



**Aplicación de técnicas de minería de datos para mejorar la gestión de los clientes,
administrados por un asistente virtual (Chatbot) en la empresa SICE**

Jumbo Carrión, Roberth Mauricio

Departamento de Ciencias de la Computación

Carrera de Ingeniería de Sistemas e Informática

Trabajo de titulación, previo a la obtención del título de Ingeniero en Sistemas e Informática

Ing. Díaz Zúñiga, Magi Paúl

5 de agosto del 2022

Análisis de similitud de contenidos



Tesis_V7_Jumbo_Roberth.pdf

Scanned on: 0:7 July 22, 2022 UTC



Firmado electrónicamente por:
**MAGI PAUL
DIAZ**



Overall Similarity Score



Results Found



Total Words in Text

Identical Words	114
Words with Minor Changes	74
Paraphrased Words	725
Omitted Words	2414



Departamento de Ciencias de la Computación

Carrera de Ingeniería de Sistemas e Informática

Certificación

Certifico que el trabajo de titulación, **“Aplicación de técnicas de minería de datos para mejorar la gestión de los clientes, administrados por un asistente virtual (Chatbot) en la empresa SICE”** fue realizado por el señor **Jumbo Carrión, Roberth Mauricio**; el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que losustente públicamente.

Sangolquí, 21 de julio de 2022

Firma:



Firmado electrónicamente por:
**MAGI PAUL
DIAZ**

.....
Ing. Magi Paul Diaz Zuñiga MSc.

C.C: 1707249072



**Departamento de Ciencias de la Computación
Carrera de Ingeniería de Sistemas e Informática**

Responsabilidad de Autoría

Yo, **Jumbo Carrión, Roberth Mauricio** con cédula de ciudadanía N° **1104982549** declaro que el contenido, ideas y criterios del trabajo de titulación: **“Aplicación de técnicas de minería de datos para mejorar la gestión de los clientes, administrados por un asistente virtual (Chatbot) en la empresa SICE”** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 21 de julio de 2022

.....
Jumbo Carrión, Roberth Mauricio.

C.C: 1104982549



Departamento de Ciencias de la Computación
Carrera de Ingeniería de Sistemas e Informática

Autorización de Publicación

Yo, **Jumbo Carrión, Roberth Mauricio** con cédula de ciudadanía N° **1104982549**, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: Título: **“Aplicación de técnicas de minería de datos para mejorar la gestión de los clientes, administrados por un asistente virtual (Chatbot) en la empresa SICE”** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 21 de julio de 2022

Firma:

.....
Jumbo Carrión, Roberth Mauricio

C.C.: 1104982549

Dedicatoria

Este presente trabajo de titulación está dedicado a:

A mis padres Salustino y Esthela; mi padre que desde el cielo fue mi guía y protección en toda mi vida estudiantil, fue ese ángel que siempre que le pedía fuerzas y motivación para seguir adelante en los momentos difíciles de mi carrera, siempre estuvo ahí, a mi madre por ser mi todo, mi guía, mi fuente de motivación, quien, con su amor, esfuerzo me dio de lo que poco que tenía para que yo pudiera al fin lograr este logro académico, mi inmenso amor y gratitud hacia ti mamá.

A mis hermanos, quienes con su cariño y apoyo me motivaban a seguir siempre adelante, una especial dedicatoria a mis hermanos: Iván, Vinicio, Cristóbal y Raúl, Reina, Cheo, Cecilia y Amada quienes muchas de las veces se esforzaron por mi para que no me falte nada durante toda mi vida estudiantil este logro también es de ustedes ya que sin su apoyo no hubiera podido llegar hasta aquí.

A mis tíos, primos quienes de una u otra forma me apoyaron para que pueda alcanzar este objetivo.

Finalmente, a todos quienes estuvieron a mi lado brindándome su apoyo para que pudiera alcanzar este logro académico.

Agradecimiento

Primeramente, a Dios por darme salud, sabiduría e inteligencia, a la Virgen en su advocación del Cisme por guiarme y protegerme en cada paso que he dado a lo largo de toda mi vida.

Un agradecimiento especial a mi mamá y mis hermanos quienes son las personas más importantes en mi vida, por siempre estar a mi lado apoyándome, dándome ánimos para que nunca me dé por vencido y para que en estos momentos este celebrando con ustedes este logro, mis más sinceros agradecimientos sin su apoyo no hubiera sido imposible alcanzar este logro.

De igual manera mi agradecimiento a mis amigos que pude conocer a lo largo de toda mi vida estudiantil, amigos de aventuras, de estudio, gracias por su amistad, consejos y apoyo incondicional.

Finalmente quiero agradecer al Ing. Paúl Díaz quien con su guía, apoyo y conocimiento a lo largo de mi vida universitaria permitió el planteamiento, desarrollo e implementación de este trabajo de investigación.

Índice de contenido

Análisis de similitud de contenidos.....	2
Dedicatoria.....	6
Agradecimiento.....	7
Índice de contenido	8
Índice de tablas	11
Índice de figuras.....	12
Resumen	14
Abstract.....	15
Capítulo I	16
Introducción.....	16
Antecedentes	16
Planteamiento del Problema	18
Justificación.....	20
Objetivos	20
<i>General</i>	20
<i>Objetivos Específicos</i>	21
Alcance	21
Hipótesis	23
Capítulo II	24
Marco metodológico.....	24
Estado del Arte.....	24
<i>Planteamiento del estudio sistemático de literatura</i>	24
<i>Definición del grupo de control y extracción de términos</i>	25
<i>Construcción de la cadena de búsqueda</i>	27
<i>Selección de los estudios primarios</i>	28
<i>Elaboración del estado del arte</i>	30
Metodología.....	35
<i>Fase 1. Comprensión del giro del negocio</i>	35
<i>Fase 2. Estudio y comprensión de los datos</i>	35
<i>Fase 3. Análisis de los datos y selección de las características</i>	35

<i>Fase 4. Modelado</i>	35
<i>Fase 5. Evaluación</i>	36
<i>Fase 6 Despliegue o Implementación</i>	36
Marco Teórico	36
Inteligencia Artificial	36
<i>Aplicaciones de la IA</i>	37
<i>Técnicas de la Inteligencia Artificial</i>	38
Minería de Datos.....	42
<i>Aplicaciones de la minería de datos</i>	43
Machine Learning.....	45
<i>Aprendizaje no supervisado</i>	45
<i>Algoritmos de Clustering</i>	46
<i>K-Means</i>	48
<i>K-Medoids</i>	48
<i>Apriori</i>	49
Procesamiento de lenguaje natural	49
Sistema de Peticiones, Quejas, Reclamos o Sugerencias (PQRS)	53
Análisis RFM	54
Chatbot	55
<i>Tipos de Chatbot</i>	55
Herramientas para la obtención de datos desde los canales conversacionales.....	56
<i>Dialogflow</i>	59
Herramientas para el almacenamiento de los datos.....	60
<i>MongoDB</i>	63
Herramientas para el análisis de datos.....	63
<i>Lenguaje R</i>	65
Otras herramientas	66
<i>Heroku</i>	66
<i>Node.js</i>	66
Capitulo III	67
Propuesta e implementación del Chatbot	67
Desarrollo del modelo en base a la metodología CRISP-DM	103
Capítulo IX.....	131

Implementación y análisis de resultados.....	131
Ejecución.....	133
Análisis de los resultados.....	139
Capítulo V.....	143
Conclusiones y recomendaciones.....	143
Conclusiones.....	143
Recomendaciones.....	144
Bibliografía.....	146

Índice de tablas

Tabla 1 Artículos que conforman el grupo de control	25
Tabla 2 Versiones de la cadena de búsqueda.....	28
Tabla 3 Estudios primarios.....	29
Tabla 4 Comparación entre Dialogflow, IBM Watson y AmazonLex	57
Tabla 5 Comparación SQL y NoSQL.....	60
Tabla 6 Comparación Casanda, MongoDB y Redis.....	61
Tabla 7 Comparación Weka y R	63
Tabla 8 Preguntas frecuentes SICE	78
Tabla 9 Estructura dataframe df_RFM	107
Tabla 10 Puntuaciones en función de cuartiles.....	112
Tabla 11 Resumen clústeres K-means	123
Tabla 12 Estructura de la colección Knowledges.....	131
Tabla 13 Estructura de la colección frequents	132

Índice de figuras

Figura 1 Árbol de problemas	19
Figura 2 Aprendizaje no supervisado.....	45
Figura 3 Análisis Léxico	50
Figura 4 Análisis sintáctico	51
Figura 5 Aprendizaje profundo	52
Figura 6 Arquitectura Dialogflow.....	59
Figura 7 Flujo de la conversación entre el usuario y Chatbot.....	72
Figura 8 Estructura de los documentos de la colección courses	73
Figura 9 Estructura de los documentos de la colección Chatbotusers.....	74
Figura 10 Estructura JSON colección courses.....	74
Figura 11 Arquitectura del Chatbot "One Click Integration"	76
Figura 12 Fulfillment en Dialogflow	77
Figura 13 Arquitectura Fulfillment con Webhook externo (Node.js) y MongoDB.....	83
Figura 14 Proceso de la intención Certificaciones.....	84
Figura 15 Proyecto en MongoDB.	88
Figura 16 Configuración proyecto MongoDB.....	88
Figura 17 Insert documento en la colección courses	89
Figura 18 MongoDB documentos de la colección "courses".....	89
Figura 19 Configuración general del Chatbot "ChatBotSICE"	90
Figura 20 Creación de la intención de bienvenida.....	91
Figura 21 Creación de la intención "CertificacionesSice" en Dialogflow	92
Figura 22 Creación de la intención en Dialogflow denominada "EmailUsuario"	92
Figura 23 Configuración "action" en cada una de las intenciones.....	93
Figura 24 Configuración app de facebook developers.....	94
Figura 25 Configuraciones de token y Webhooks para Facebook Messenger.....	95
Figura 26 Integración Facebook en Dialogflow Integrations	96
Figura 27 Activación Webhook en Dialogflow Fulfillment	97
Figura 28 Estructura servidor en Node.js.....	98
Figura 29 Dependencias necesarias para ejecutar el proyecto en Node.js.....	99
Figura 30 Archivo de configuraciones de credenciales config.js.....	99
Figura 31 Configuración Servidor Express & Conexión MongoDB.....	100
Figura 32 Configuración global en Google Cloud.	100
Figura 33 Definición esquemas (Esquema CertificationSchema)	101
Figura 34 facebookBot.js en Node.js	102
Figura 35 Creación y configuración servidor en la nube en Heroku.	103
Figura 36 Creación campo Recencia RFM	106
Figura 37 Cálculo de la frecuencia RFM	107
Figura 38 Cálculo del monto RFM.....	107
Figura 39 Histograma de frecuencia de compra	109
Figura 40 Histograma de monto de compra.....	109
Figura 41 Histograma recencia de compra	110
Figura 42 Diagrama variable frecuencia RFM.....	111
Figura 43 Diagrama variable monto RFM	111

Figura 44	Diagrama variable recencia RFM	112
Figura 45	Normalización de las puntuaciones.....	114
Figura 46	Nuevo Data frame con datos normalizados	115
Figura 47	Código cálculo de la matriz de distancias RStudio	116
Figura 48	Visualización matriz de distancias	116
Figura 49	Cálculo de número de clústeres RStudio	117
Figura 50	Número de clústeres optimo RStudio	117
Figura 51	Aplicación algoritmo K-means.....	118
Figura 52	Gráfico clústeres agrupados.....	119
Figura 53	Gráfica 2 clústeres agrupados	120
Figura 54	Función para la visualización del dendograma resultante	120
Figura 55	Dendograma Sice.....	121
Figura 56	Representación centros y fórmula para el cálculo de la distancia al punto cero.....	122
Figura 57	Código cálculo de los centros RStudio	122
Figura 58	Data frame final k-means	124
Figura 59	Data frame inicial Apriori	126
Figura 60	Limpieza del dataframe sice_apriori.....	127
Figura 61	Data frame Lista_cursos	127
Figura 62	Comando para la creación del histograma de cursos más vendidos	128
Figura 63	Histograma cursos más vendidos SICE	128
Figura 64	Obtención reglas de asociación.....	129
Figura 65	Limpieza de las reglas de asociación	129
Figura 66	Conjunto de reglas de asociación finales	130
Figura 67	Entidad TipoCertificación	135
Figura 68	Intención CertificacionesSiceInfo.....	136
Figura 69	Chatbot SICE	137
Figura 70	Regla de asociación activada en el mensaje del cliente.....	137
Figura 71	Sugerencia de compra al cliente en base al modelado con el algoritmo Apriori.....	138
Figura 72	Flujo final usuarios/Chatbot	139
Figura 73	Estadística Mensajes Fan Page SICE antes de aplicar los resultados	140
Figura 74	Estadística alcance de la página SICE	141
Figura 75	Estadística interacciones en la página SICE	141
Figura 76	Estadística Mensajes Fan page SICE después de aplicar el modelo.....	142

Resumen

La empresa SICE requiere mejorar la comunicación y toma de decisiones con respecto a la información proporcionada por los clientes en cada uno de sus canales comunicacionales, todo esto con el fin de ofrecer una atención personalizada y eficiente a cada uno de los usuarios que se pongan en contacto con la empresa. Como primera solución se propone la implementación de un Chatbot basado en inteligencia artificial tanto en la Fan page como en la página web de la empresa; segundo un modelo que permita agrupar e identificar los patrones de comportamiento de cada uno de los clientes mediante la aplicación de técnicas de inteligencia artificial (Minería de Datos), con el fin de mejorar la atención y la toma de decisiones sobre la información proporcionada por cada uno de los clientes.

Para llevar a cabo las soluciones propuestas se usó herramientas tecnológicas como Dialogflow que es el agente conversacional de Google para establecer la comunicación con cada uno de los clientes, RStudio para el análisis, tratamiento, limpieza y entrenamiento de los datos proporcionados por la empresa. Como técnica se usó el agrupamiento o clustering que es una de las tantas técnicas de Minería de Datos y una de las más usadas cuando se trata del análisis de datos en pequeños grupos, donde cada uno de estos grupos contiene data similar y existe una significativa diferencia con respecto a los demás grupos.

La metodología usada es la CRISP-DM, ya que esta permite la comprensión del negocio, análisis y simulación a los datos como corresponda, el análisis se lo hizo con la aplicación del modelo RFM (Recencia, Frecuencia, Monto) para luego aplicar los algoritmos tanto de agrupamiento: k-means, y el algoritmo de asociación Apriori para encontrar las asociaciones que existen entre cada uno de los cursos y certificaciones que ofrece la empresa.

Palabras Clave: minería de datos, Chatbot, mejora de la toma de decisiones.

Abstract

The SICE company needs to improve communication and decision-making regarding the information provided by customers in each of its communication channels, all this in order to offer personalized and efficient attention to each of the users who contact us contact with the company. As a first solution, the implementation of a Chatbot based on artificial intelligence is proposed both on the Fan page and on the company's website; second, a model that allows grouping and identifying the behavior patterns of each one of the clients through the application of artificial intelligence techniques (Data Mining), in order to improve the attention and decision-making on the information provided by each one of the clients.

