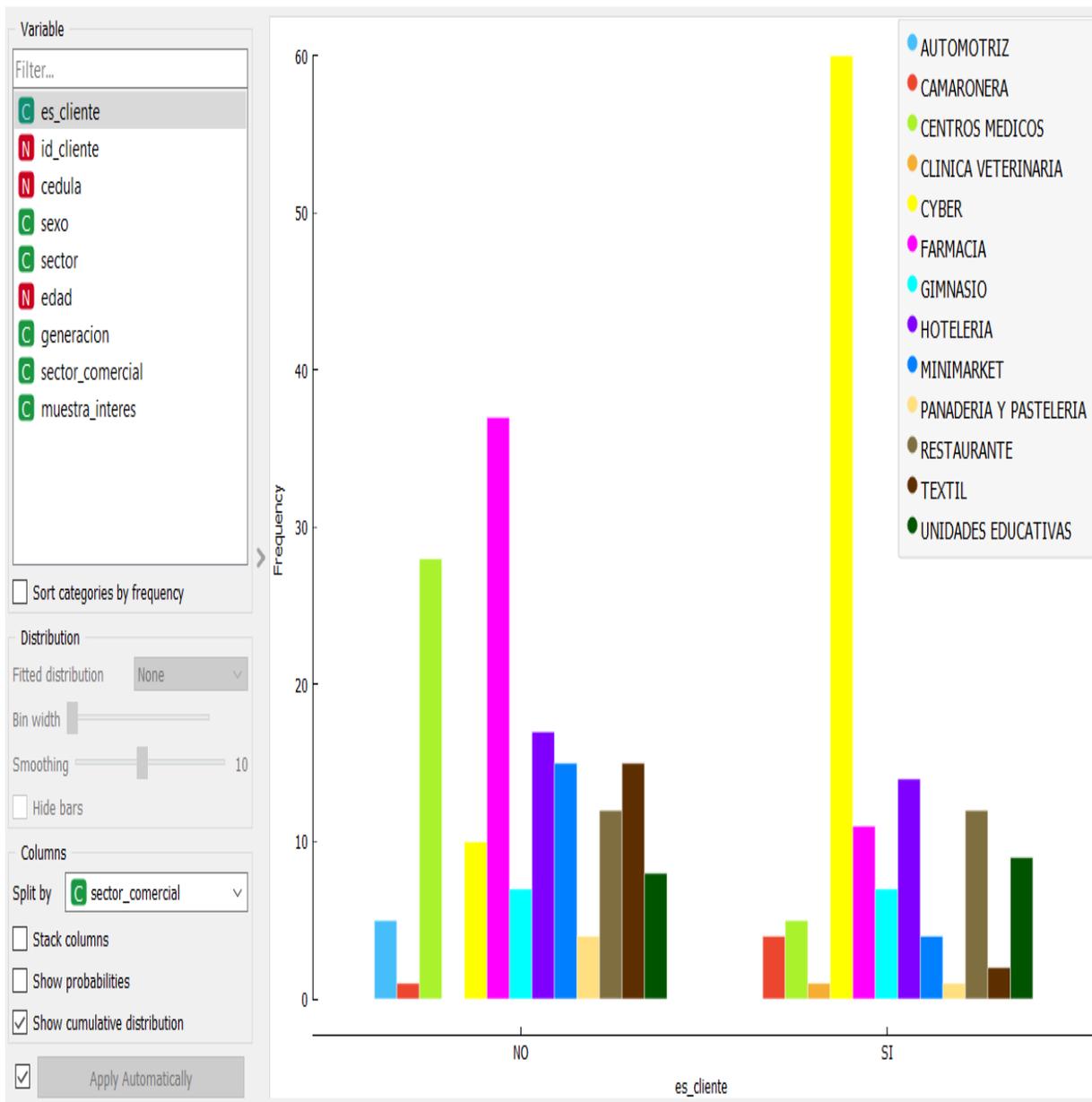


Figura 31. Sector Comercial

Fuente: elaboración propia

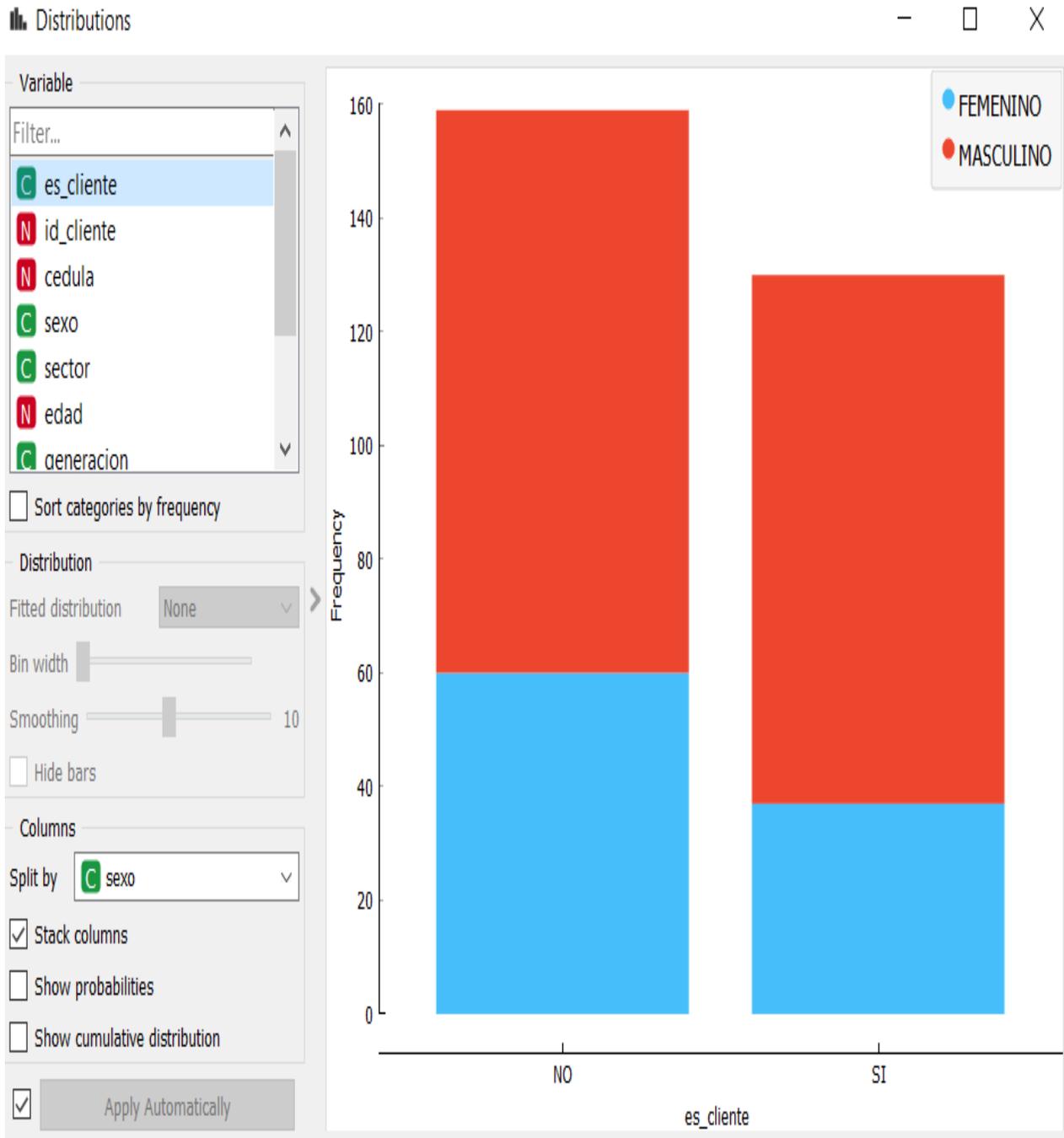


Figura 32. Sexo
Fuente: elaboración propia

Muestra el directorio donde se trabaja.