To carry out the proposed solutions, technological tools such as Dialogflow, which is Google's conversational agent, were used to establish communication with each of the clients, RStudio for the analysis, treatment, cleaning and training of the data provided by the company. As a technique, grouping or clustering was used, which is one of the many Data Mining techniques and one of the most used when it comes to data analysis in small groups, where each of these groups contains similar data and there is a significant difference. relative to the other groups.

The methodology used is the CRISP-DM, since it allows the understanding of the business, analysis and simulation of the data as appropriate, the analysis was done with the application of the RFM model (Recurrence, Frequency, Amount) to then apply the algorithms for both grouping: k-means, and the Apriori association algorithm to find the associations that exist between each of the courses and certifications offered by the company.

Keywords: data mining, Chatbot, improved decision making.

Capítulo I

Introducción

Antecedentes

Las organizaciones utilizan la flexibilidad como estrategia para adaptarse a un mercado globalizado con el fin de obtener los mejores resultados de la gestión organizacional, dando como resultado procesos que afectan sus sistemas estructurales. Por lo tanto, una empresa ágil está orientada al cliente, tiene nuevas tecnologías y es organizada e innovadora con acuerdos paralelos (Hansen & Mouritsen, 1999).

La empresa SICE ofrece servicios educativos como asesorías, capacitaciones, talleres, conversatorios, mesas de trabajo, workshop, selección de personal, entre otros; orientadas al campo educativo y otros ámbitos profesionales.

La empresa mencionada anteriormente gestiona las incidencias de los consumidores a través del sistema PQR's (Peticiónes, Quejas, Reclamaciones y Sugerencias) y encuestas de satisfacción, esto con el fin de poder tomar decisiones sobre nuevos cursos o a su vez poder resolver cualquier inquietud que surja por parte de los usuarios.

Gracias al progreso de la tecnología, actualmente se usan varias herramientas tecnológicas que ayudan al desempeño más rápido de los procesos, de esta manera se busca poder transformar y procesar datos con el fin de obtener información útil para las organizaciones.

Actualmente, la inteligencia artificial (IA) es uno de los factores fundamentales que se implementan en las empresas que se encuentran dentro de los países desarrollados y se está infiltrando gradualmente en sus estructuras. Las primeras ideas sobre la inteligencia artificial surgieron en la segunda mitad del siglo XX, cuando, desde la ciencia ficción, se pensaba que

en un futuro próximo existirían robots antropomórficos que serían utilizados con el fin de realizar todas las actividades que son designadas para los humanos. La realidad es que la situación actual es un poco diferente a lo que se planteaba en aquel entonces. Por ejemplo, los robots diseñados y alimentados por inteligencia artificial son capaces de superar con creces las capacidades humanas en problemas técnicos como las matemáticas o el ajedrez, sin embargo, se limitan a actividades puntuales y monótonas (Huimin, Yujie, Min, Kim, & Seiichi, 2018).

Albarrán Trujillo y Salvado Gallegos (2013) mencionan que uno de los factores principales de la IA es un sistema experto (SE). El surgimiento de estos sistemas fue solo unos años después de la inteligencia artificial, y los primeros artículos académicos sobre inteligencia artificial aparecieron alrededor de 1960.

El principio de los SE es prevenir que los expertos humanos se vean obligados a dedicar tiempo para actividades repetitivas. Según Almurshidi (2018), los SE reciben estímulos externos, los cuales son analizados a partir de la información codificada que poseen y generan automáticamente respuestas específicas.

El número de empresas que tiene como apoyo a las tecnologías de inteligencia artificial aplicadas a las interacciones con los clientes sigue creciendo. Estos sistemas sirven de apoyo para los analistas ya que ayuda a comprender el comportamiento del consumidor y diseñar estrategias promocionales basadas en las necesidades reales del negocio, obtener una ventaja competitiva de sus datos y mejorar su rentabilidad (Angstenberger, 1998).

El gasto en sistemas cognitivos y de IA alcanzará los 77 600 millones de dólares para 2022 (más del triple de los 24 000 dólares estimados para 2018), según la última actualización de International Data Corporation (IDC) en su guía bianual de gasto mundial en sistemas cognitivos de IA. La tasa de crecimiento anual compuesto (TCCA) para el período de pronóstico 2017-2022 es 37.3% (Solutions Artificial, 2020).

El software es considerado como una de las categorías relacionadas a la tecnología más relevante ya que posee un rápido crecimiento, lo cual representa aproximadamente el 40% del gasto cognitivo/intelectual con una TCCA del 43,1% determinado para un periodo de cinco años. Las dos áreas que contribuyen a este enfoque en las inversiones son las aplicaciones de aprendizaje profundo y aprendizaje automático, así como las aplicaciones de IA conversacionales (como asistentes personales y Chatbots) (empleadas en una amplia gama de casos de uso) (Solutions Artificial, 2020).

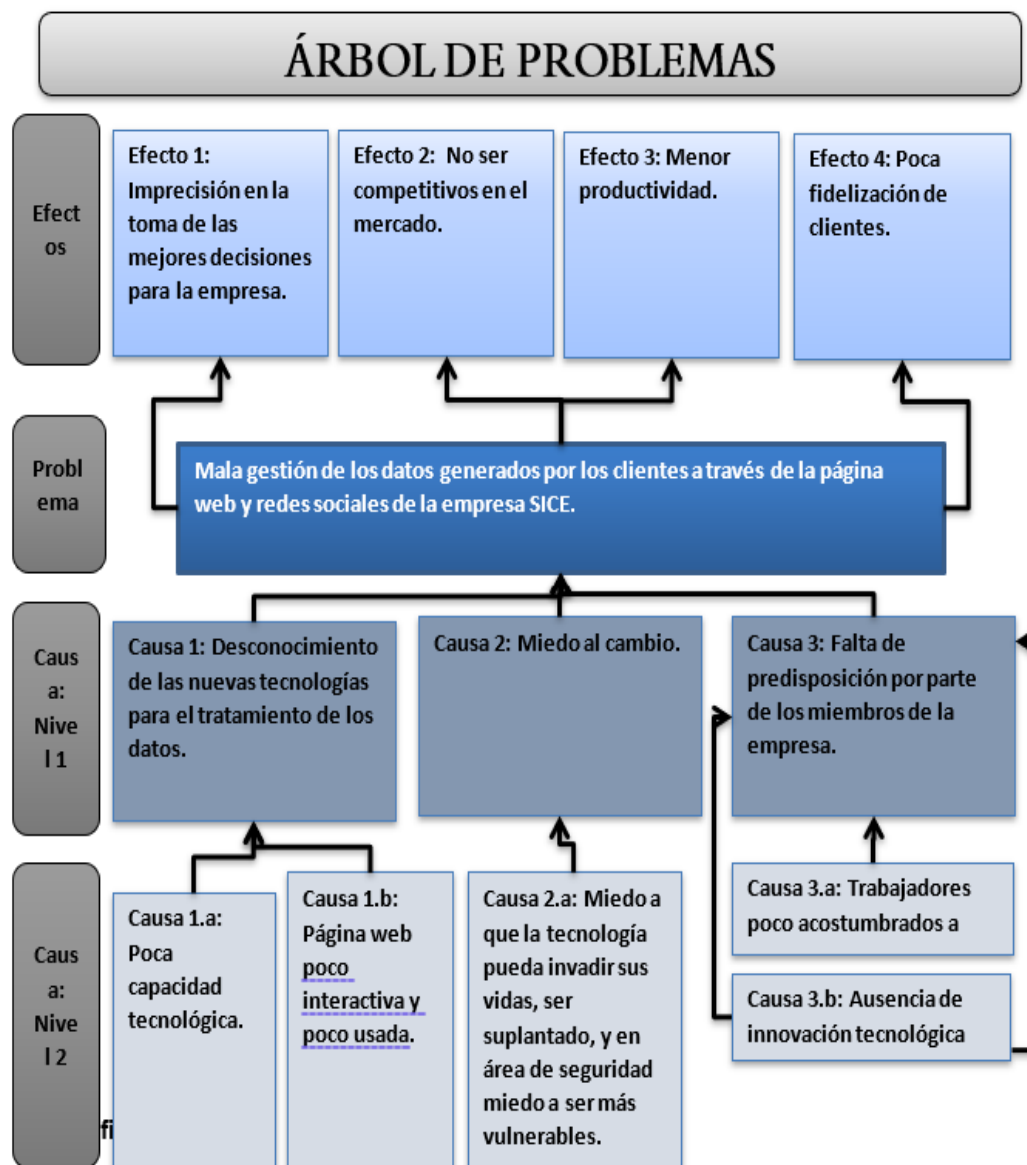
Planteamiento del Problema

En tiempos donde la información crece aceleradamente, las dinámicas van cambiando considerablemente en todos los ámbitos. El campo de la enseñanza no es la excepción ya que se ha vuelto un reto a la hora de encontrar nuevas formas de transmitir el conocimiento.

La empresa SICE que actualmente ofrece servicios educativos como asesorías, capacitaciones, talleres, conversatorios, mesas de trabajo, selección de personal, entre otros; relacionados con la educación y en áreas profesionales, gestiona la atención de sus clientes y las incidencias a través del sistema de PQR's (Peticiones, Quejas, reclamos y sugerencias), el mismo no cuenta con un tratamiento óptimo de los datos que se generan en cada una de las interacciones con sus clientes tanto en su página web como en sus redes sociales y es frecuente hallar retrasos o déficit de eficiencia al momento de recolectar solicitudes o en la atención que se proporciona al usuario y a la información desarrollada dentro del sistema PQR's; esta falta de eficiencia y retraso hace que los usuarios pierdan el interés y que la empresa no logre tener la información específica para que se tomen las mejores decisiones, además que, no se identifiquen nuevos mercados y por supuesto que la empresa no pueda ampliar su horizonte de mercado.

Figura 1

Árbol de problemas



Justificación

Con base al estudio desarrollado se evidencia que hoy en día muchas de las empresas tienen problemas al momento de tomar las mejores decisiones en base a la búsqueda de diferentes estrategias que les permitan ser más competitivas en el mercado, así como también que logren aumentar la innovación y a la vez generar cambios de manera constante en cada uno de los productos o servicios ofertados a los usuarios, adicional a esto las empresas también buscan un mejor manejo de las estrategias de comercialización y publicidad, y definir de una mejor manera las necesidades de los clientes; lo que representa una desventaja en las empresas del Ecuador.

Muchas de estas empresas poseen grandes cantidades de información almacenada de una manera incorrecta, lo que implica que estos datos no son aprovechados correctamente, esto con el fin de crear o mejorar las estrategias de negocio que contribuyan considerablemente a mejorar las ventas y a su vez ser más competitivas e innovadoras en el mercado nacional e internacional.

La implementación de diferentes métodos de procesamiento de datos (Data Mining) plantea un reto para lograr mejorar las estrategias de negocio de las empresas ya que aplicar cada una de estas técnicas nos permitirá tener un mejor manejo y análisis de los datos generados en las peticiones de cada uno de los consumidores con la ayuda de una variedad de medios de comunicación.

Objetivos

General

Implementar un asistente virtual basado en técnicas de minería de datos e Inteligencia Artificial mediante el uso de la tecnología Chatbot con el propósito de tomar decisiones sobre la información proporcionada por el cliente.

Objetivos Específicos

- Realizar un estudio de literatura para identificar los métodos, técnicas y tecnologías de minería de datos que influyen en el comportamiento y seguimiento de clientes, aceptación de productos y tendencias.
- Caracterizar las tecnologías de NPL (Natural Language Processing), al Machine learning usadas en Chatbots.
- Implementar un prototipo funcional para los canales conversacionales de la empresa SICE, de acuerdo a las necesidades de cada cliente, creando con esto un medio de comunicación más eficiente.
- Correlacionar métricas (Interacciones por usuario, tiempo de respuesta, tasas de retención, tasa de satisfacción, etc) previos a la implementación Chatbot.

Alcance

Para demarcar el alcance del proyecto de investigación a desarrollarse, el mismo se ha dividido en las siguientes fases:

1. Realizar una investigación exploratoria de la situación actual, en torno a las encuestas peticiones, quejas, reclamos y sugerencias realizadas por los usuarios.
 - a. Identificar las causas y efectos de la empresa SICE al no poseer un tratamiento óptimo de los datos que se generan en cada una de las interacciones con sus clientes.
 - b. Investigar cómo se realiza y cuánto tiempo se tarda la empresa en solventar la gestión de incidencias.
2. Obtener información en estudios relacionados sobre mejores y recientes formas de automatizar y mejorar los procesos mediante la Minería de Datos e Inteligencia Artificial.

3. Implementar un asistente virtual basado en técnicas de Minería de Datos e Inteligencia Artificial con el propósito de tomar decisiones respecto a la información que el cliente proporciona a la empresa SICE y gestionar las incidencias casi que, en tiempo real.
 - a. Extraer información de la base de datos de clientes:
 - i. Identificar las características comunes de los clientes.
 - b. Recolectar los datos:
 - i. Agrupar la información de los mensajes, peticiones y recomendaciones recibidos por la institución en las encuestas de satisfacción realizadas.
 - c. Caracterización de técnicas y tecnologías NPL (“Procesamiento del Lenguaje Natural)
 - i. Extraer la información necesaria para entender las técnicas NPL que más se ajusten a la necesidad de la empresa.
 - ii. Dialogflow (API.AI)
 1. La API se implementa para la comprensión del lenguaje de Google, llamado Dialogflow, genera varios grupos de respuestas para un mensaje definido, utiliza diferentes técnicas de NPL, que facilitan el análisis del texto recibido y desarrolla respuestas aleatorias basadas en las características. Google permite que los desarrolladores de Python utilicen esta API mediante la descarga de la biblioteca api.ai, que puede enviar diferentes textos y obtener respuestas, según la configuración proporcionada a través de la consola de administración.
 - d. Constituir el aprendizaje del sistema:
 - i. Obtención de los datos.
 - ii. Estandarización de la información obtenida
 - iii. Durante el proceso de vectorización, tomando en cuenta que la máquina de soporte de vectores no procesa texto, es primordial la aplicación del método

TF-IDF (Term frequency – Inverse document frequency), mediante el cual se representa la frecuencia (mensaje) de los términos del documento en forma numérica, en la categoría correspondiente para evaluar la importancia de las palabras de un documento dentro de una categoría.

- iv. Búsqueda de los parámetros óptimos del clasificador
- v. Proceso de entrenamiento y validación cruzada
 - 1. Usando el vector de características obtenido en el proceso TF-IDF, es fundamental muestrear el SVM para encontrar los parámetros óptimos del clasificador. Según los autores (RUIZ, 2017) y (FLORIAN NORIEGA, 2013), se aplica un kernel lineal porque su investigación es sobre la clasificación de texto.
- e. Implementación del Chatbot para la empresa SICE (Servicios Educativos).

Hipótesis

La implementación de un asistente virtual basado en metodologías de procesamiento de datos e IA mejorara la toma de decisiones sobre la información proporcionada por el cliente.

Capítulo II

Marco metodológico

Estado del Arte

Para la revisión de literatura de la presente investigación, se efectuó el mapeo sistemático haciendo referencia a las guías propuestas por Kitchenham, las mismas que se mencionan a continuación: (1) Planificación sistemática del estudio de la literatura; (2) Definición del grupo de control y extracción del término ;(3) Construir de la cadena de búsqueda ;(4) Selección de estudios primarios y finales; (5) Desarrollo del estado del arte (Kitchenham & Charters, 2007). Las fases mencionadas anteriormente se describen a continuación:

Planteamiento del estudio sistemático de literatura

En la etapa inicial del mapeo sistemático de la literatura, primero se describe la pregunta central del proyecto de investigación, luego se definen los objetivos de búsqueda y por supuesto la definición de cada pregunta de investigación detallada en la Tabla 4, finalmente se determinan los criterios de inclusión y exclusión, y se lleva a cabo la investigación preliminar.

Los criterios de inclusión y exclusión que se establecieron para la investigación son los siguientes:

Criterios de inclusión:

Artículos que se encuentren en el rango del 2015 al presente año, esto permite tener información actual.

Artículos en los que se detalle soluciones para la problemática en torno a la falta de estrategias de seguimiento a clientes y fidelización de los mismos.

Artículos en los que se mencione como mejorar las estrategias de marketing con respecto a la relación empresa cliente.

Criterios de exclusión

Artículos donde no se trate de soluciones aplicando minería de datos.

Artículos donde el enfoque no esté totalmente centrado en los detalles técnicos de una tecnología y no se aborde una solución al problema planteado.

Definición del grupo de control y extracción de términos

En la segunda fase, se identificó cada estudio que constituía el grupo control, el cual debía estar estrictamente relacionado con los criterios de inclusión y exclusión.

Después de analizar varios estudios científicos, se seleccionaron los artículos que constituyeron el grupo de control, como se muestra en la Tabla 1:

Tabla 1

Artículos que conforman el grupo de control

Título	Cita	Palabras clave
Managing Customer Relationships in the Social Media Era: Introducing the Social CRM House	(Edward C.Malthouse, MichaelHaenlein, BerndSkiera, EgbertWege, MichaelZhang, 2013)	Customer relationship management, Social media, Engagement, Information technology, Customer insight, Employees, Key performance indicator

Titulo	Cita	Palabras clave
Mining the network value of customers	(Domingos Pedro, Richardson Matt, 2001)	Applied computing, Law, social and behavioral sciences, Information systems, Information systems applications, Data mining
Data Mining Techniques: For Marketing, Sales, and Customer Relationship management	(Gordon Linoff, Michael J A Berry, 2011)	Business -- Data processing, Data mining, Marketing -- Data processing, Business & economics -- Green Business, Data Mining, Marketing, Kundenberatung, Marknadsföring – databehandling, Computers and IT
Application of data mining techniques in customer relationship management: A literature review and classification	(E.W.T. Ngai, LiXiu, D.C.K. Chau, 2008)	Data mining, Customer relationship management, Literature review, Classification
Customer Care Excellence: How to Create an Effective Customer Focus.	(Sarah Cook, 2010)	Customer relations, Customer services, Total quality management, Marketing & Sales,

Título	Cita	Palabras clave
Advanced Customer	(Brent Kitchens, David	Commerce, Business & Economics
Analytics: Strategic Value	Dobolyi, Jingjing	big data,
Through Integration of	Li & Ahmed Abbasi, 2018)	customer acquisition,
Relationship-Oriented Big		customer analytics,
Data		customer expansion,
		data integration,
		data management,
		design science,
		IT strategic value,
		relationship marketing,
		customer retention

Se llegó a conformar el grupo de control (GC) de seis artículos, cada artículo permitió la selección de las palabras claves para poder conformar la cadena de búsqueda.

Construcción de la cadena de búsqueda

Una vez definidos los términos claves en el grupo de control, en esta sección se genera las cadenas de búsqueda en la base digital seleccionada, para el caso de esta investigación la base digital seleccionada es IEEE XPLORE. La cadena de búsqueda tuvo varias versiones, en la cadena de búsqueda definitiva se encuentran los términos que van acorde con la presente investigación.

La primera y segunda cadena de búsqueda fueron descartadas porque se obtuvieron pocos artículos y la mayoría de ellos no presentaban una relación con la problemática en sí, por otro lado, la cadena número tres se excluyó, ya que no se encontraban todos los contextos en la cadena de búsqueda y la cantidad de palabras clave eran limitadas además de que también se obtuvo muy pocos resultados. Para finalizar, la cuarta versión de la cadena fue la final ya

que contemplaba todos los contextos y los resultados contenían las palabras claves además de que se obtuvo un número manejable de artículos, todas estas versiones de cadena de búsqueda están detalladas en la tabla 2.

Tabla 2

Versiones de la cadena de búsqueda

Versión	Cadena de búsqueda	# Resultados
1	(Customer relationship management OR Customer insight) AND (Information technology) AND (Social media)	57
2	(Social and behavioral sciences OR Customer insight) AND (Information systems applications) AND (Data mining)	66
3	(Social and behavioral sciences OR Customer insight) AND (Information systems applications) AND (Data mining OR Data processing) AND (Marketing)	35
4	(Customer insight OR social media) AND (Data mining OR Big Data) AND (relationship marketing OR Customer relationship management OR customer acquisition OR customer expansion OR customer retention)	124

Selección de los estudios primarios

La cadena de búsqueda retorno 124 artículos científicos candidatos dentro de la base digital IEEE XPLORE.

Para adaptar cada uno de los criterios de inclusión y exclusión, se realizó un minucioso análisis tanto de los títulos como de los resúmenes de los estudios candidatos a través de una validación cruzada entre tres investigadores. Una vez aplicado todos los criterios finalmente se obtuvieron 86 estudios relevantes, mismo que se procedió a descargar y leer para verificar que

cumplan con todos los criterios establecidos, luego de este análisis se obtuvieron finalmente 8 estudios primarios los mismos que fueron utilizados para desarrollar el estado del arte de la presente investigación y que se detallan en la tabla 3.

Tabla 3

Estudios primarios

Código	Título	Cita
EP1	Social Media Data Aggregation and Mining for Internet-Scale Customer Relationship Management	(Stephen Wan, Cecile Paris, Dimitrios Georgakopoulos, 2015)
EP2	Clustering and profiling of customers using RFM for customer relationship management recommendations	(Ina Maryani, Dwiza Riana, 2017)
EP3	Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study	(Mohammadreza Tavakoli, Mohammadreza Molavi, Vahid Mosoumi, Majid Mobini, Sadegh Etemad, Rouhollah Rahmani, 2018)
EP4	Market Basket Analysis Approach to Machine Learning	(Abu Hasnat Patwary, Md Tamim Eshan, Prazzal Debnath, Abdus Sattar, 2021)
EP5	A Cloud-Based System for Improving Retention Marketing Loyalty Programs in Industry 4.0: A Study on Big Data Storage Implications	(Antonino Galletta, Lorenzo Carnevale, Antonio Celesti, Maria Fazio, Massimo Villari, 2017)

EP6	Analyzing Customer Journey with Process Mining: From Discovery to Recommendations	(Alessandro Terragni, Marwan Hassani, 2018)
EP7	Survey on Customer Centric Sales Analysis and Prediction	(B. Ida Seraphim, Lavi Samuel Rao, Shiwani Joshi, 2018)
EP8	Social CRM using web mining for Indonesian academic institution	(Nyoman Karna, Iping Supriana, Nur Maulidevi, 2015)

Elaboración del estado del arte

EP1(Stephen Wan, Cecile Paris, Dimitrios Georgakopoulos, 2015): **Social Media Data Aggregation and Mining for Internet-Scale Customer Relationship Management**

Este artículo está enfocado en la explotación de las redes sociales para CRM, y describen el desarrollo y la funcionalidad de VIZIE – que es un sistema de control de redes sociales diseñado para ayudar al monitoreo de medios a destilar las percepciones reales y potenciales de los clientes sobre el valor de los productos y servicios existentes, identificar y corregir la información errónea de los clientes, descubrir la mejor manera de cumplir con las expectativas de los clientes y (si es necesario) a través de la implementación de las redes sociales preferidas del cliente aplicar un foro de medios. Además, se proporciona una visión general de los requisitos del mundo real que informaron el diseño de este sistema que ha sido utilizado por más de veinte organizaciones y describimos un enfoque general de su arquitectura, los métodos de análisis de datos empleados en su implementación y su

orquestración para lograr la respuesta en tiempo real en un ámbito de análisis de datos sociales de gran nivel de Internet.

EP2(Ina Maryani, Dwiza Riana, 2017): Clustering and profiling of customers using RFM for customer relationship management recommendations

En este trabajo de investigación se aborda en primera instancia los problemas que las empresas enfrentan en la actualidad al momento de identificar sus clientes potenciales, así como también aplicar de una manera correcta la “Gestión de relación con el cliente - CRM” con el fin de poder llevar a cabo una correcta estrategia de marketing. El estudio tiene como finalidad realizar una agrupación dependiendo del perfil del cliente mediante el uso del modelo de Frecuencia de Recencia y Monetaria (RFM) con el fin de proporcionar recomendaciones CRM a la empresa industrial media.

Se llevan a cabo una serie de pasos: primero se empieza con la minería de datos a partir de datos históricos de las ventas, segundo se procede a realizar el modelado de minería de datos usando RFM con el algoritmo K-Means, tercero se realiza la clasificación de clientes usando arboles de decisiones y se establece el nivel de lealtad del cliente y las recomendaciones de CMR.

EP3(Mohammadreza Tavakoli, Mohammadreza Molavi, Vahid Mosoumi, Majid Mobini, Sadegh Etemad, Rouhollah Rahmani, 2018): Customer Segmentation and Strategy Development Based on User Behavior Analysis, RFM Model and Data Mining Techniques: A Case Study

El propósito es analizar efectivamente las decisiones con el fin de dirigir estrategias de marketing adecuadas de acuerdo con el comportamiento anterior de los clientes, en este trabajo se propone un modelo RFM que configura la segmentación según los cambios de negocio y agrupa al cliente de la empresa usando K-Means. Además, se construye estrategias para cada uno de los segmentos y se ejecuta una campaña de SMS de acuerdo con las

estrategias encontradas en el modelo. Los resultados de la campaña mostraron que nuestro Modelo de Segmentación mejoró el número de compras y el promedio monetario de las canastas.

EP4 (Abu Hasnat Patwary, Md Tamim Eshan, Prazzal Debnath, Abdus Sattar, 2021):

Market Basket Analysis Approach to Machine Learning

Está centrado en el análisis de la cesta de compra y en el desarrollo de promociones de ventas para varios segmentos de consumidores para aumentar la lealtad del cliente y como resultado el beneficio. Se busca que los comerciantes puedan mejorar aún más su negocio, al mismo tiempo puedan generar muchos más ingresos económicos. Los conjuntos de elementos frecuentes son extraídos de una base de datos haciendo uso del algoritmo Apriori para finalmente generar las reglas de asociación que nos servirán para ver el comportamiento que tienen los clientes en cada una de las ventas.

EP5 (Antonino Galletta, Lorenzo Carnevale, Antonio Celesti, Maria Fazio, Massimo Villari, 2017): **A Cloud-Based System for Improving Retention Marketing Loyalty Programs in Industry 4.0: A Study on Big Data Storage Implications**

En este artículo se hace énfasis en cada una de las estrategias de comercialización de "retención" dirigidas no sólo a la adquisición de nuevos clientes, sino también a la rentabilidad de los existentes, todo esto permite a las industrias aplicar estrategias de producción específicas para maximizar sus ingresos. Esto es posible mediante el análisis de diversos tipos de información procedente de clientes, productos, compras, etc. En particular, se propone un software basado en la nube como una arquitectura de servicio que almacena y analiza big data relacionados con las compras y las filas de productos con el fin de proporcionar a los clientes una lista de productos recomendados. Los experimentos se centran en un prototipo de flujo de trabajo humano a máquina para la preselección de clientes implementados en escenarios de nube privada e híbrida.

EP6 (Alessandro Terragni, Marwan Hassani, 2018): Analyzing Customer Journey with Process Mining: From Discovery to Recommendations

Se realiza un análisis del recorrido del cliente y como este es un tema candente en marketing. Entender cómo se comportan los clientes es crucial y se toma en consideración como uno de los factores principales del éxito empresarial. En este documento se contribuye a un enfoque novedoso para aplicar técnicas de minería de procesos para el análisis del recorrido del cliente de registro web. A través de la minería de procesos se puede (i) descubrir el proceso que describe mejor el comportamiento del usuario, (ii) encontrar información útil, (iii) comparar los procesos de diferentes clústeres de usuarios, y luego (iv) utilizar este análisis para mejorar el recorrido mediante la optimización de algunos KPI (Indicadores clave de rendimiento) a través de recomendaciones personalizadas basadas en el comportamiento del usuario. Finalmente se muestra a través de un caso práctico de la vida real una prueba de la exactitud del concepto introducido mejorando la precisión del recomendado al incorporar información de contexto adicional sobre el viaje tal como se extrae del modelo de proceso.

EP7 (B. Ida Seraphim, Lavi Samuel Rao, Shiwani Joshi, 2018): Survey on Customer Centric Sales Analysis and Prediction

Este artículo aborda inicialmente las técnicas y algoritmos de minería de datos usadas en la predicción de ventas. Pero para predecir estas ventas hay un requisito de datos de clientes que nos den una idea de cómo el cliente percibe los productos o servicios que se ofrecen. En este artículo se trata de comparar varias técnicas de minería de datos y algoritmos utilizados en el pasado para predecir estas ventas, y luego se trata de proponer mejores maneras de resolver el mismo problema. Este documento también compara la eficiencia de las técnicas ANN en comparación con la técnica de minería de datos. Hasta ahora, se han realizado una serie de mejoras y logros en este campo mediante el análisis de cesta de la compra.

EP8 (Nyoman Karna, Iping Supriana, Nur Maulidevi, 2015): Social CRM using web mining for Indonesian academic institution

Se evidencia un modelo y aplicación de la extensión de la minería de redes de CRM social, principalmente relacionadas con instituciones académicas en Indonesia. El modelo y los elementos de minería de datos web pueden enfocar las estrategias de marketing al segmentar a los candidatos potenciales.

Características del estado del arte

Existe variedad de estudios que se enfocan básicamente en el análisis de Big Data, y cuán importante es este para mejorar la relación con el cliente, en su mayoría los artículos encontrados plantean las estrategias de marketing que se deben aplicar luego de analizar la gran cantidad de datos que los mismos clientes generan al momento de hacer contacto con una determinada empresa, por otro lado también hay artículos que recomiendan la aplicación de algoritmos de minería de datos que nos permitan segmentar de una mejor manera a los clientes. La mayoría de artículos mencionados anteriormente brindan algunos métodos y técnicas de minería de datos que se pueden aplicar hoy en día en las empresas para mejorar sustancialmente la relación con los clientes, así mismo se identificó que no se hace uso de ninguna herramienta que permita extraer datos de cada uno de los medios o canales conversacionales de una manera fácil, centralizada y que sobre todo esté disponible las 24 horas del día, ofreciendo respuesta a todas las inquietudes de los usuarios. Esta dificultad lleva a proponer un agente conversacional (Chatbot) basado en minería de datos e inteligencia artificial que permita a la empresa conocer el comportamiento, los deseos, quejas y sugerencias de cada uno de sus clientes que visitan la página web y redes sociales, de esta manera se mejorará la toma de decisiones y mejor tiempo y calidad de respuesta al usuario.