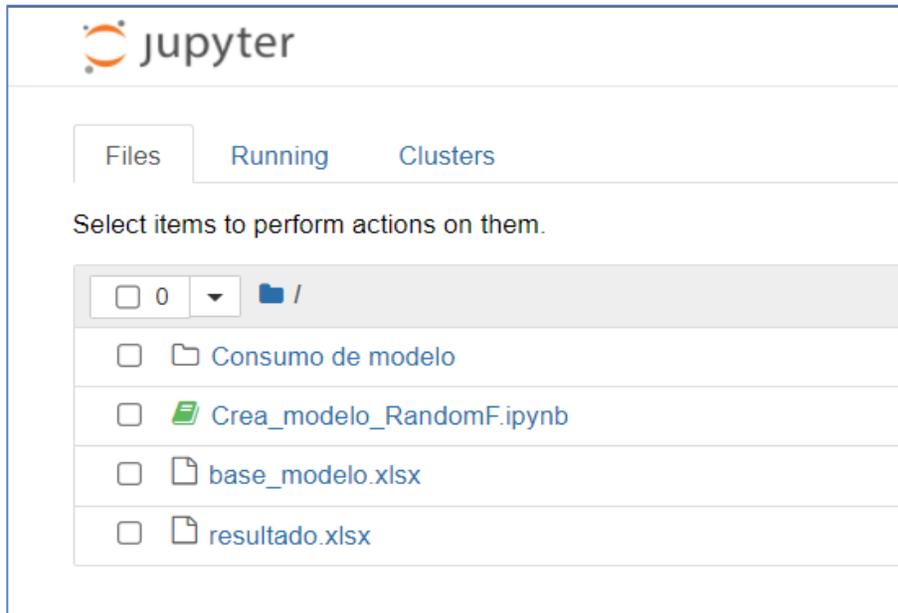


Figura 33. Directorio crea modelo
Fuente: elaboración propia

In7: Se importan las librerías necesarias para el modelo.

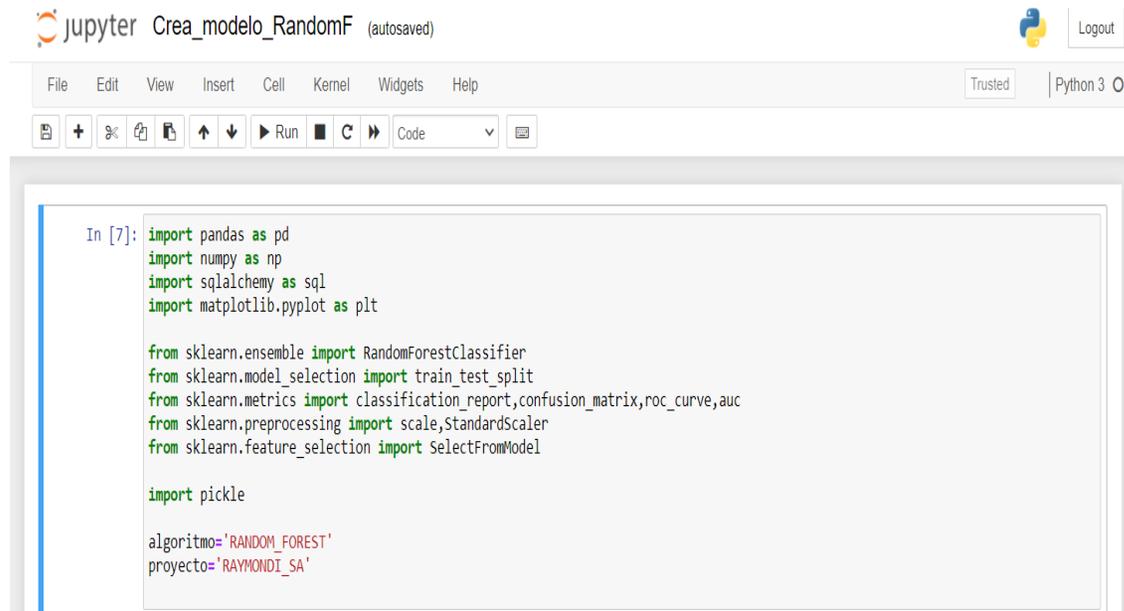


Figura 34. crea_modelo1
Fuente: elaboración propia

In8: dentro de la variable data se importa el Excel que contiene la información.

```
In [8]: data=pd.read_excel('./base_modelo.xlsx','Hoja1')

In [9]: data.columns

Out[9]: Index(['id_cliente', 'cedula', 'nombres_apellidos', 'sexo', 'sector', 'edad',
              'generacion', 'sector_comercial', 'muestra_interes', 'es_cliente'],
              dtype='object')
```

```
In [10]: y=data.es_cliente
x_cols=list(['sexo','sector','edad','generacion','sector_comercial','muestra_interes'])
x=data[x_cols]
y[y==2]=0
x.head()

<ipython-input-10-b46c6e81e12b>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
y[y==2]=0
```

```
Out[10]:
```

	sexo	sector	edad	generacion	sector_comercial	muestra_interes
0	1	1	43	1	1	1
1	2	1	39	2	2	1
2	2	1	49	1	2	2
3	1	1	32	2	3	1
4	1	1	50	1	3	2

Figura 35. Crea_modelo2
Fuente: elaboración propia

Mostramos las variables con su cantidad de registros no vacíos y el tipo de dato

```
In [11]: x.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 289 entries, 0 to 288
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   sexo                  289 non-null    int64
1   sector                289 non-null    int64
2   edad                  289 non-null    int64
3   generacion            289 non-null    int64
4   sector_comercial      289 non-null    int64
5   muestra_interes       289 non-null    int64
dtypes: int64(6)
memory usage: 13.7 KB
```

```
In [12]: x.describe()

Out[12]:
```

	sexo	sector	edad	generacion	sector_comercial	muestra_interes
count	289.000000	289.000000	289.000000	289.000000	289.000000	289.000000
mean	1.335640	1.937716	42.051903	1.470588	4.245675	1.519031
std	0.473033	0.653194	6.665683	0.500000	3.113343	0.500504
min	1.000000	1.000000	28.000000	1.000000	1.000000	1.000000
25%	1.000000	2.000000	37.000000	1.000000	2.000000	1.000000
50%	1.000000	2.000000	42.000000	1.000000	3.000000	2.000000
75%	2.000000	2.000000	48.000000	2.000000	6.000000	2.000000
max	2.000000	3.000000	56.000000	2.000000	13.000000	2.000000

Figura 36. crea_modelo3
Fuente: elaboración propia

Se cuenta la cantidad de registros para saber si es o no cliente en $y=es_cliente$

In14: Se busca no tener valores nulos, esto es ningún registro sin llenar.

```
In [13]: #Verificar target 0 no es cliente, 1 es cliente de la empresa
y.groupby(y).count()

Out[13]: es_cliente
0      159
1      130
Name: es_cliente, dtype: int64

In [14]: #Verificar que no tenga valores NaN (NULL)
x.muestra_interes[x.muestra_interes.isna()]=0
x.describe()

C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\series.py:1020: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
self._where(~key, value, inplace=True)

Out[14]:
```

	sexo	sector	edad	generacion	sector_comercial	muestra_interes
count	289.000000	289.000000	289.000000	289.000000	289.000000	289.000000
mean	1.335640	1.937716	42.051903	1.470588	4.245675	1.519031
std	0.473033	0.653194	6.665683	0.500000	3.113343	0.500504
min	1.000000	1.000000	28.000000	1.000000	1.000000	1.000000
25%	1.000000	2.000000	37.000000	1.000000	2.000000	1.000000
50%	1.000000	2.000000	42.000000	1.000000	3.000000	2.000000
75%	2.000000	2.000000	48.000000	2.000000	6.000000	2.000000
max	2.000000	3.000000	56.000000	2.000000	13.000000	2.000000

Figura 37. Crea_modelo4
Fuente: elaboración propia

In18: Se muestra la cantidad de registros y cantidad de columnas .

In20: Se define el modelo de prediccion random forest Clasificador y definimos los parametros para la creacion del modelo.

```
In [18]: x.shape

Out[18]: (289, 6)

In [20]: clf=RandomForestClassifier()
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=9)
clf.fit(x_train,y_train)

Out[20]: RandomForestClassifier()
```

Figura 38 Crea_modelo5
Fuente: elaboración propia

In22: el valor de y lo dará clf (random forest), usando la función predict en la prueba de x.

Se realiza una matriz de confusión con el valor real de y contra y predecida.

##score AUC, devuelve la precisión del modelo en nuestro caso es 86%.

```
In [22]: #Algoritmo RandomForest
y_pred=clf.predict(x_test)

##Matriz de confusion
print(confusion_matrix(y_test,y_pred))

##Reporte de clasificacion
print(classification_report(y_test,y_pred))

y_pred_prob=clf.predict_proba(x_test)[:,-1]
fpr, tpr, thresholds=roc_curve(y_test,y_pred_prob)
plt.plot([0,1],[0,1], 'k--')
plt.plot(fpr,tpr,label='Random Forest Classifier')
plt.xlabel('False positive rate')
plt.ylabel('True positive rate')
plt.title('Random Forest Classifier')

## Curva ROC
plt.show()

## Score AUC
print(roc_auc_score(y_test,y_pred_prob))
```

	precision	recall	f1-score	support
0	0.82	0.85	0.84	48
1	0.81	0.77	0.79	39
accuracy			0.82	87
macro avg	0.82	0.81	0.81	87
weighted avg	0.82	0.82	0.82	87

Figura 39 crea_modelo6
Fuente: elaboración propia

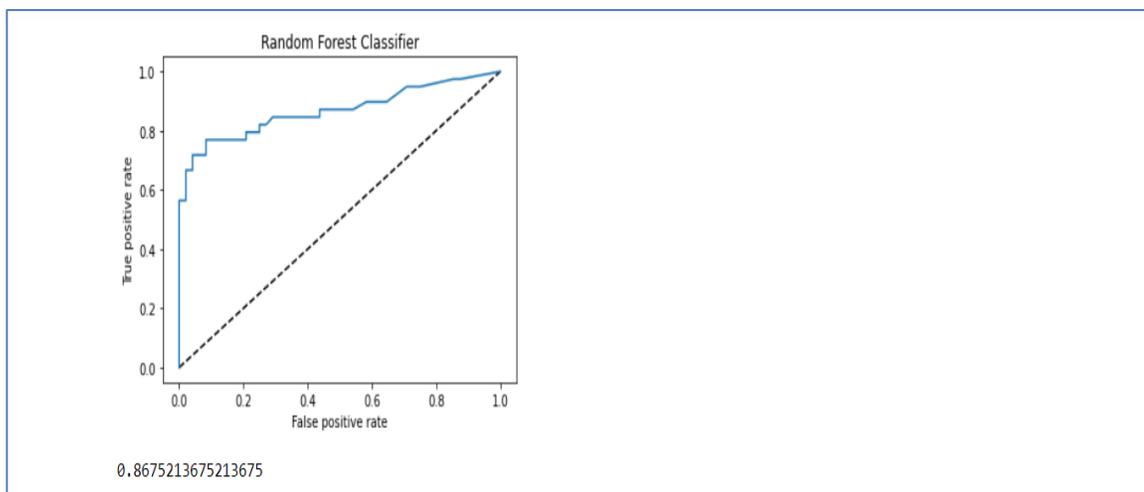


Figura 40 Crea_modelo7
Fuente: elaboración propia

In25: Mejora la precisión del modelo haciendo que entrene más con otros parámetros de entrenamiento.

```
In [25]: from sklearn.model_selection import GridSearchCV

param_grid={
    'bootstrap':[True],
    'max_depth':[80,90,100,110],
    'max_features':[2,3,4],
    'min_samples_leaf':[3,4,5],
    'min_samples_split':[8,9,10,11],
    'n_estimators':np.linspace(50,200,3).astype(int),
    'criterion':['gini']
}

rf=RandomForestClassifier()

grid_search2=GridSearchCV(estimator=rf,param_grid=param_grid,cv=5,n_jobs=3, verbose=2)
grid_search2.fit(x,y)

Fitting 5 folds for each of 432 candidates, totalling 2160 fits

[Parallel(n_jobs=3)]: Using backend LokyBackend with 3 concurrent workers.
[Parallel(n_jobs=3)]: Done 35 tasks      | elapsed: 2.6s
[Parallel(n_jobs=3)]: Done 156 tasks   | elapsed: 7.8s
[Parallel(n_jobs=3)]: Done 359 tasks   | elapsed: 16.6s
[Parallel(n_jobs=3)]: Done 642 tasks   | elapsed: 28.9s
[Parallel(n_jobs=3)]: Done 1007 tasks  | elapsed: 44.2s
[Parallel(n_jobs=3)]: Done 1452 tasks  | elapsed: 1.0min
[Parallel(n_jobs=3)]: Done 1979 tasks  | elapsed: 1.4min
[Parallel(n_jobs=3)]: Done 2160 out of 2160 | elapsed: 1.5min finished

Out[25]: GridSearchCV(cv=5, estimator=RandomForestClassifier(), n_jobs=3,
                  param_grid={'bootstrap': [True], 'criterion': ['gini'],
                              'max_depth': [80, 90, 100, 110],
                              'max_features': [2, 3, 4],
                              'min_samples_leaf': [3, 4, 5],
                              'min_samples_split': [8, 9, 10, 11],
                              'n_estimators': array([ 50, 125, 200])},
                  verbose=2)
```

Figura 41. Crea_modelo8
Fuente: elaboración propia

In26: Se repite la elaboración de una matriz de confusión y nos devuelve una precisión en el modelo del 94%.

```
In [26]: y_pred=grid_search2.predict(x_test)

print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))

y_pred_prob=grid_search2.predict_proba(x_test)[:,:1]
fpr,tpr,treshholds=roc_curve(y_test,y_pred_prob)
plt.plot([0,1],[0,1],k--')
plt.plot(fpr,tpr,label='Random Forest Classifier')
plt.xlabel('False positive rate')
plt.ylabel('True positive rate')
plt.title('Random Forest Classifier')
plt.show()
print(auc(fpr,tpr))

[[45  3]
 [ 9 30]]

              precision    recall  f1-score   support

     0       0.83      0.94      0.88         48
     1       0.91      0.77      0.83         39

 accuracy          0.87
 macro avg          0.87
 weighted avg       0.87
```