Metodología

En la presente investigación se hace uso de la metodología Cross Industry Standard Process for Data Mining, también conocida por sus siglas CRISP-DM, la misma que se encarga de abarcar las tareas, subtareas y etapas por desarrollar en la presente investigación, a breves rasgos se puede mencionar que el ciclo de vida para un proyecto donde interviene la minería de datos consta de seis etapas o fases, mismas que son descritas a continuación (Bellini Saibene & Volpacchio, 2014):

Fase 1. Comprensión del giro del negocio

Dentro de esta fase se contempla la comprensión total de los objetivos que son detallados en el proyecto, además se centra en la creación de un plan preliminar que se encuentre diseñado para cumplir con todas las metas que se han propuesto.

Fase 2. Estudio y comprensión de los datos

Iniciación de la recopilación de los datos que luego van hacer analizados, continuando con las tareas que admite la familiarización por parte de los investigadores. Permite por una parte obtener data que puede ser considerada como conocimiento y por otra la manera de llegar a mostrar diferentes conjuntos interesantes que estén ocultos dentro de la data.

Fase 3. Análisis de los datos y selección de las características

El estudio se centra en el tratamiento que se les da a los datos de entrada, para que luego esta se convierta en información útil y necesaria, y que sea capaz de construir un conjunto que ingrese al análisis de una manera definitiva, es decir datos que sirvan para la toma de decisiones en base al conocimiento.

Fase 4. Modelado

Selección y aplicación de métodos que son necesarios para dar solución al problema identificado, así mismo como la mejora de toma de decisiones para la empresa SICE. En esta

fase de debe tener en cuenta que es posible volver a la fase tres de la metodología ya que en ocasiones es necesario recurrir a datos que en otros análisis podrían ser irrelevantes.

Fase 5. Evaluación

En esta fase del proyecto y luego de haber construido diferentes modelos con la capacidad de solventar el problema de la mejora de toma de decisiones con respecto a los clientes de la empresa SICE, por lo que es imprescindible realizar una evaluación minuciosa así mismo verificar cada uno de los pasos seguidos para llegar a cada uno de los modelos obtenidos durante todo el análisis posterior.

Fase 6 Despliegue o Implementación

En esta última fase se intenta crear un nuevo conocimiento acerca de la data de entrada la misma que ha sido analizada y sometida a varios tipos de modelos, por lo que toda esta data tendrá que obligatoriamente ser evaluada una vez finalizado el análisis de una manera completa, durante esa evaluación se deberá verificar cada uno de los pasos ejecutados antes de la culminación del proyecto.

Marco Teórico

Inteligencia Artificial

La inteligencia artificial (IA) es actualmente uno de los componentes más importantes implementados por las organizaciones dentro de los países desarrollados y está penetrando poco a poco en las naciones en via de desarrollo. Las primeras ideas sobre la inteligencia artificial aparecieron en la segunda mitad del siglo XX, cuando, desde la ciencia ficción, se creía que en un futuro próximo existirían robots antropomórficos para realizar todas las actividades que realizan los humanos. La realidad es que las cosas son un poco diferentes hoy de lo que eran entonces. Por ejemplo, los robots diseñados y alimentados por inteligencia artificial pueden superar con creces las capacidades humanas en problemas técnicos como las

matemáticas o el ajedrez, sin embargo, se limitan a actividades puntuales y monótonas (Lu, Li, Chen, Kim y Serikawa, 2018).

En informática, el concepto de inteligencia artificial fue propuesto por McCarthy, Minsky, Rochester y Shannon en 1955; se refiere a la imitación de las actividades humanas por parte de algún tipo de sistema informático, sin embargo, existe un debate a partir desde que se planteó si la inteligencia artificial tiene la capacidad de superar las capacidades humanas, lo que pone en entredicho el primer planteamiento. Otra definición sostiene que la inteligencia artificial es "la capacidad de un sistema para interpretar datos externos, aprender de ellos y usarlos de manera flexible para lograr objetivos específicos" (Kaplan y Haenelin, 2018: 3).

Una tercera definición ve a la IA como "la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran los humanos" (Philip C, 2019).

Aplicaciones de la IA

El campo de la inteligencia artificial se ha desarrollado rápidamente en los últimos años. Los tres principales desencadenantes de este fenómeno de acuerdo con (Brynjolfsson, 2017) son: el desarrollo de algoritmos y circuitos electrónicos especializados (con mayor poder de procesamiento), el crecimiento de los datos disponibles y el aumento de los recursos humanos y financieros destinados a su desarrollo. Se estima que el mercado global de inteligencia artificial tendrá un valor de casi \$ 126 mil millones para el año 2050 (Bughin, y otros, 2017).

- Reconocimiento visual: Tiene la capacidad de procesar imágenes y videos para reconocer y rastrear objetos y personas.
- Reconocimiento del lenguaje natural: Reconocen, reproducen de manera artificial y descifran el lenguaje hablado. La traducción automática se hace posible en diferentes idiomas.
- Estrategia y planeación: Generan estrategias aptas para la resolución de problemas complejos, son capaces de brindar apoyo en logística y manufactura, videojuegos o navegar a través de espacios físicos.

- Diagnóstico y apoyo en la toma de decisiones: Analiza problemas complejos y sirve de apoyo al tomar decisiones, por ejemplo, en medicina, en la detección de enfermedades o la elección del tratamiento más adecuado
- Colaboración humano-computadora: consiste en integrar sistemas inteligentes en los equipos de trabajo humano. Por ejemplo, se han desarrollado sistemas que pueden analizar vistas aéreas de las áreas afectadas para determinar dónde se necesita más asistencia para ayudarnos a responder a los desastres naturales más rápidamente.

Técnicas de la Inteligencia Artificial

Las más destacadas son (Pajares Martinsanz & Santos Peñas, 2005):

- Sistemas Expertos
- Redes neuronales
- Algoritmos Genéticos
- Lógica Difusa

La combinación de estas técnicas es aplicada al para generar soluciones óptimas del problema en estudio.

Sistemas expertos

Son las herramientas de inteligencia artificial más utilizadas desde sus inicios, y corresponden a programas informáticos que recopilan el conocimiento de los expertos en un programa informático.

Redes Neuronales

Sustenta la premisa de que los secretos del aprendizaje y del conocimiento residen en axiomas o verdades incuestionables, que el conocimiento es independiente de la estructura que rige los símbolos, que la representación del conocimiento proviene del nivel más básico de

la inteligencia: el cerebro en particular y sus neuronas y las múltiples interconexiones que existen entre ellas. La red neuronal artificial constituye una tecnología de procesamiento de información paralela a gran escala que simula las características básicas de la estructura neuronal del cerebro biológico. Si bien las redes neuronales artificiales tienen limitaciones para representar todas las características del cerebro humano, tales como: desarrollar la capacidad de aprendizaje adaptativo, autoorganización, tolerancia a fallas, operación en tiempo real y otras cualidades, estas redes constituyen una poderosa herramienta tecnológica para procesamiento de información cuyos resultados pueden tomar decisiones efectivas y oportunas.

Algoritmos Genéticos

Son el resultado de los recientes avances en computación evolutiva y genética, y son una de las principales herramientas tecnológicas de la inteligencia artificial. Estos algoritmos utilizan información histórica para simular los mecanismos de la selección natural y la genética para encontrar nuevos puntos de búsqueda de las mejores soluciones para obtener soluciones a problemas que no tienen una solución precisa debido a su complejidad.

Lógica Difusa

Basada en la noción de que todo es una cuestión de grado, la lógica difusa permite el manejo de información ambigua o difícil de especificar importante para la resolución de problemas a través de una serie de reglas de "sentido común" aprendidas por un sistema informático. Nutrir a través de observaciones humanas o recetas de expertos humanos.

Dado que la certeza propuesta es una cuestión de grado y, como resultado, forma parte de la base del razonamiento aproximado en lugar del razonamiento exacto como en la lógica clásica, la característica central de las técnicas de lógica dispersa es su capacidad para reproducir los hábitos de razonamiento humano de una manera aceptable y eficaz.

Aprendizaje Automático (Machine Learning)

El aprendizaje automático se plantea la interrogante de cómo construir computadoras que a través de la experiencia sean capaces de evolucionar. Es uno de los campos tecnológicos de más rápido crecimiento en la actualidad, en la intersección de la informática y la estadística y en el corazón de la inteligencia artificial y la ciencia de datos. Los avances recientes en el aprendizaje automático han sido impulsados por el desarrollo de nuevos algoritmos y teorías de aprendizaje y por la continua explosión en la disponibilidad de datos en línea y computación de bajo costo (Jordan & Mitchell, 2015).

Sistemas de recomendación

Son programas informáticos que tienen como objetivo el posibilitar a los clientes un acceso mucho más rápido a sus distintos intereses, estos sistemas trabajan de tres maneras (Ospina Quintero, 2015):

- A partir de los aspectos de búsqueda que el usuario ha utilizado previamente, es posible determinar qué complementos se pueden realizar a sus intereses.
- A través del aspecto de búsqueda colaborativa, se pueden establecer sugerencias de información mediante la participación de usuarios que ejecutan procedimientos o consultas similares.
- Al mezclar los dos primeros aspectos, se puede obtener información precisa sobre las necesidades del usuario.
- Los sistemas basados en recomendaciones facilitan la toma de decisiones y optimizan los procesos de la empresa para brindar a los usuarios lo que necesitan en tiempo y forma.

Implementar estos sistemas en la gestión de PQR significa un gran apoyo para que las empresas demuestren de manera rápida y eficiente los procesos que los clientes necesitan, esto

por la información obtenida a través de solicitudes previas del usuario, o solicitudes similares de otros usuarios.

Aprendizaje automático

Estos algoritmos están diseñados para posibilitar el desarrollo de sistemas informáticos, robots, etc., para la toma de decisiones racionales previa formación (TECHTARGET, 2017).

Al igual que la minería de datos, estos algoritmos prueban en recopilar información o datos del pasado con el fin de poder realizar predicciones, se clasifican en supervisados y no supervisados. Capacidad para adaptarse al contexto actual o crear inferencias a partir de conjuntos de datos individualmente basados en información pasada (TECHTARGET, 2017).

Tipos de aprendizaje automático

- Aprendizaje inductivo: redes neuronales artificiales diseñadas para identificar una descripción para las combinaciones de conjuntos de datos.
- Aprendizaje analítico: Uso de probabilidades que preceden la explicación de un ejemplo, se logra establecer una relación entre la causa y el efecto de una situación establecida.
- Aprendizaje genético: Se pueden encontrar diferentes soluciones a las combinaciones de datos presentados cuando se usan algoritmos que recrean varias teorías evolutivas.
- Aprendizaje conexionista: redes neuronales artificiales diseñadas para encontrar una descripción para las combinaciones de conjuntos de datos.

Los algoritmos de aprendizaje automático permiten determinar rápidamente posibles soluciones a situaciones nunca antes vistas en función del entrenamiento previo. Al aplicarlo al sistema de gestión de PQR, es posible determinar qué procesos o áreas responden a los eventos de los usuarios. Esto es gracias a un ejercicio preliminar del sistema para comprender cada uno de los procesos y PQRs que se originan con una mayor frecuencia para cada zona

de la organización. La información que se puede categorizar de manera ágil y exitosa, expresada de formas inusuales, cada vez es más fácil.

Minería de Datos

La revolución digital hace realidad la recopilación sencilla de la información digitalizada para que sea procesada, almacenada, distribuida, y transmitida (Mitra & Acharya, 2003).

La demanda de la tecnología de Internet va incrementando y por consiguiente es necesario desarrollar una metodología de minería de datos más avanzadas de tal manera que sea posible procesar la información y analizar los datos distribuidos por todo el mundo (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004).

La escasez de metodologías para el análisis inteligente de datos, las cuales puedan descubrir conocimiento útil de los datos. El término KDD (iniciales de Knowledge Discovery in Databases), creados en 1989 plantea que todo el proceso de extracción de conocimiento de las bases de datos, marcando un cambio de paradigma donde es importante que seamos capaces de descubrir conocimiento útil a partir de los datos. El primer estado del arte en la región, [Fayy96] se dice (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004):

"La mayor parte del trabajo anterior sobre KDD se ha centrado en la [...] fase de extracción de datos. Sin embargo, otros pasos son muy importantes para el éxito de las aplicaciones de KDD en la práctica".

Esto muestra claramente la importancia de incluir el preprocesamiento de datos o la formalización del conocimiento por descubrimiento en la metodología.

La extracción de conocimiento de base de datos y minería de datos (KDD) se define como el proceso de identificar patrones significativos en datos que son válidos, llamativos, útiles y entendibles para los usuarios (Gordon S & Berry, 2011) (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004) (Mitra & Acharya, 2003) (Tang & MacLennan, 2005). El

proceso general implica convertir los conocimientos básicos en conocimientos avanzados. El proceso KDD es interactivo e iterativo y consta de los siguientes pasos:

1. Comprender el dominio de la aplicación: Conocimiento previo relevante y los objetivos de la aplicación.
2. Extracción de la base de datos de destino: recopilación, evaluación, calidad y familiaridad de los datos con el análisis exploratorio.
3. Preparación de datos: Limpieza, transformación, integración y reducción de datos. La finalidad es pulir los datos reduciendo el tiempo para los algoritmos de aprendizaje aplicados posteriormente.
4. Minería de datos: Consiste en la clasificación, regresión, agrupamiento, agregación, recuperación de imágenes, extracción de reglas, etc.
5. Interpretación: Explicar los patrones encontrados y la posibilidad de visualizarlos.
6. Utilice el conocimiento descubierto: Utilización provechosa del modelo creado.

Aplicaciones de la minería de datos

Las tareas principales de la minería de datos incluyen la identificación de aplicaciones de tecnologías y el desarrollo de nuevas tecnologías para áreas de aplicación tradicionales o nuevas, como el comercio electrónico y la bioinformática. Hay muchas áreas donde se puede aplicar la minería de datos, casi en todas las actividades humanas que generan datos (Hernández et al., 2004).

- Comercio y Banca: Segmentación de clientes, proyección de ventas, análisis de riesgo.
- Medicina y Farmacia: Efectividad en el Diagnóstico y Tratamiento de la Enfermedad.
- Seguridad y detección de fraude: Reconocimiento facial, biometría, acceso no autorizado a la red y más.
- Recuperación de información no numérica: Minería de texto y web, búsqueda de imágenes, videos, voz y texto y reconocimiento de bases de datos multimedia.