Figura 42. Crea_modelo9
Fuente: elaboración propia

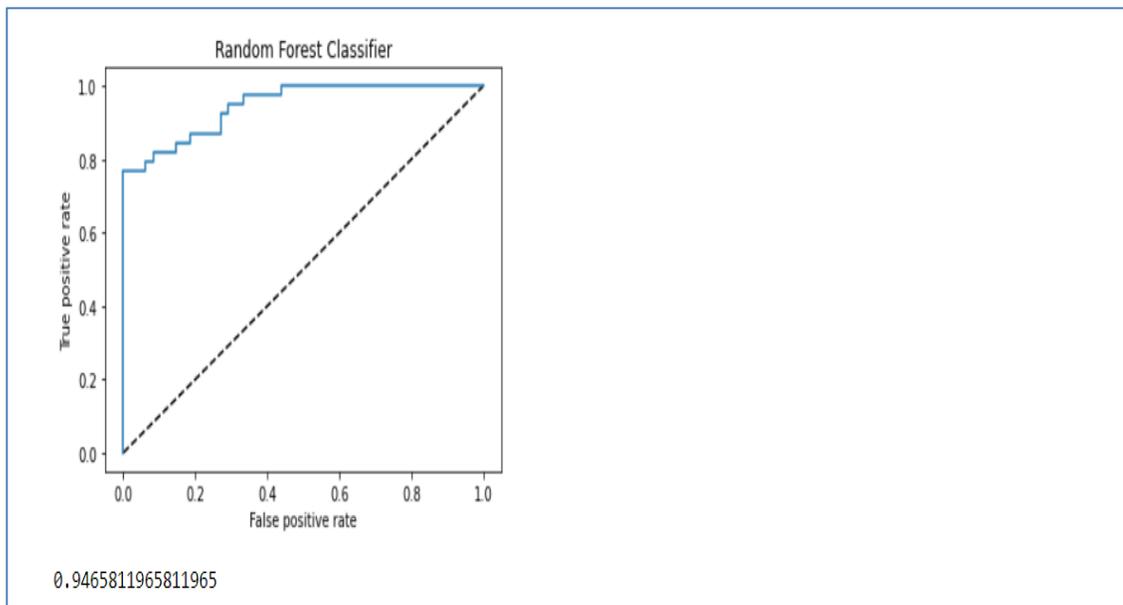


Figura 43. Crea_modelo10
 Fuente: elaboración propia

```
In [27]: y_pred_total=grid_search2.predict(x)

print(confusion_matrix(y,y_pred_total))
print(classification_report(y,y_pred_total))

y_pred_prob_t=grid_search2.predict_proba(x)[:,:1]

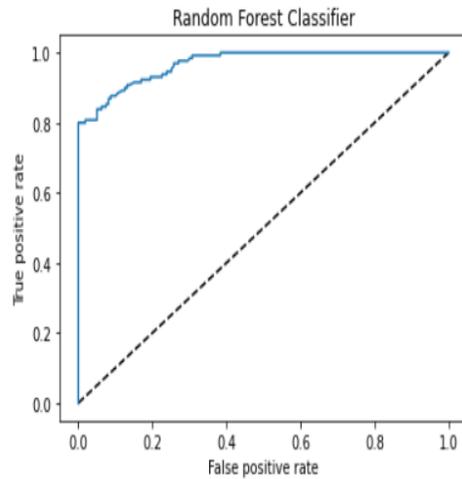
fpr,tpr,tresholds=roc_curve(y,y_pred_prob_t)
plt.plot([0,1],[0,1], 'k--')
plt.plot(fpr,tpr,label='Random Forest Classifier')
plt.xlabel('False positive rate')
plt.ylabel('True positive rate')
plt.title('Random Forest Classifier')
plt.show()
print(auc(fpr,tpr))

[[151  8]
 [ 22 108]]
      precision    recall  f1-score   support

   0       0.87     0.95     0.91     159
   1       0.93     0.83     0.88     130

 accuracy          0.90
 macro avg          0.90     0.89     0.89     289
 weighted avg          0.90     0.90     0.90     289
```

Figura 44. Crea_modelo11
 Fuente: elaboración propia



0.968577648766328

Figura 45 Crea_modelo12

Fuente: elaboración propia

In28: con la data que es el archivo de Excel (base_modelo.xls), devuelve otro archivo Excel (resultado.xls) con el campo si es o no probable que sea cliente.

```
In [28]: y_0=grid_search2.predict_proba(x)[:,:0]
y_1=grid_search2.predict_proba(x)[:,:1]
```

```
data['prb_acept']=y_1
```

```
data.to_excel('resultado.xlsx')
```

```
In [29]: nombre_archivo_modelo=algoritmo+'_'+proyecto+'.sav'
pickle.dump(grid_search2,open(nombre_archivo_modelo,'wb'))
```

Figura 46. Crea_modelo13

Fuente: elaboración propia

Este es el directorio donde están los archivos necesarios para consumir el modelo random forest creado previamente.



Files

Running

Clusters

Select items to perform actions on them.



Figura 47. Directorio consume modelo
Fuente: elaboración propia

Se importan las librerías necesarias para el modelo, se define el nombre del archivo modelo con el nombre del modelo entrenado, luego se cargan y se hace el llamado al modelo. En una variable data importamos el Excel (data) a predecir.

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

import pickle

In [5]: nombre_archivo_modelo='./RANDOM_FOREST_RAYMONDI_SA.sav'
model=pickle.load(open(nombre_archivo_modelo,'rb'))

data=pd.read_excel('./base_consumo_modelo.xlsx','Hoja1')
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id_cliente            100 non-null    int64
1   cedula                100 non-null    int64
2   nombres_apellidos     100 non-null    object
3   sexo                  100 non-null    int64
4   sector                100 non-null    int64
5   edad                  100 non-null    int64
6   generacion            100 non-null    int64
7   sector_comercial      100 non-null    int64
8   muestra_interes       100 non-null    int64
dtypes: int64(8), object(1)
memory usage: 7.2+ KB
```

Figura 48. Consume_modelo01
Fuente: elaboración propia

Mostramos las columnas que tiene la variable data

In8: Se definen las variables independientes

```
In [6]: data.columns

Out[6]: Index(['id_cliente', 'cedula', 'nombres_apellidos', 'sexo', 'sector', 'edad',
              'generacion', 'sector_comercial', 'muestra_interes'],
              dtype='object')

In [8]: x_cols=['sexo','sector','edad','generacion','sector_comercial','muestra_interes']

x=data[x_cols]

x.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   sexo                  100 non-null    int64
1   sector                100 non-null    int64
2   edad                  100 non-null    int64
3   generacion            100 non-null    int64
4   sector_comercial      100 non-null    int64
5   muestra_interes       100 non-null    int64
dtypes: int64(6)
memory usage: 4.8 KB
```

Figura 49. Consume_modelo02
Fuente: elaboración propia

Se predice utilizando el modelo entrenado.

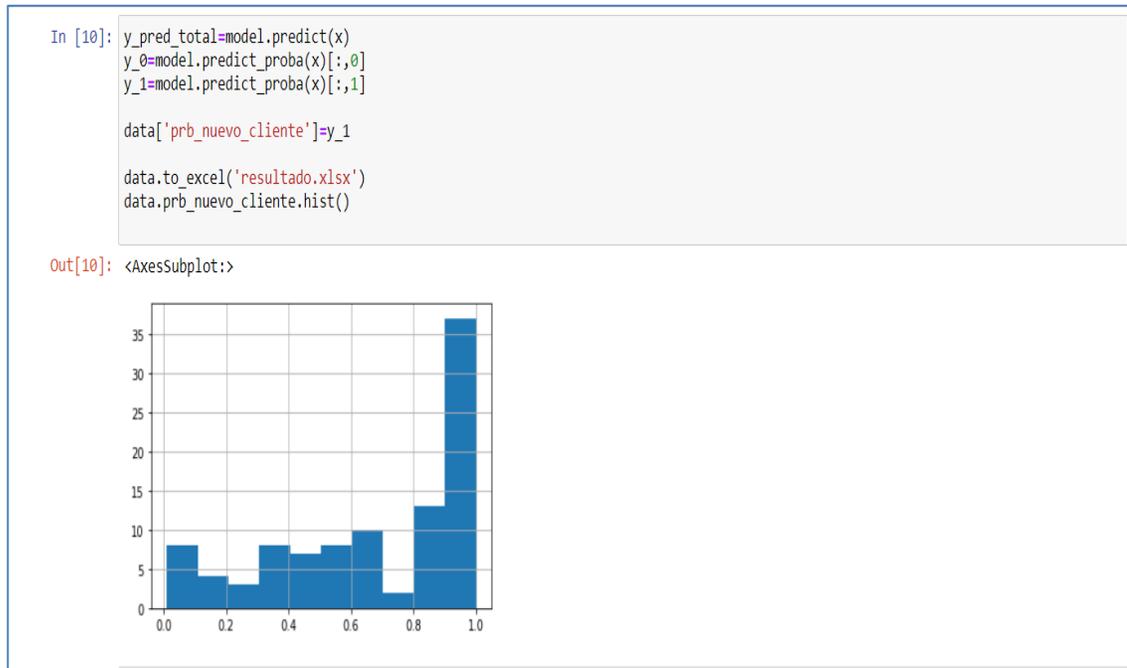


Figura 50. Consume_modelo03
Fuente: elaboración propia

5.3. Desarrollo del prototipo del Cuadro de Mando en Power BI

A continuación, para el desarrollo del Dashboard se utiliza la herramienta Power BI, se usan los indicadores planteados, se realizó la construcción del cuadro de mando, que permitirá visualizar de forma gráfica, así como facilitar la comprensión de la información y dar apoyo a la toma de decisiones.

Iniciamos con la conexión de la herramienta de la base de datos SQL creada ingresando el nombre de nuestro servidor SQL SERVER, como se muestra en la siguiente imagen.

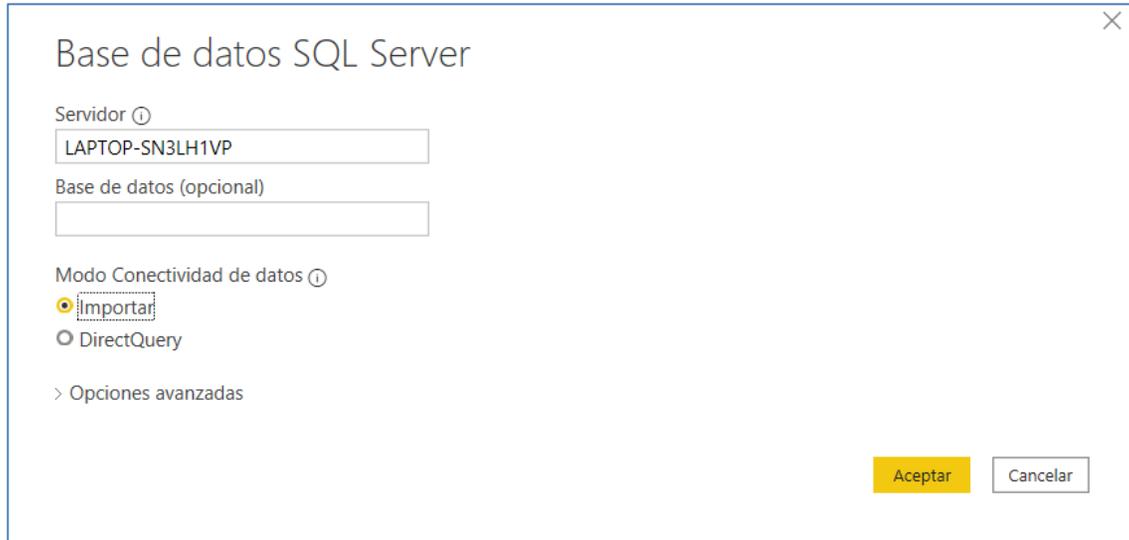


Figura 51. Conexión con la base de datos SQL Server
Fuente: elaboración propia

Una vez creada la conexión, continuamos seleccionando las tablas de nuestra base de datos para la construcción del cuadro de mando, como se muestra en la siguiente imagen.

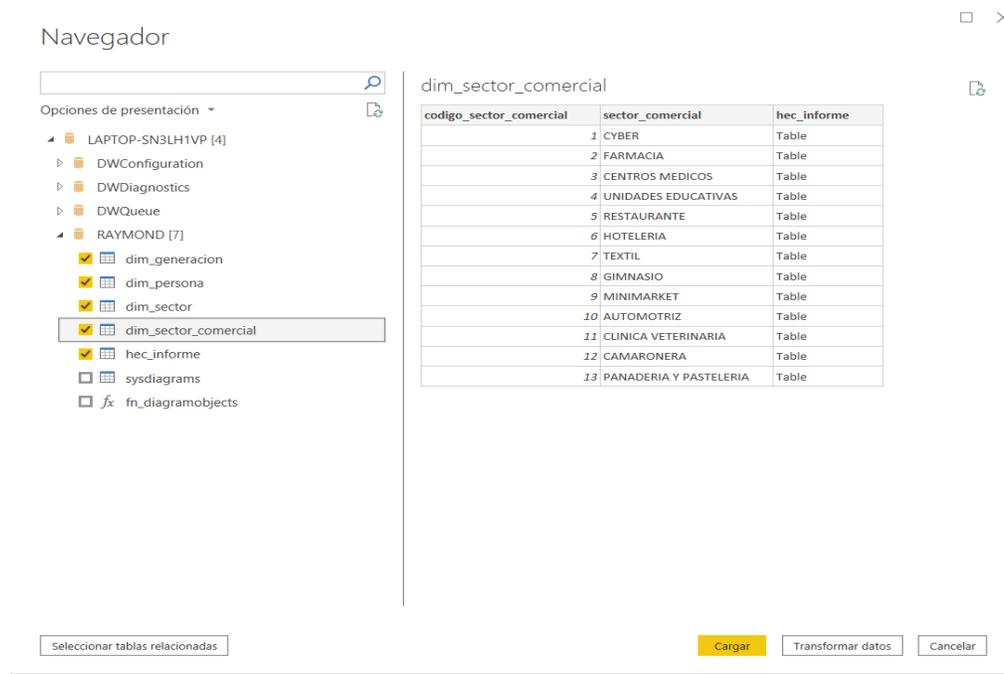


Figura 52. Cargar tablas de la base de datos
Fuente: elaboración propia

Se puede observar en la imagen (figura 92) las dimensiones con sus respectivas conexiones a la tabla de hechos.

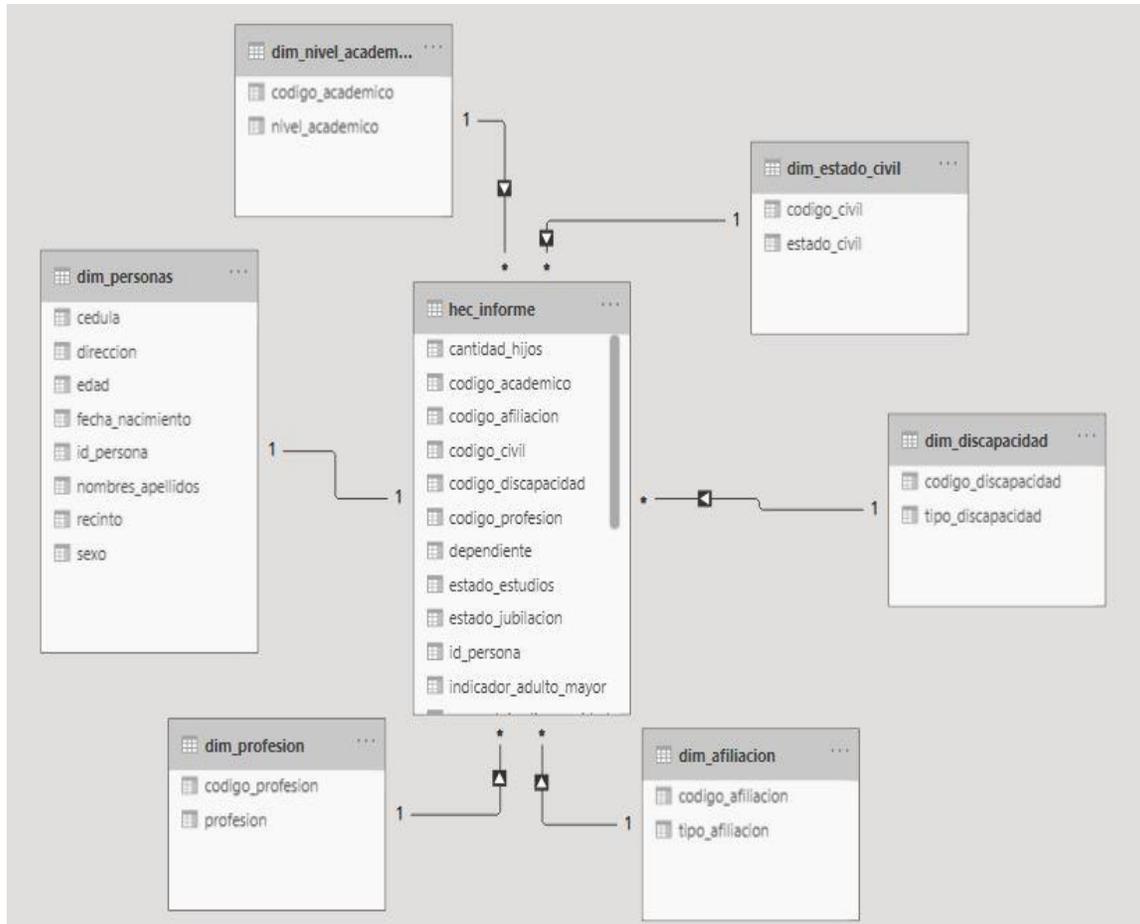


Figura 53. Diagrama Estrella en Power BI
Fuente: elaboración propia

Una vez cargadas las tablas de la base de datos, se procede a crear la presentación de los resultados como se muestra a continuación.

En este dashboard se visualizan los diferentes sectores comerciales a los que pertenecen las empresas, distribuidas tanto para el grupo de los no clientes (lado izquierdo) y los clientes

(lado derecho). Mediante barras se observa el volumen de registros que pertenece a los distintos rangos según la probabilidad de ser cliente.

El dashboard fue generado con la intención de ver el comportamiento de los registros de acuerdo con las probabilidades y el estatus de cliente o no cliente para confirmar la validez del modelo; es decir, que las probabilidades más altas deben concentrarse en los registros que históricamente han sido clientes.

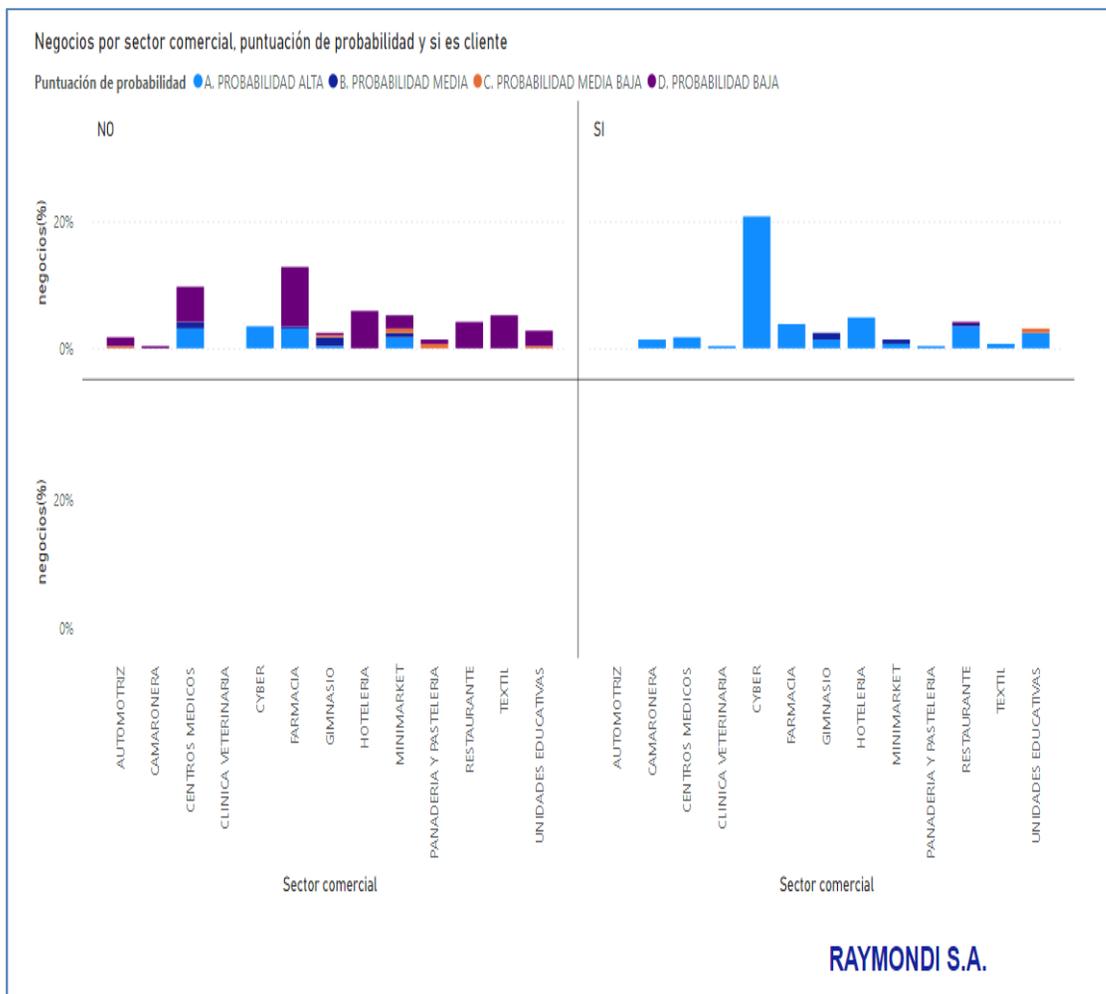


Figura 54. Pantalla 1
 Fuente: elaboración propia