- Astronomía: Identificación de nuevas estrellas y galaxias.
- Geología, Minería, Agricultura y Pesca: Identifica áreas de uso para diferentes cultivos o extracción pesquera o minera en bases de datos de imágenes satelitales.
- Ciencias Ambientales: Modelos funcionales de ecosistemas naturales y/o artificiales (ej. plantas de tratamiento de aguas residuales) para mejorar su observación, manejo y/o control.
- Ciencias sociales: un estudio de la movilidad de la opinión pública.
- Urbanismo: Barrios en conflicto a partir de valores sociodemográficos.

Técnicas usadas por la minería de datos La Minería de Datos

Se puede abstraer en la construcción de un modelo que se ajusta en base a unos datos, aportando conocimiento. Se establece dos pasos principales en la tarea MD, la selección del modelo por un lado y el ajuste final a los datos. La elección del modelo se desarrolla básicamente por dos condiciones: el tipo de datos y el objetivo a conseguir. Así, por ejemplo, no es apropiado aplicar regresión a datos que consisten en texto o modelos basados en distancia a datos simbólicos. En cuanto a la relación del modelo objetivo, la literatura propone un catálogo de diferentes modelos para diferentes objetivos. Entonces, si hay un problema de clasificación, usará una máquina de vectores de soporte o un árbol de decisión, si es un problema de regresión, puede usar un árbol de regresión o una red neuronal, si necesita agrupamiento, puede usar un modelo jerárquico o de correlación, etc. También es importante en esta elección el nivel de comprensibilidad que desea del modelo final, ya que algunos modelos son fáciles de "explicar" al usuario, como las reglas de asociación, mientras que otros implican dificultades obvias, como las redes neuronales o el vector de soporte (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004).

Machine Learning

El aprendizaje automático (Machine learning) es una rama de la inteligencia artificial que se encarga de desarrollar una serie de algoritmos que permiten a las máquinas aprender por sí mismas a través de los datos, es decir, buscando patrones implícitos en ellos y sabiendo que están estructurados, todo con el fin de crear un nuevo conocimiento para tomar decisiones sobre un nuevo conjunto de datos. En pocas palabras, el aprendizaje automático permite que una máquina o un programa informático tenga inteligencia artificial (Alpaydin, 2014).

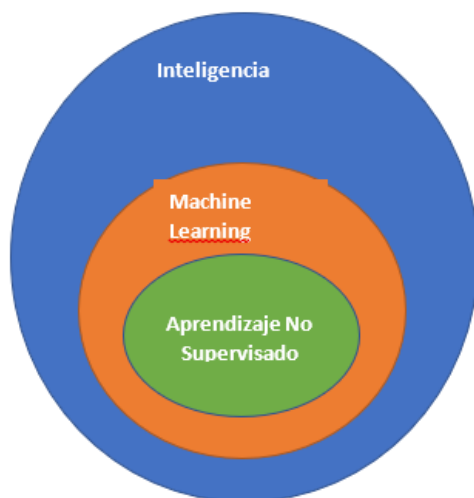
El aprendizaje automático se divide en tres categorías: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. El enfoque de este estudio es el aprendizaje no supervisado.

Aprendizaje no supervisado

Se llama aprendizaje no supervisado porque no tiene en cuenta los datos de entrada, y en eso se diferencia del aprendizaje supervisado. Este tipo de aprendizaje se utiliza directamente para la abstracción y compresión de patrones de datos. La figura 2 muestra que el aprendizaje no supervisado es aprendizaje automático, que a su vez es inteligencia artificial.

Figura 2

Aprendizaje no supervisado



Nota. La figura representa como está relacionado el aprendizaje no supervisado tanto con el machine learning como con la inteligencia artificial. Elaboración propia.

El aprendizaje no supervisado puede ser de dos tipos:

- **Reglas de Asociación:** Se utiliza básicamente para establecer posibles relaciones entre distintas acciones aparentemente independientes, tratando de reconocer cómo la ocurrencia de un determinado evento o acción produce la aparición de otros. Utilícelos cuando el objetivo sea un análisis exploratorio para encontrar relaciones en grandes conjuntos de datos (Molina López & García Herrero, 2006).
- **Clustering:** Permite identificar tipos o grupos donde cada elemento tiene similitudes y diferencias significativas con otros grupos (Miguel, Cuadrado, Sicilia, Rodríguez, & Rejas, 2007). Algunos ejemplos obvios son la segmentación de un conjunto de clientes, la suma de valores y métricas financieras, la recolección de áreas forestales dentro de un país, la recolección de sucursales y sus empleados, etc. (Jindal & Kharb, 2013). La agrupación en clústeres, también conocida como partición de datos, tiene una amplia variedad de objetivos relacionados con agrupar o dividir un conjunto de objetos en pequeños conjuntos o clústeres, donde los objetos dentro de cada clúster están más relacionados que los objetos asignados a diferentes clústeres (Guojun, Chaoqun, & Jianhong, 2007) (Halkidi, Batistakis, & Vazirg, 2001).

El estudio se desarrollará usando el método de Clustering para lo que es la agrupación de clientes de la empresa y reglas de asociación para obtener diversas reglas en base a las transacciones de la empresa SICE.

Algoritmos de Clustering

Algoritmos Jerárquicos

Este tipo de métodos tienen un enfoque que se desarrolla a partir de la construcción de un árbol o dendograma, los elementos del conjunto de ejemplos se representan por medio de las hojas y los nodos hacen referencia a los subconjuntos de ejemplos que pueden ser usados como particionamiento del espacio (Prakash & Aarohi, 2015). Existen varios algoritmos jerárquicos como, por ejemplo: AGNES (Aglomerative NESTing), DIANA (Divisia ANALysis), CURE (Clustering Using Representatives), CHAMALEON, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchical) (Sudipto, Rajeev, & Kyuseok, 2001) (Andritsos, 2002).

Existen dos métodos al momento de construir el árbol:

- Aglomerativos: El árbol se construye a partir de cada hoja que lo compone, hasta llegar finalmente a la raíz. En el primer caso, cada ejemplo es nuevamente un grupo, y estos conjuntos juntos forman más y más conjuntos hasta que finalmente se llega a la raíz (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004).
- Desaglomerativos o divisivos: Por otro lado, estos comienzan en la raíz, que es un solo grupo que contiene todos los ejemplos, y luego se dividen gradualmente hasta llegar a las hojas, que son representaciones de la situación en cada ejemplo (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004).

Algoritmos de particionamiento

Estos algoritmos están diseñados para la clasificación de individuos principalmente en K grupos. El proceso que se lleva a cabo es primeramente elegir una partición de los individuos en K grupos, luego se procede a intercambiar los miembros de los clústeres para así obtener una mejor partición (Mythili & Madhiya, 2014) (Pandey & Dubey, 2013). Entre los algoritmos de este tipo tenemos, por ejemplo: k-means (Hernández Orallo, Ramírez Quintana , & Ferri Ramírez, 2004), k-medoids (Kaufman & Rousseeuw, Finding Groups in Data: An Introduction

To Cluster Analysis, 1990), EM (Expectation Maximization) (Tsai, Hu, & Lu, 2013), Self-Organizing Maps SOM.

K-Means

Es implementado en la mayoría de los casos en la actualidad y es adecuado para situaciones en las que las variables involucradas son cuantitativas, y la distancia euclidiana generalmente se elige como la medida de disimilitud (Tsai, Hu, & Lu, 2013). El algoritmo K-means es un método de agrupación de vecindarios que comienza con un número determinado de prototipos y un conjunto de ejemplos, que se pueden recopilar sin etiquetarlos. Es el más popular de todos los métodos de agrupación en clústeres de partición, a diferencia de los métodos de agrupación en clústeres jerárquicos.

La idea principal de este algoritmo es ubicar cada prototipo en el espacio para que los datos pertenecientes al mismo tipo tengan características iguales o similares (Hernández Orallo, Ramírez Quintana, & Ferri Ramírez, 2004). El método puede ser descrito por medio del siguiente algoritmo (Pandey & Shukla, 2015):

1. Divida el conjunto de elementos en K grupos.
2. Calcula la distancia euclidiana de cada elemento a los k centros, luego asígnalos al grupo con el centro más cercano. Luego, vuelva a calcular los nuevos centroides después de cada asignación de un nuevo elemento al grupo al que va y al grupo al que llega.
3. Definir criterios óptimos y comprobar si la nueva reasignación puede mejorarlos. En este caso particular, se vuelve al paso 2.

K-Medoids

Este tipo de algoritmo es con base a agrupamientos de partición que se convierten ligeramente del algoritmo K-Means (Kaufman & Rousseeuw, Finding Groups in Data: An Introduction To Cluster Analysis, 1990). El algoritmo elige la media como centroide, pero en K-Medoids, los puntos de datos sin procesar se eligen como medoids. Un punto central se define como un objeto en un grupo que tiene una pequeña diferencia promedio de otros objetos en el

grupo (Jin & Han, 2011). Cada objeto restante se agrupa con el punto central más cercano, y estos algoritmos iterativamente hacen todos los intercambios posibles entre el objeto representativo y el objeto no representativo, todo hasta la suma de k-medoids y k-medoids La medida de la diferencia entre ellos se reduce. Los vectores de observación que componen el conglomerado. En este grupo se pueden encontrar varios algoritmos, tales como: PAM (particionamiento alrededor de Medoids) y CLARA (agrupamiento de grandes aplicaciones) (Kaufman & Rousseeuw, Finding Groups in Data: An Introduction To Cluster Analysis, 1990).

Reglas de Asociación

Este tipo de algoritmos buscan principalmente encontrar relaciones dentro de grupo extenso de transacciones, entendiendo como transacciones a cada uno de los grupos de eventos que se asocian de alguna u otra forma. Cada uno de los elementos que conforman una transacción se los denomina “item” y a un conjunto de ítems se denomina “itemset”. Cada una de las transacciones analizada puede tener uno o varios ítems, en este caso específico cada posible subconjunto es un “itemset” distinto.

Apriori

Este algoritmo es el uno de los más usados y uno de los primeros en ser creado y que tiene como objetivo realizar la búsqueda de reglas de asociación, este algoritmo tiene dos etapas bien marcadas:

1. La identificación de itemsets que se dan con frecuencia sobre un límite determinado de itemsets frecuentes.
2. Se convierte cada uno de los itemsets encontrados en la etapa anterior a una regla de asociación.

Procesamiento de lenguaje natural

Natural Language Processing (NLP) “es la manera por la cual las máquinas pueden analizar, comprender y obtener significado del lenguaje humano” (Icapps, 2017). Se usa para la

descripción de capacidades de las máquinas para comprender lo que se dice, desglosarlo y comprender y determinar su significado y, en última instancia, responder en un idioma que el usuario pueda entender. Este sistema se basa en siete pasos que se detallan a continuación.

1. Tokenización

El primer paso que se debe hacer antes de procesar cualquier información escrita es dividir el texto en palabras y oraciones para facilitar el análisis. La tokenización es preprocesamiento, y aunque este paso pueda parecer básico, debe hacerse precisamente para que el resto del análisis sea lo más relevante posible.

2. Análisis Léxico

Cuando la realización del proceso de tokenización es exitoso y las palabras están debidamente separadas la una de la otra, se procede a clasificar cada token para procesarlas de forma rápida. La figura 3 muestra el proceso de tokenización de una oración en sí, que visualiza cómo se divide en artículos, adjetivos, sustantivos, adverbios y verbos.

Figura 3

Análisis Léxico



Nota. La figura representa el análisis léxico que se da en la tokenización. La lingüística detrás de los Chatbots. ICapps (2017).

Primero, es conveniente que la coincidencia entre una palabra y sus propiedades gramaticales se dé fácilmente en un diccionario. Dado que una gramática es un conjunto de

reglas que gobiernan un idioma, la clasificación de tokens utilizando criterios gramaticales ayuda en los pasos posteriores, especialmente en el paso de análisis sintáctico.

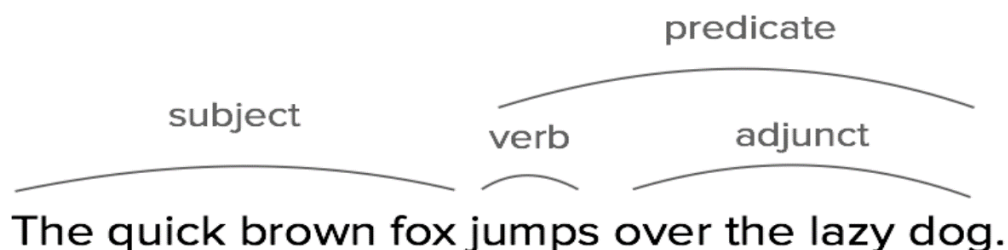
3. Análisis Sintáctico

Si bien los pasos anteriores ocurren a nivel de palabra, el análisis salta al nivel de oración para identificar la relación entre cada palabra.

La figura 4 muestra la descomposición que se produce en las oraciones a nivel sintáctico.

Figura 4

Análisis sintáctico



Nota. La figura representa el análisis sintáctico en una oración. La lingüística detrás de los Chatbots. ICapps (2017).

Dentro del análisis sintáctico se determina el orden y estructura de cada oración. La identificación del sujeto es importante para uno de los siguientes pasos.

4. Análisis Semántico

La computadora busca lo que significa cada una de las palabras, lo que para un humano puede ser una cosa sencilla para una computadora es un poco más complicado. Algunas de las palabras pueden resultar sencillas, y por lo tanto fáciles de interpretar.

5. Integración del discurso

Analiza el significado de las oraciones, para poder lograr esto los pronombres deben reconocerse bien, para luego vincularlos al antecedente relevante.

6. Análisis Pragmático

La pragmática se conoce como el estudio de las variaciones del contexto que influye en el significado. Este paso es el más difíciles de los seis pasos ya que trata de interpretar lo que se dice y que quiere decir, la ambigüedad es difícil de manejar a nivel de las máquinas, dependiendo del idioma y situación el contexto puede variar.

En el procesamiento de lenguaje natural (NLP) se pueden calcular dos tipos de diferentes conversaciones.

7. Conversaciones basadas en intenciones

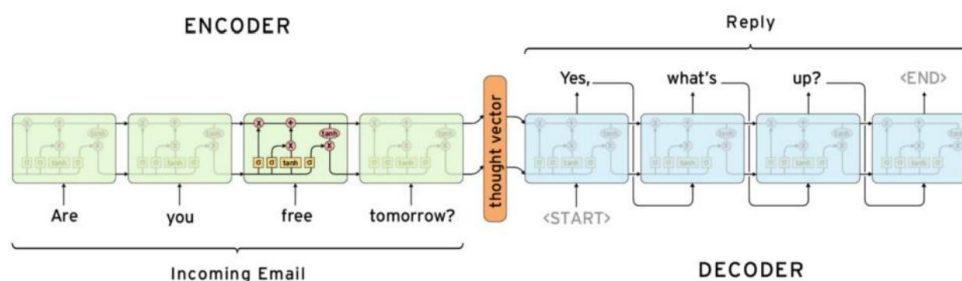
En este punto los algoritmos de NLP aprovechan las intenciones y las entidades para la conversación, y los bots pueden realizar acciones eficientes al reconocer sustantivos y verbos en el texto que escribe el usuario y luego cruzarlo con su diccionario.