El dashboard indica que en la parte de cyber es muy alta la aceptación de los servicios ofrecidos por Distecom, se dividen también entre sexo (masculino y femenino), generación (millennials, centennials) y sectores de la ciudad, para una mejor toma de decisiones.

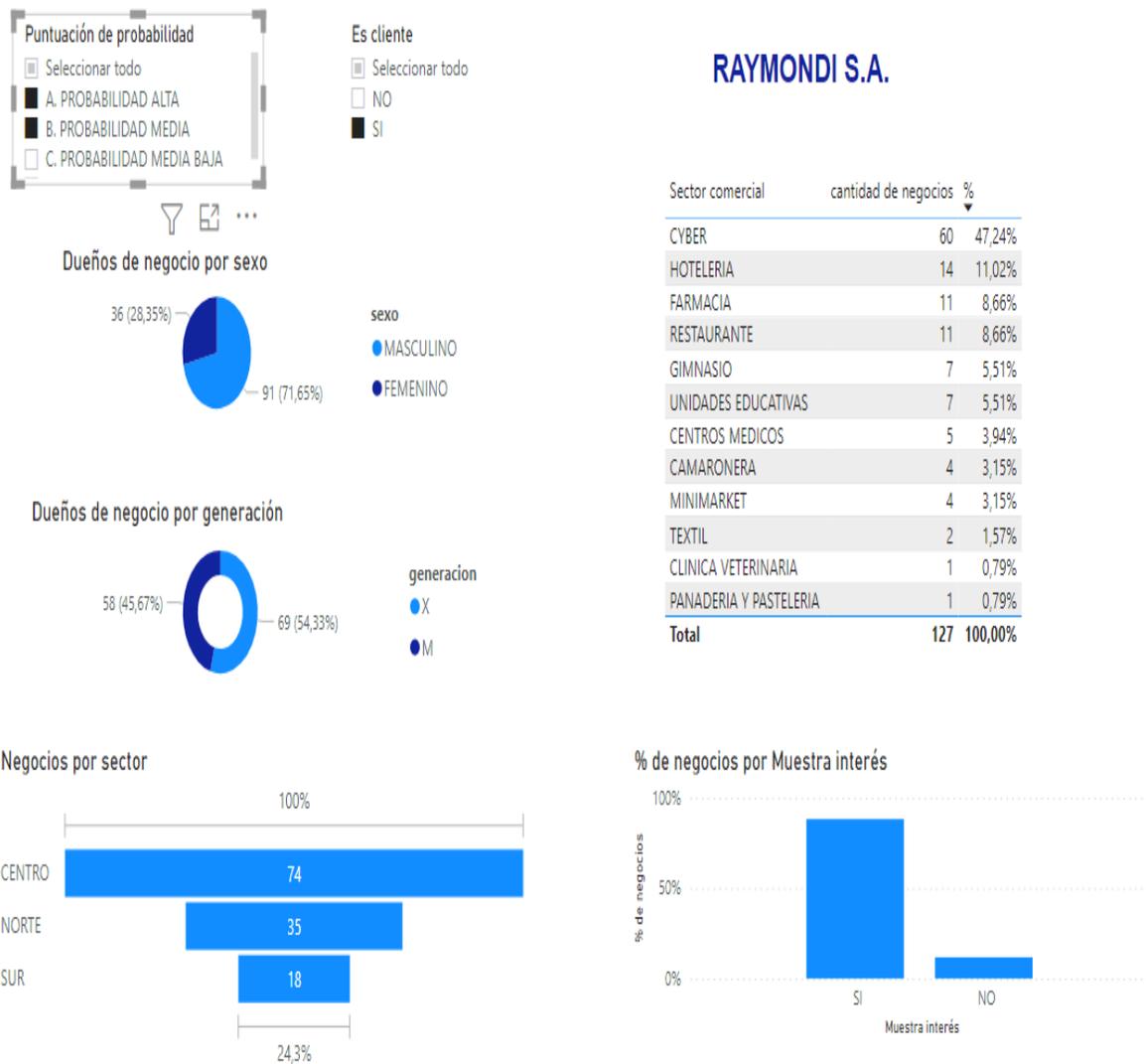


Figura 55. Pantalla 2
Fuente: elaboración propia

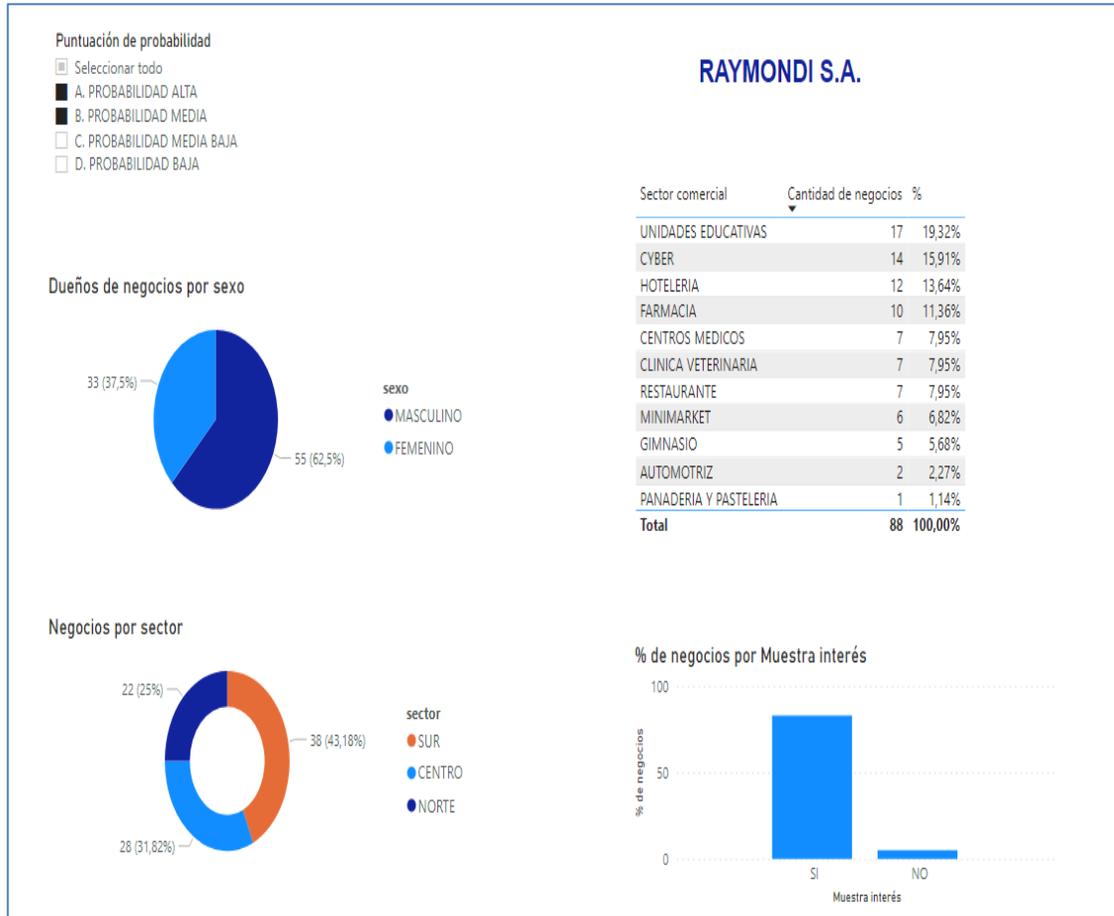


Figura 56. Pantalla 3
Fuente: elaboración propia



Figura 57. Columna calculada
Fuente: elaboración propia

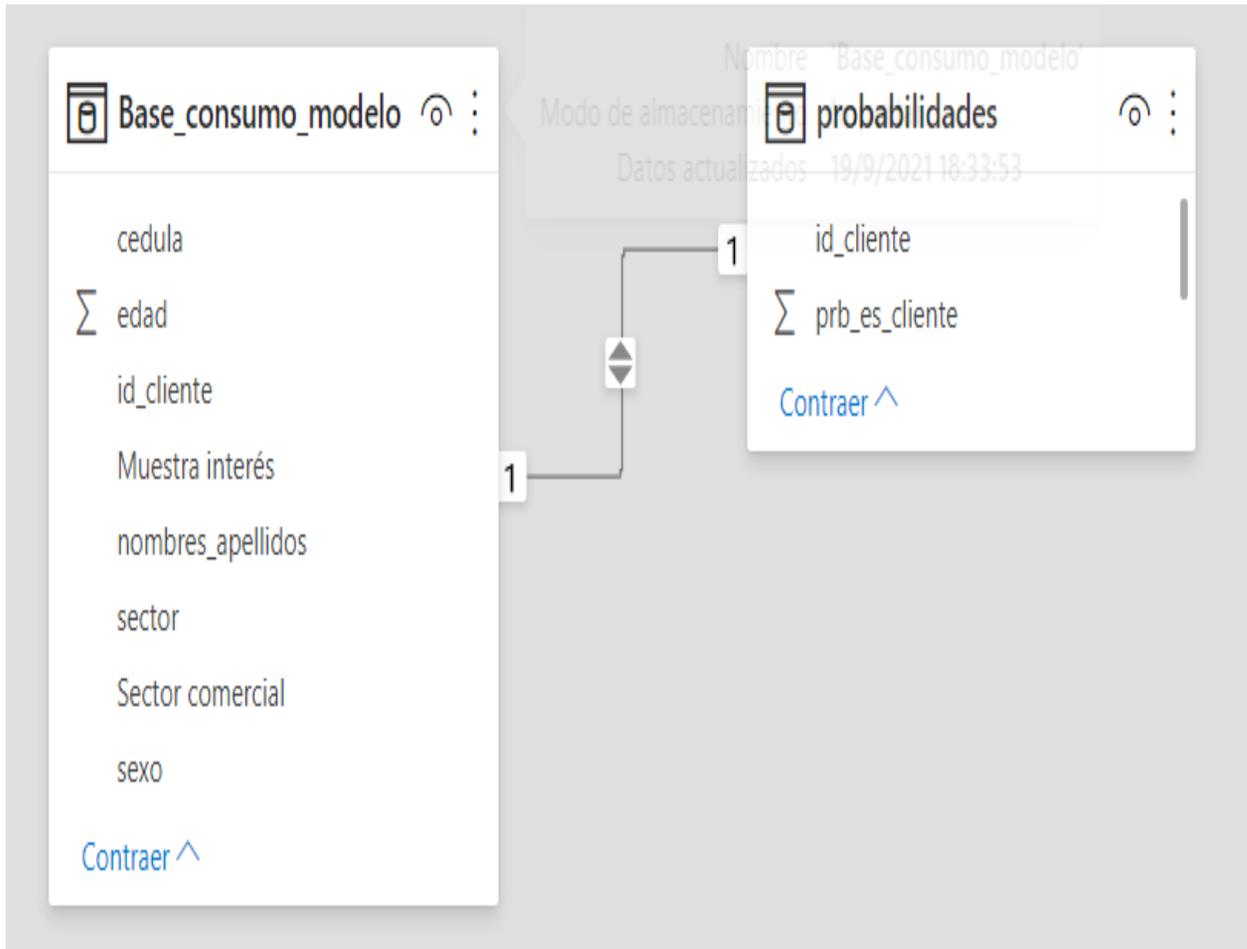


Figura 58. Modelo
Fuente: elaboración propia



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación
William Rafael Raymondi Lomas
Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023





Conclusiones

La realización del presente proyecto sirvió como guía para la futura adaptación de un modelo predictivo que permita identificar la recomendación de los clientes, efectuándose así el objetivo general del trabajo.

Del mismo modo, se cumplieron con los objetivos específicos del proyecto de investigación a lo largo del estudio ejecutado, que consistían en determinar los factores que influyen en la recomendación de los clientes de la empresa Distecom, identificar el comportamiento de los clientes para desarrollar un modelo predictivo con el propósito de mantener la rentabilidad del negocio, y por último, validar el modelo de análisis predictivo que permita pronosticar la referencia de los clientes, utilizando Python, por ello, la compilación de la información respalda la posibilidad de desarrollar y viabilidad del proyecto.

El desarrollo del análisis predictivo utilizando Python resultó una herramienta de gran utilidad para la empresa Distecom, ante ello, la planeación ayudó a determinar las debilidades y fortalezas del proyecto, aperturando campo para futuras tomas de decisiones, con el fin de que se pueda desarrollar en el negocio de manera más eficiente y apropiada posible.

Los factores que influyen en los clientes para que estos recomienden la marca Distecom son principalmente la confianza, la alta calidad de los productos y servicios, y su utilidad para satisfacer las necesidades de los clientes. A esto se suma la atención al cliente.



Cuando se cumplen las condiciones para que los clientes refieran a prospectos, es extremadamente probable que se transforme en cliente frecuente, esto es que vuelva a demandar bienes y servicios ofertados por Distecom.

El modelo de análisis predictivo establece una puntuación de probabilidad de referencia por parte de un cliente, se segmenta en cuatro tramos, uno de probabilidad alta, media, media-baja y baja. También indica una segmentación por rama económica, mientras más vinculado su modelo de negocio al hardware es más probable que refieran, tal es el caso de los cyber que tienen una alta puntuación. En cuanto al sexo, también es relevante.