8. Conversaciones basadas en flujo

Este tipo de conversaciones son el siguiente nivel de comunicación inteligente, donde las redes neuronales recurrentes (RNN) aprenderán de muchas otras conversaciones entre personas. En la Figura 5, la red neuronal visualiza el modelo de secuencia a secuencia.

Figura 5

Aprendizaje profundo



Nota. La figura representa una red neural. Denny Britz (2016).

Sistema de Peticiones, Quejas, Reclamos o Sugerencias (PQRS)

El sistema PQR son las actividades básicas que son parte del proceso de servicio al cliente desarrolladas en las oficinas comerciales y que dan respuesta a cada una de las necesidades que los clientes manifiestan (Aplextm, 2015).

Petición: Son aquellas solicitudes de servicio o de información y que está asociada a la prestación de los mismo por parte del usuario, dicha solicitud puede ser de manera verbal o escrita.

Queja: una queja está ligada a la calidad en la atención del usuario, es decir es cualquier manifestación de no conformidad y esta puede ser oral o escrita, está asociada a restricciones en las que se ha prestado o dejado de ejercer algún servicio.

Reclamos: Un reclamo es básicamente la disconformidad sobre algún tipo de servicio que ha sido adquirido por parte del cliente de una empresa.

Este sistema es una herramienta que se usa para comprender e identificar las cualificaciones e inhabilidades de los usuarios, por lo que es imperativo brindar respuestas oportunas y mejorar los servicios para la plena satisfacción del usuario.

El seguimiento adecuado de PQRS a través de un programa informático permitirá la ejecución de los informes pertinentes por parte de la empresa u organización con el fin de tomar medidas para mejorar la toma de decisiones en interés de la empresa. Los beneficios del sistema PQRS para nosotros incluyen:

- Prestar un servicio a los usuarios de calidad mediante la automatización de cada uno de los procesos.
- Llevar a cabo un correcto control de las PQRS.
- Minimizar considerablemente los costos de la empresa u organización.
- Mejora de la satisfacción del usuario y de la comunicación entre las partes que están involucradas.

El objetivo principal de este software es descentralizar las prácticas relacionadas con la gestión de servicios. Una parte importante es el excelente servicio al cliente, que será un factor importante en el éxito y crecimiento de una empresa u organización, manteniendo la calidad del servicio y la eficiencia de los procesos (Aranda, 2017).

Análisis RFM

Por sus siglas en inglés RFM (Recency, Frequency, Monetary) no es más que una técnica del marketing que es usada para llevar a cabo el análisis del comportamiento de los clientes de una empresa u organización (Birant, 2011). Para llevar a cabo este análisis se empieza evaluando las compras realizadas por el cliente implementando los tres factores del modelo: (R) Recencia de compra, (F) la Frecuencia de compra y (M) el monto de la compra (MORELO TAPIAS, 2014).

Según varios estudios y encuestas, los clientes que gastan más dinero o compran con más frecuencia para una empresa son los que acaban siendo más sensibles a la información que les transmite la empresa y responden mejor a cada plan de marketing (Birant, 2011).

El análisis RFM se basa en la “Ley de Pareto” o también conocida como regla del 80/20 (Pareto, 1896), cuyo autor vio que en su país el 80% de la tierra era propiedad del 20% de la población. Con esta fundamentación en Pareto, se empieza a analizar más extensamente en muchas situaciones cotidianas. En el análisis RFM podemos decir que el 80% de los clientes de una empresa compran al 20% de los clientes de esa empresa, de igual forma podemos decir que el 20% de los clientes generan el 80% de las ventas de esa empresa (MORELO TAPIAS, 2014).

En la aplicación del análisis RFM a cada uno de los clientes se le concede un rango o categoría que va de 1 a 5, esto con el fin de poderlos clasificar por los factores mencionados anteriormente. A los tres factores juntos se los denomina celdas RFM, los datos de los clientes son analizados para poder determinar cuáles son los mejores y peores clientes en un tiempo

estipulado, en el que obtienen una calificación de 5-5-5 son los ideales para la empresa (Birant, 2011).

Chatbot

Los asistentes virtuales Chatbot son programas informáticos, que permiten simular una conversación con respuestas razonables y que se encuentran dentro de un contexto determinado a través de la implementación de machine learning y metodologías que permiten el desglose del lenguaje natural, dando la sensación de una conversación humana (Manaure, 2017), (COGNIAPPS, 2016).

Un Chatbot según lo expusieron Luzardo & Hernández (2010) forman parte de los entornos virtuales inteligentes y se comprende “Como entidad software, a partir del conocimiento que tiene de su entorno, realiza una serie de acciones encaminadas a lograr un objetivo, ya sea de manera proactiva o porque una situación particular lo requiere”.

Tipos de Chatbot

Son dispositivos que han avanzado basados en la inteligencia artificial, el aprendizaje automático y el procesamiento del lenguaje natural, y han avanzado tanto que hoy en día se pueden categorizar de otras formas según los servicios que brindan. Los tipos de Chatbot que podemos encontrar son (Manaure, 2017):

- Los Chatbots para ventas son aquellos que se enfocan en comercializar el producto o servicio que ofrece la empresa que los implementa.
- Chatbot de servicio al cliente, estos se orientan a brindar atención respecto a dudas que los usuarios tengan respecto a un determinado producto o servicio.
- Chatbot de noticias y contenido, son aquellos que se han implementado en canales conversacionales instantáneos, y su objetivo es el envío de contenido de una manera masiva.

De igual forma la evolución de los Chatbot ha logrado diferentes aplicaciones hoy en día, y están inmersos en el día a día algunos de estos son, por ejemplo: Siri de Apple, Google Now de Google, entre otros (COGNIAPPS, 2016).

Herramientas para la obtención de datos desde los canales conversacionales.

Para el análisis de los datos generados por los usuarios de la empresa SICE se consideró los Chatbots más populares hoy en día, dentro de la infinidad de agentes disponibles en el mercado se procedió a comparar y estudiar tres de ellos. En la tabla 4 se detalla el análisis realizado entre los Chatbots Dialogflow, IBM Watson y AmazonLex.

Tabla 4*Comparación entre Dialogflow, IBM Watson y AmazonLex*

	Canales	Integraciones	Web y Móvil Integraciones	Lenguajes	Costo
Dialogflow	Voz, texto	<ul style="list-style-type: none"> • Asistente de Google • Slack • Viber • Facebook Messenger • Twitter • Twilio • Wordpress, etc... 	<ul style="list-style-type: none"> • Integración sin código • Integración básica incorporada 	Soporte para más de 20 lenguajes, incluido inglés, español, portugués, etc.	<ul style="list-style-type: none"> • Gratis para el plan estándar • Para el plan empresarial: \$0.002/respuesta

IBM Watson	Voz y texto	<ul style="list-style-type: none"> • Agente de voz • Slack • Facebook Messenger • Wordpress 	<ul style="list-style-type: none"> • Interfaz de usuario de chat básica para sitios web 	<p>Soporte para 10 lenguajes, incluido inglés, español, portugués, etc.</p>	<ul style="list-style-type: none"> • Plan estándar gratis con ciertas restricciones • Plan de pago: \$0.0025/mensaje
AmazonLex	Voz y texto	<ul style="list-style-type: none"> • SMS • Slack • Kik • Twilio 	<ul style="list-style-type: none"> • Proveedor de interfaz de usuario de chat básico para probar el sitio web 	<p>En la actualidad soporta el idioma inglés.</p>	<ul style="list-style-type: none"> • En el primer año 10k respuestas de texto • Voz: \$0.004/respuesta • Texto: \$0.00075/respuesta

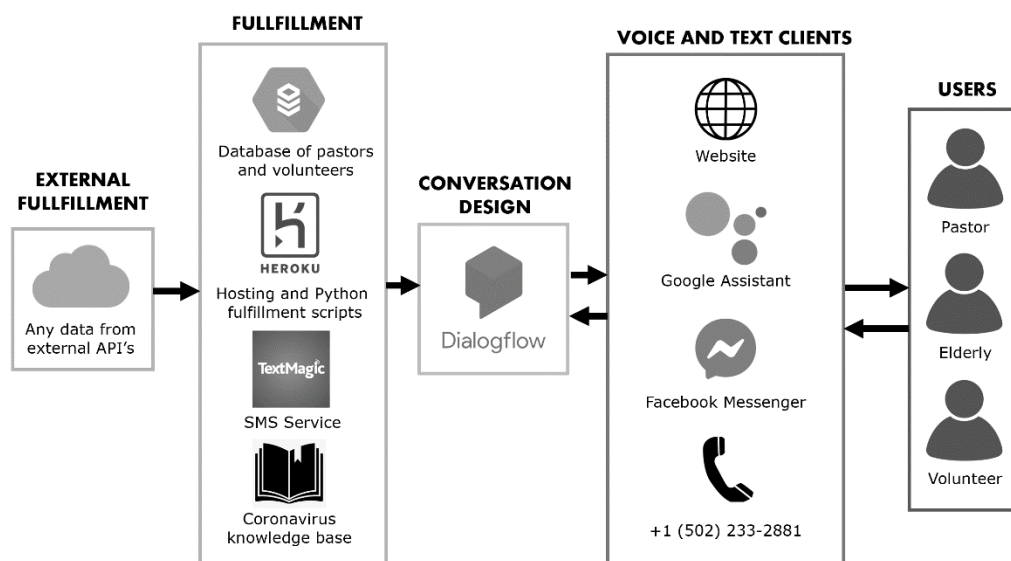
En base a las características analizadas en la tabla anterior se toma la decisión de usar el Chatbot Dialogflow, ya que este nos permite acceder a múltiples funciones de una manera gratuita, además que es apto para implementar tanto en canales conversacionales de redes sociales como de páginas web.

Dialogflow

La API para el procesamiento del lenguaje natural de Google se llama Dialogflow, anteriormente conocida como Api.ai, y la plataforma se conecta y permite la creación de diferentes conjuntos de respuestas para mensajes específicos utilizando diferentes técnicas de NPL. Dialogflow usa intenciones, acciones parametrizadas, entidades y contextos en modo de texto o modo de voz. El diagrama muestra la arquitectura de Dialogflow y su biblioteca llamada "Fullfillment" y cómo interactúa con diferentes plataformas.

Figura 6

Arquitectura Dialogflow



Nota. La figura representa la arquitectura de Dialogflow para la comunicación con diferentes plataformas. Tomado de Cytotoxic Tees (2020).

Dentro del sinnúmero de características que ofrece dialogflow tenemos dos que son esenciales al momento de caracterizar las tecnologías NPL:

Codificación rápida. Dado que la plataforma integra un editor de código basado en javascript, las tareas relacionadas con el código se pueden realizar rápida y rápidamente, lo que permite a los desarrolladores vincular diferentes herramientas con Chatbots, además de crear sus propios Webhookss.

Aprendizaje automático accionado. Es de útil para el apoyo de los desarrolladores ya que les permite capacitar a sus agentes conversacionales para comprender de una mejor manera lo que el usuario escribe.

Herramientas para el almacenamiento de los datos

Para almacenar las interacciones que se dan entre el usuario y el agente conversacional primeramente se analizó que tipo de gestor se adapta mejor a las necesidades del negocio y la empresa, para analizar y elegir tenemos los sistemas gestores de base de datos relacionales y los de datos no relacionales, siendo más eficiente y el que más se adapta al giro del negocio el gestor de datos no relacionales ya que este nos permitirá tener un alto rendimiento y una disponibilidad alta de los datos.

Tabla 5

Comparación SQL y NoSQL

Característica	Relacional	No relacional
Rendimiento	Bajo	Alto
Disponibilidad	Bueno	Bueno
Consistencia	Bueno	Bajo
Confiablez	Bueno	Bueno

Característica	Relacional	No relacional
Almacenamiento de datos	Bueno para BBDD de mediano tamaño	Optimizado para cantidades masiva de datos
Escalabilidad	Alto (más costoso)	Alto

Una vez seleccionado el tipo de gestor, es importante realizar una comparación de que gestor de base de datos no relacional es el más conveniente y que nos permita obtener mejores resultados de una manera sencilla y económica. En este análisis entran a comparación tres grandes gestores cuyas características son muy similares.

Tabla 6

Comparación Casanda, MongoDB y Redis

Nombre	Cassandra	MongoDB	Redis
Descripción	Almacén de columnas anchas basado en ideas de BigTable y DynamoDB	Uno de los almacenes de documentos más populares disponibles tanto como un servicio en la nube totalmente administrado como para la implementación	Popular plataforma de datos en memoria utilizada como caché, agente de mensajes y base de datos que se puede implementar en las instalaciones, en nubes y en entornos híbridos

Nombre	Cassandra	MongoDB	Redis
		en infraestructura autogestionado	
Modelo de base de datos principal	Almacenamiento de columnas anchas	Almacenamiento de documentos	Almacenamiento clave-valor
Solo basado en la nube	No	No	No
Lenguaje de implementación	Java	C++	C
Métodos de replicación	Factor de replicación seleccionable	Implementaciones de múltiples fuentes con MongoDB Atlas Global Clusters Replicación de réplicas de origen	Replicación multiproyecto Replicación de réplica de origen
Concurrencia	Si	Si	Si
Durabilidad	Si	Si	Si
Métodos de partición	Fragmentación	Fragmentación	Fragmentación

En base a las características y parámetros analizados en la tabla anterior se toma la decisión de usar el gestor de base de datos no relacional MongoDB, ya que posee una amplia documentación además que es muy conocida mundialmente, además de que nos permitirá a futuro un escalamiento sencillo de acoplar según las necesidades de la empresa.

MongoDB

MongoDB es una base de datos no relacional, que permite tener una alta disponibilidad, escalabilidad horizontal y una distribución geográfica flexible lo que la hace fácil de usar. MongoDB es de uso gratuito y almacena los datos en documentos similares a JSON y las consultas ad hoc ofrecen maneras potentes para acceder a los datos.

Herramientas para el análisis de datos

Para llevar a cabo el análisis de los datos de la empresa SICE se tomó en cuenta diversas herramientas populares hoy en día, pero finalmente se procedió a estudiar y comparar dos de ellas tanto para el proceso de análisis y visualización de los datos.

Para el proceso de Minería de datos se analizó Weka y el lenguaje de programación R, si bien nos encontramos con dos herramientas poderosas y similares en cuanto a características para aplicar la minería de datos, esta vez seleccionaremos R ya es un artilugio fuerte de fácil usabilidad y gratuito lo que nos permite una infinidad de posibilidades para realizar nuestro proyecto de investigación.

Tabla 7

Comparación Weka y R

Característica	Weka	R
	Agrupación	
EM	X	X
KMeans	X	X
XMeans	X	

Hierarchical clustering		X
Bagged clustering		X
Cluster ensembles		X
	Arboles	
ID3	X	
C4.5	X	
Cart	X	X
Arboles de decisión	X	X
Arboles aleatorios	X	X
	Funciones	
Regresión Lineal	X	X
Regresión Logística	X	X
Regresión Isotónica	X	X
Procesos Gaussianos	X	X
	Visualización	
Estadística descriptiva	X	X
Tabla de Frecuencias	X	X
Gráfico de dispersión	X	X
Gráfico de dispersión con matrices	X	X
Histogramas	X	X
Gráficos/Arboles	X	X
Tabla de elevación	X	X

Lenguaje R

R es un lenguaje diseñado para entornos computacionales y gráficos estadísticos. Es de código abierto oferta diferentes metodologías (análisis de series en tiempo real, modelado no lineal, pruebas de clasificación y agrupamiento, etc.). R permite trabajar tanto en línea de comandos como en interfaz gráfica.

El entorno R

No es más que un conjunto de servicios de software que nos permite manipular datos, así como el cálculo y la representación gráfica. El entorno R incluye:

- Conjunto de operadores para el cálculo de matrices.
- Gran colección integrada para el procesamiento de datos.
- Manejo eficaz y facilidad de almacenamiento de los datos.
- Visualización de los datos en forma gráfica.

RStudio

Es un IDE que básicamente genera una interfaz potente de usuario para R.

Paquetes de Minería de Datos R

Caret: Está conformado por la agrupación de funciones que intentan minimizar en gran medida el proceso de creación de modelos predictivos. Además, contiene herramientas para la partición de datos, el preprocesamiento, la selección de funciones, el ajuste de modelos y más.

Rattle: Este paquete nos permite realizar algunos resúmenes estadísticos, así como la visualización de datos, es decir, modifica y recrea los datos en una forma que puede ser modelada y comprensible, construye modelos supervisados y no supervisados de los datos.

FactoMineR: Es un paquete de R con un enfoque exclusivamente al análisis multivariado exploratorio de datos. Las metodologías delimitadas en este paquete son enfoques multivariantes exploratorios, como el análisis de componentes principales, análisis de correspondencia o agrupación.

Otras herramientas

Heroku

Es una plataforma que se encuentra en la nube y que permite a las organizaciones construir, supervisar sus aplicaciones de una manera segura y a un bajo costo. Al ser una plataforma que se encuentra en la nube a los desarrolladores les evita la preocupación y gastos por la infraestructura, Heroku permite desplegar aplicaciones en cualquier lenguaje de programación: Java, Node.js, PHP, etc.

Node.js

Node.js es un entorno de ejecución de JavaScript y es dirigido básicamente a la asincronía de las situaciones, está diseñado para la creación de aplicaciones network escalables. Node.js es muy similar en cuanto a diseño a sistemas como Event Machine de Ruby y Twisted de Python, lo que lo diferencia de estos es que Node.js lleva el modelo de eventos a otro nivel ya que incluye bucles de eventos como runtime de ejecución en lugar de solo tener una biblioteca encargado de eso. Si bien Node.js está diseñado para trabajar sin hilos se puede aprovechar sus múltiples núcleos que vienen en su entorno, se pueden generar subprocessos o procesos haciendo uso de la API `child_process.fork()`, Node.js en conclusión es un entorno evolutivo que crece continuamente en los últimos años ya que muchos desarrolladores lo usan para la creación de sus aplicaciones (Node.js, 2022).

Capítulo III

En el presente capítulo se presenta el tanto los requisitos funcionales como no funcionales del agente conversacional (Chatbot), primeramente, se empieza haciendo una propuesta formal de como estará diseñada la arquitectura del Chatbot, así mismo se detalla mediante diagramas de flujo de procesos de como el Chatbot interactuará con los usuarios de la empresa SICE. Finalmente, con los datos recogidos a lo largo de las interacciones del Chatbot con los usuarios se procederá a aplicar las técnicas de minería de datos para mejorar la atención a los clientes.

Propuesta e implementación del Chatbot

Planteamiento

SICE carece de un Chatbot para responder las inquietudes que se dan por parte de sus usuarios en cada uno de sus canales conversacionales y evidenciando los posibles beneficios, pasa por un proceso para la ejecución y aplicación de un plan estratégico enfocado en el desenvolvimiento tecnológico con la intención de mejorar la atención que le brinda a cada uno de sus clientes.

Según lo expuesto en el capítulo anterior se ha optado por seleccionar a Dialogflow es el pilar para la implementación del Chatbot. Anteriormente se mencionó las características y beneficios lo que lo convierten en idóneo para la realización del presente trabajo de investigación.

Con la empresa se ha llegado al acuerdo que el Chatbot será capaz de:

Requerimientos no funcionales:

RNF0. Consultar los cursos y certificaciones que la empresa ofrece en una base de datos no relacional alojada en MongoDB.

RNF3. Desplegarse en un servidor de la nube.

Requerimientos funcionales:

RNF1. Responder a las preguntas frecuentes (FAQ).

RNF2. Buscar cursos/certificaciones de la empresa.

Como se mencionó anteriormente las funcionalidades están distribuidas dependiendo de la complejidad e impacto.

Para iniciar se llevará a cabo las siguientes actividades:

- Se procederá a normalizar la información de los cursos/certificaciones de SICE mismo que se encuentran en formato csv.
- Se creará una cuenta en MongoDB y posteriormente un proyecto el mismo que nos permitirá hacer uso de las colecciones en tiempo real.
- Una vez normalizada la información se procederá a cargar a MongoDB, cabe recalcar que los datos deben estar en formato JSON para no tener ningún tipo de problemas en el proceso de la importación.
- Se procederá a configurar las reglas y backups de la base de datos para el proceso de autenticación.
- Se creará una cuenta en Google Cloud Console.
- Se creará un proyecto en Dialogflow.
- Se creará el agente conversacional en Dialogflow.
- Se procederá a crear cada una de las intenciones.
- En esta sección se procederá a entrenar a cada una de las intenciones del Chatbot.
- Cada pregunta tendrá una intención definida y entrenada para que se active inmediatamente el Chatbot la detecte y pueda ofrecer una respuesta coherente.
- Se procederá a habilitar la integración mediante de Webhook propia de Dialogflow esta nos permitirá conectar el Chatbot con Facebook Messenger y la página web de la empresa.

- Se procederá a configurar una aplicación de desarrollador en la página de Facebook developers, esto permite conectarse a la página de Facebook de la empresa con el Webhook de Dialogflow.
- Una vez configurada la app en Facebook developers se procederá a configurar el plugin de chat en la página de Facebook, esto servirá para agregar el chat a la página web de la empresa SICE.

Dentro de Dialogflow existen términos mismos que se detallan a continuación para una mejor comprensión:

- **Agente:** El término en Dialogflow básicamente se refiere a los módulos de comprensión del lenguaje natural (NLU) que se incluyen en una aplicación, sitio web, producto o servicio y traducir las solicitudes de texto o voz proporcionadas por el usuario en datos procesables. Esta traducción ocurre solo si el texto recibido o la expresión de voz coincide con la intención definida en el agente.
- **Intención:** En estos se asigna la entrada, es decir, lo que el usuario puede escribir. Dentro de cada intención definida en el Chatbot, se pueden definir ejemplos de expresiones de usuario y cómo esas expresiones activan la intención. Además, es posible determinar qué extraer de la expresión y cómo responder.
- **Fulfillment o Cumplimiento:** Distribuye la solicitud que llega al bot y la transporta a una fuente externa para así generar una respuesta y reenviarla al usuario todo esto a través de Webhook. La configuración permite llevar información de una intención que coincide a un servicio y así obtener una respuesta.

- **Entidades:** Estas se usan para la extracción de los datos que son importantes dentro de un texto o voz, dentro de Dialogflow existen tres tipos de entidades, las de sistema, de desarrollador y las de usuario.
- **Contextos:** Proyectan la situación real de la solicitud realizada por el usuario y ayuda a que el agente conversacional pase información de un intento a otro. Se aplican varias combinaciones de contextos de entrada y salida para controlar la ruta del usuario a través del diálogo para una conversación determinada.
- **Eventos:** Dialogflow nos permite admitir eventos de diferentes plataformas como por ejemplo Google Assistant, Slack, etc. Dichos eventos estipulan diferentes intenciones basándose en algún evento que haya sucedido en lugar de lo que el usuario trata de transmitir con su mensaje de texto o voz.
- **Frases de entrenamiento:** Son básicamente ejemplos de lo que el usuario podría llegar a decir, para de esta manera hacer coincidir con una intención en particular.
- **Acción y parámetros:** La información relevante se define aquí, es decir, los parámetros extraídos de la declaración del usuario. Algunos ejemplos de dicha información incluyen ubicación, nombre, correo electrónico, fecha, hora y similares.
- **Respuesta:** Una expresión de texto o voz que se muestra al usuario final.

Una vez culminadas el conjunto de actividades detalladas anteriormente, se procede a realizar las siguientes tareas para continuar con el flujo de desarrollo e implementación del Chatbot:

- Se procederá a crear un servidor en el entorno de desenvolvimiento de Node.js.
- Se genera un servicio web sencilla con un Webhook usando Node.js Express.

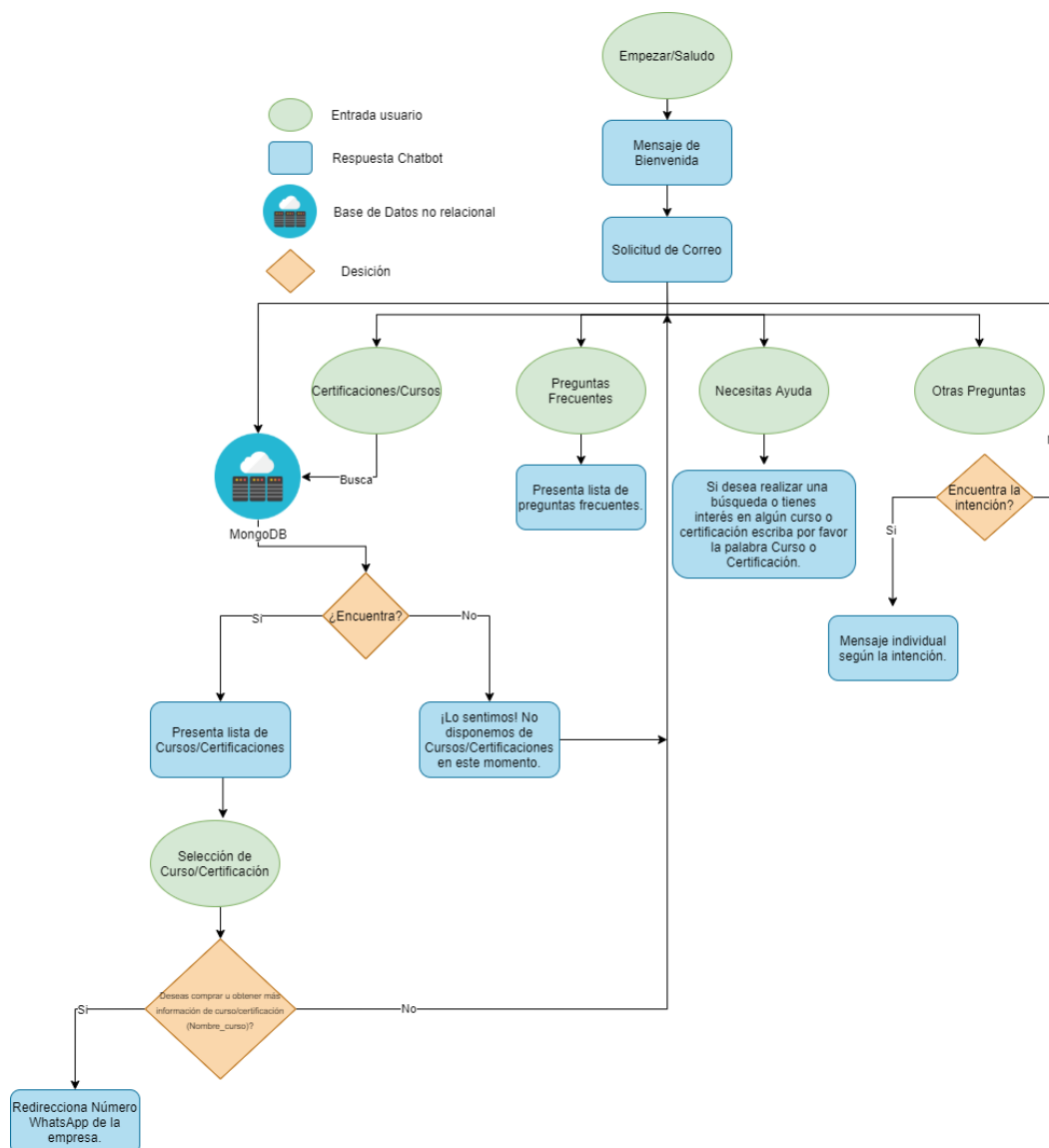
- Se obtiene las credenciales del api de Facebook y de Google para configurar correctamente el servidor y este pueda recibir los mensajes de los usuarios.
- Se procederá a crear el primer endpoint que se usará en la validación de Facebook y Google.
- Se procederá a crear un segundo endpoint que servirá en la gestión de los mensajes de texto.
- Se procederá a conectar la app de desarrollador de Facebook misma que se creó en la versión anterior, lo siguiente será conectar el servidor Webhook y las páginas tanto de Facebook como la web, en otras palabras, funcionará como un middleware entre ellos.
- Se procederá a configurar en el servidor la autenticación de Dialogflow.
- Se creará los métodos necesarios en Node.js para poder recibir los mensajes de los usuarios.
- Se creará la conexión a la base de datos no relacional MongoDB.
- Se procederá a crear un método para hacer el CRUD a la base de datos.
- Se entrenará cada una de las entidades.
- Se procederá a desplegar el servidor en la plataforma como servicio denominado Heroku.
- Se habilitará a los usuarios el Chatbot tanto en la página de Facebook con en la página web.

Diseño

En la Figura 7, se puede observar el flujo que llevará cada una de las conversaciones que se den entre el Chatbot y los usuarios.

Figura 7

Flujo de la conversación entre el usuario y Chatbot



Nota. La figura representa el flujo que llevarán a cabo las conversaciones entre el Chatbot y el usuario. Autor. (2022).

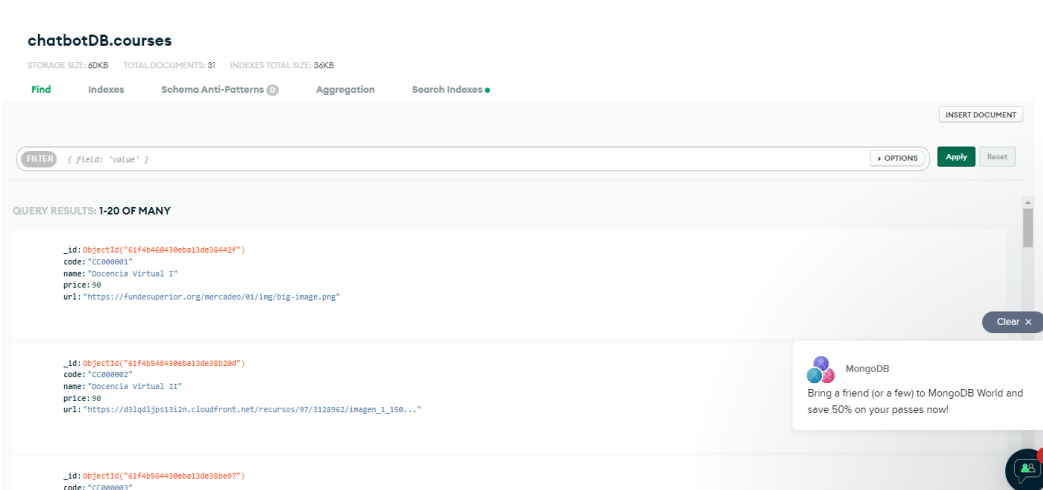
Estructura de la base de datos en MongoDB

Dado que la empresa SICE guardaba los datos de la organización y sus clientes en libros de Excel, es necesario primeramente hacer una normalización y estructuración de los datos.