Lo que se buscó con esta propuesta fue principalmente identificar la posibilidad de que los clientes puedan referir los productos y servicios que promueve la empresa Distecom para conservar la estabilidad y rentabilidad del negocio.



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación
William Rafael Raymondi Lomas
Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023





Recomendaciones

Es necesario que, a partir de la implementación de un modelo predictivo, la empresa inicie un proceso de recopilación de información del cliente más amplio, tales como su perfil personal, profesional y de su negocio, en esta última categoría es particularmente relevante su tamaño, su rama económica, ubicación geográfica, entre otros aspectos, monto promedio de transacciones, tipo de demanda de bienes o servicios a la empresa. Lo antedicho permitirá disponer de un perfil de donde obtener los determinantes.

Para el comportamiento del cliente es recomendable disponer de un sistema más estricto y sistemático de nivel de expectativas, percepción del servicio al cliente y nivel de satisfacción. Esto permitirá establecer una relación entre perfil y comportamiento para generar referidos.

Las labores de validación se realizarán estableciendo un seguimiento y relación entre los resultados que ofrece el modelo y el incremento real de clientes por la vía de la validación. Esto exige un esfuerzo del área de marketing para al tiempo que se establece la relación se enriquezca mediante la incorporación de nuevas variables que fortalezcan el proceso.



Book Citation Index

Web of Science/Core Collection

Modelos predictivos para identificar probabilidades de recomendación
William Rafael Raymondi Lomas
Sergio Israel Peña Guano



9 780311 000470

Recepción: 12-03-2023

Aprobación: 11-06-2023



Bibliografía

- Altuz. (2020). <https://altuz.com/altuz/big-data-la-importancia-de-los-modelos-predictivos-para-potenciar-tu-negocio/>
- Bedoya, O., López, M., & Marulanda, C. (2019). Modelo predictivo para la identificación de factores socioculturales asociados al tiempo de búsqueda del primer empleo en egresados universitarios. *Virtual Universidad Católica del Norte* (58), 3-18.
- Bernal, C. (2010). *Metodología de la Investigación*. Tercera.
- Cabrera, G. (2017). *Metodología para la construcción de un sistema de ayuda a la decisión para Instituciones Financieras del Sector de la Economía Popular y Solidaria: Un enfoque basado en conceptos de Datawarehouse*. Universidad de Cuenca.