En la figura 8 y 9 se visualiza cada una de las colecciones y sus documentos, la primera colección consta con la información de los cursos y certificaciones que dispone la empresa y a la cual se ha nombrado como “courses” figura 8, por otro lado, se puede ver la colección denominada “Chatbotusers” que guardará los datos del cliente y las intenciones que ha activado en el agente conversacional figura 9, además se procedió a la creación de la colección “customers” que contiene los datos de los clientes, la colección “invoice” que contendrá los datos de cada una de las transacciones de la empresa SICE, finalmente las colecciones “frequents” y “knowledges” que nos servirán para guardar los datos obtenidos en la fase de modelado.

Figura 8

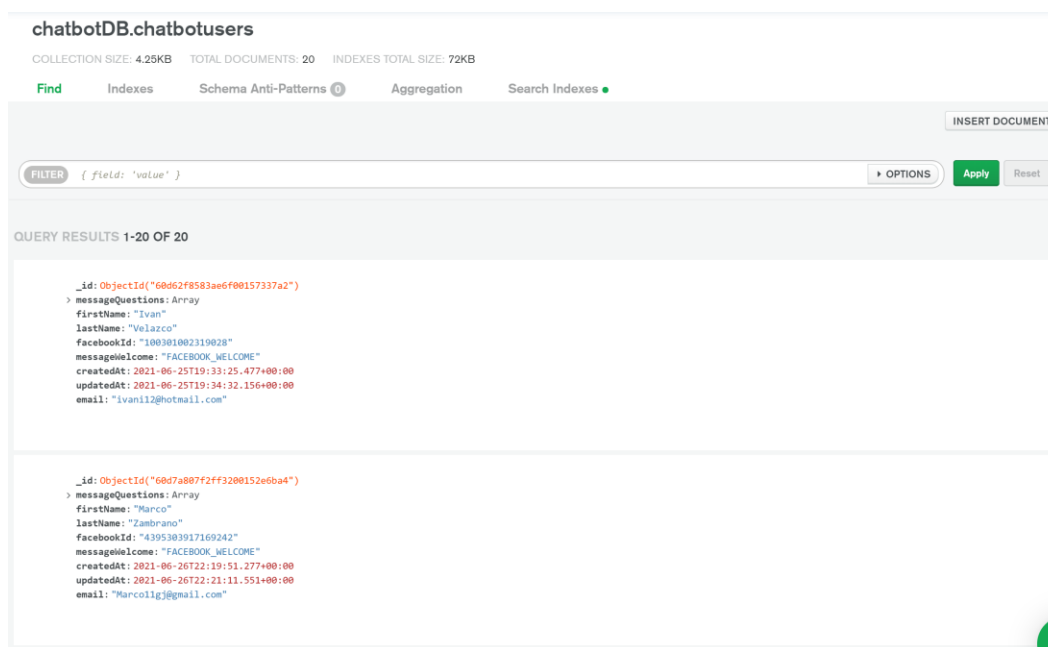
Estructura de los documentos de la colección courses



Nota. La figura representa la estructura de los documentos de la colección certifications. Autor. (2022).

Figura 9

Estructura de los documentos de la colección Chatbotusers



Nota. La figura representa la estructura de los documentos de la colección Chatbotusers. Autor. (2022).

En las figuras 10 y 11 se muestra la estructura del objeto JSON que tendrá la colección “courses” y “Chatbotusers” respectivamente.

Figura 10

Estructura JSON colección courses

```

_id: ObjectId("6153ca653fdefd2a49e1bca8")
name: "Certificaciones en Marketing"
description: "Certificación Internacional en Negocios Digitales con Mención en Estrategias de Marketing Digital."
price: 750
status: "Active"
type: "Certificación"

```

ObjectId
String
String
Double
String
String

Nota. La figura representa la estructura y el tipado del objeto JSON de la colección certifications. Autor. (2022).

Figura 11

Estructura JSON colección Chatbotusers

```
_id: ObjectId("61362c2ff4516a0015c3e41c")
> messageQuestions : Array
facebookId: "189271994826516"
messageWelcome: "FACEBOOK_WELCOME"
createdAt: 2021-09-06T14:56:47.012+00:00
updatedAt: 2021-09-06T14:57:30.675+00:00
email: "oswaldo_1990@yahoo.es"
messageFormasPago: "Formas de pago"
```

ObjectId
Array
String
String
Date
Date
String
String

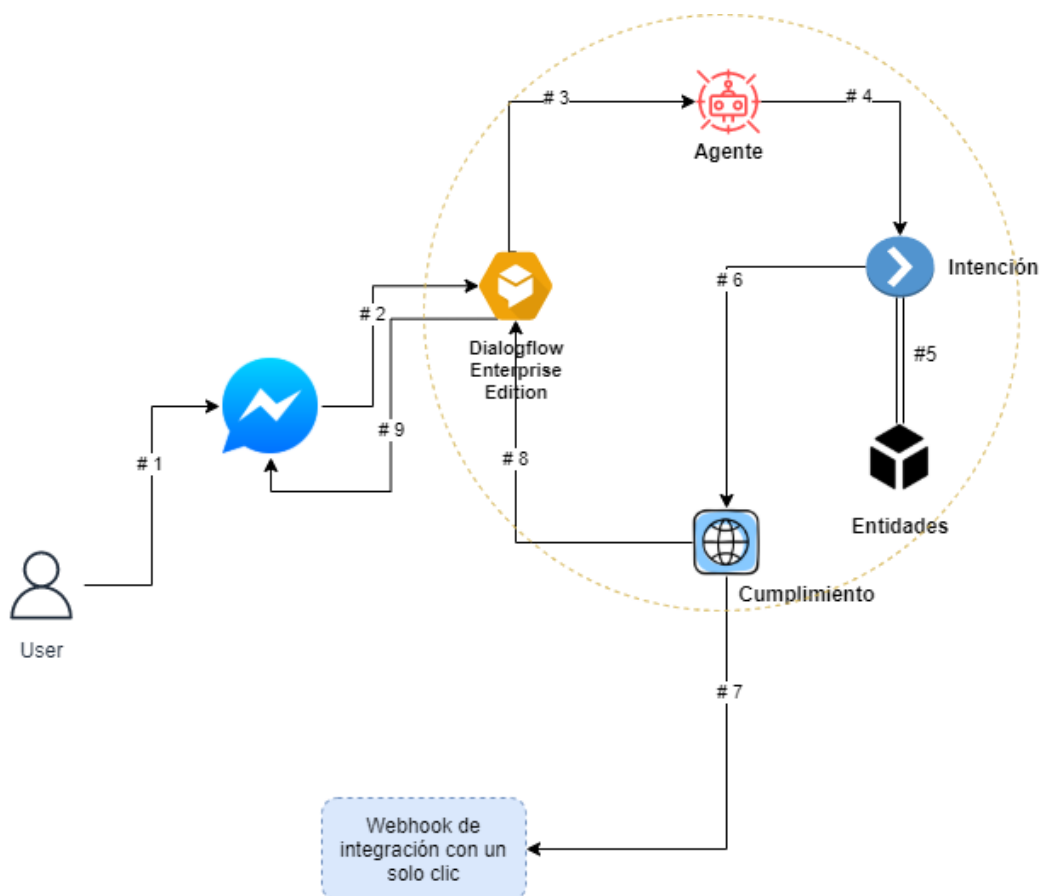
Nota. La figura representa la estructura y el tipado del objeto JSON de la colección Chatbotusers. Autor. (2022).

Arquitectura

En la figura 12, tenemos la arquitectura que se fundamenta en servicios en los cuales el flujo inicia con interacciones que se dan entre un usuario y Facebook Messenger, en el círculo podemos ver todos los componentes de Dialogflow que intervienen, hasta finalizar en un Webhook "self-hosted" de Dialogflow.

Figura 11

Arquitectura del Chatbot "One Click Integration"



Nota. La figura representa la arquitectura denominada "One click integration" de C. Autor. (2022).

A continuación, se detalla cómo se desarrolla el flujo de una conversación:

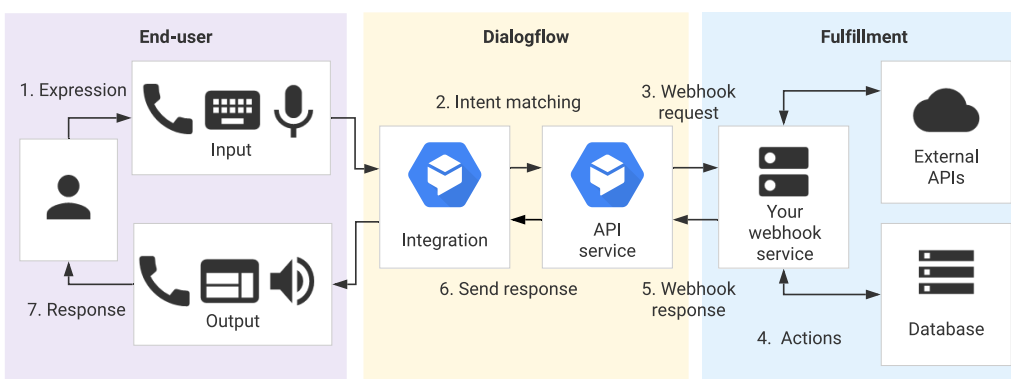
- El usuario escribe un mensaje.
- Facebook Messenger envía el mensaje recibido en forma de texto.
- Dialogflow recibe ese texto y procede a determinar a qué agente lo enviará.
- El agente procede a primeramente a identificar la intención del usuario, luego pasa a la intención correcta.

- Las intenciones de Dialogflow usan entidades que son usadas para almacenar los valores de los parámetros en caso de estar configurados.
- Las intenciones de Dialogflow pasan las solicitudes recibidas junto con las entidades hacia el Fulfillment/Cumplimiento.
- Fulfillment usa un Webhook de “one-click-integration/webhook de integración en un solo Click” para Facebook Messenger.
- Fulfillment entrega la respuesta en formato JSON desde el Webhook.
- Finalmente, Dialogflow envía la respuesta hacia Facebook Messenger.

Es importante describir el proceso que lleva a cabo el cumplimiento/fulfillment dentro de Dialogflow y como este nos permite responder las inquietudes a los usuarios. En la figura 12 se detalla este flujo.

Figura 12

Fulfillment en Dialogflow



Nota. La figura representa la arquitectura denominada “One click integration” de Chatbot. Dialogflow. (2019).

RNF1 “Responder a preguntas frecuentes”


Descripción: Este caso de uso le permitirá al usuario consultar las preguntas más frecuentes que la empresa SICE tiene a disposición. Cabe recalcar que cada pregunta dentro

de la opción preguntas frecuentes puede ser invocada por el usuario de una manera individual, estas preguntas y respuestas se detallan en la Tabla 8.

El Chatbot mostrará al usuario la totalidad de interrogantes con sus respectivas respuestas en forma de lista con botones, mismo que permitirá al usuario elegir cualquier pregunta con un solo Click. Si el Chatbot detecta una intención distinta en las especificadas en la siguiente tabla las respuestas serán de tipo texto o si es que esa respuesta debe redirigir a un sitio web o enlace de una red social en ese caso serán las respuestas rápidas de tipo botón.

Tabla 8

Preguntas frecuentes SICE

N°	Pregunta	Respuesta	Tipo respuest a	Intención/Intent
1	¿Información de las certificaciones ?	Haremos llegar a tu Whatsapp el Brochure con los contenidos y perfil del ponente. Adicional se adjunta un botón con la redirección hacia el WhatsApp de la empresa.	Texto y respuesta rápida tipo lista con botón	InformacionCertificaciones
2	¿Tipo de aval y valor curricular de	Todas nuestras certificaciones tienen aval de SICE. 	Texto	TipoAval

N°	Pregunta	Respuesta	Tipo respuest a	Intención/Intent
las certificaciones ?	<p>Algunas cuentan con el aval de la Universidad Hemisferios.</p> <p>Algunas cuentan con el aval de la SETEC.</p> <p>Otras son Certificaciones Internacionales, con aval de BMF, World Business & Marketing Federation y Samarketing 360 de México.</p> <p>Gastronomía Mexicana cuenta con el aval de Lycée Culinaire de Cancún.</p> <p>Algunas de Educación, cuentan con el aval de UDESA, Universidad</p>			

N°	Pregunta	Respuesta	Tipo respuest a	Intención/Intent
		de San Andrés de Buenos Aires.		
3	¿Duración de las certificaciones ?	Tenemos cursos cortos entre 8 y 20 horas académicas. Las certificaciones son a partir de 40 horas académicas, hasta 300 horas académicas.	Texto	Tiempo de Duración
4	¿Modalidad?	Las modalidades son: presencial, semipresencial y Online con clases en vivo.	Texto	Modalidad cursos
5	¿Formas de pago?	Los pagos puedes hacerlo en efectivo, con depósito o transferencia.	Texto	Formas de Pago

N°	Pregunta	Respuesta	Tipo	Intención/Intent
		También con tarjeta de crédito, a través de la aplicación PayPhone.		
6	¿Número de contacto?	Se despliega un botón que redirigirá al usuario al WhatsApp de la empresa.	Botón	WhatsAppSice
7	¿Disponen de algún correo electrónico?	Escríbenos al siguiente correo: info@siceduccion.org	Texto	CorreoSice

Actores: Usuario SICE, Messenger, Messenger Web, Dialogflow.

Para entender el flujo principal del Chatbot se detallan a continuación los siguientes pasos:

- El usuario de SICE selecciona una de las opciones del menú y activa la intención descrita en la tabla 8.
- Messenger captura la selección del usuario y envía el mensaje como texto.

- Dialogflow recibe el mensaje que le envió el usuario desde Messenger o la página web y lo envía al Chatbot.
- Dialogflow según el mensaje la intención que debe activarse y luego pasa esa intención hacia el Webhook que está corriendo en nuestro servidor en Heroku.
- Cuando se activa la intención de “PreguntasFrecuentes” que es la que despliega todas las opciones descritas en la anterior tabla, el Fulfillment en Dialogflow entrega la respuesta al Webhook una lista en formato JSON con cada una de las preguntas frecuentes identificadas en la empresa SICE.
- Finalmente, Dialogflow devuelve la respuesta a