- Calzada, J. (2015). *Metodología Agile y las técnicas de desarrollo del Kimball Group*. <https://www.certia.net/metodologia-agile-y-las-tecnicas-de-desarrollo-del-kimball-group/>
- Carrillo, S. (2020). *Enroke*. <https://blog.grupoenroke.com/como-implementar-un-sistema-de-referidos>
- Carvajal, N. (2020). *Desarrollo de un modelo predictivo de morosidad de cartera de la línea de fianzas de una empresa de seguros*. Universidad de las Fuerzas Armadas. Sangolquí: ESPE.

Casermeiro, M. (2014). *Técnicas de información y atención al cliente consumidor*. España:

IC.

Castallano, M. (2006). *Inteligencia de negocios*. México.

Chagcha, L., & Urvina, K. (2016). *Herramienta informática de business intelligence para el departamento de ventas en la empresa Mascorona*. Universidad Técnica de Ambato. Ambato: UTA.

Chang, L., Rosete, F., Charre, J., & Mas, J. (2020). *Validación de modelos predictivos de cambio de cubierta y uso del suelo en la península de Baja California, México*. Investigaciones Geográficas.

Espinosa, J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería, Investigación y Tecnología*, 21(1).

Keyrus. (2021). <https://keyrus.com/sp/es/insights/las-11-tecnicas-mas-utilizadas-en-el-modelado-de-analisis-predictivos>

Larc, A. (2016). *Ocho pasos para incrementar tus ventas fácilmente*. Barcelona, Mestas.

Laseca, E. (2016). *Venta por referidos*. Developing the business. <https://developingthebusiness.com/venta-por-referidos/>

López, P., & Fachelli, S. (2015). *Metodología de la investigación social cuantitativa*. Barcelona, España: UB.

MathWorks. (2020). <https://es.mathworks.com/discovery/predictive-analytics.html>



Maya, N., Orona, M., & Barron, M. (2019). Diseño de un sistema de análisis de datos para trayectorias educativas UTVT. *Teoría Educativa*, 3(9), 10-24.

Silva, G., Zapata, V. M., & Toaquiza, L. (2019). Análisis de metodologías para desarrollar Data Warehouse. *Ciencia Digital*, 3(3.4), 397-418.

Villena, J. (2016). *CRISP-DM: La metodología para poner orden en los proyectos*.
<https://www.sngular.com/es/data-science-crisp-dm-metodologia/>

Modelos predictivos para identificar probabilidades de recomendación



Editorial Tecnocientífica Americana

Domicilio legal: calle 613sw 15th, en Amarillo, Texas. **ZIP:** 79104, EEUU

Teléfono: 7867769991

Fecha de publicación: 27 junio de 2023

Código BIC: KNTX

Código EAN: 9780311000470

Código UPC: 978031100047

ISBN: 978-0-3110-0047-0

La Editorial Tecnocientífica Americana se encuentra indizada en, referenciada en o tiene convenios con, entre otras, las siguientes bases de datos:

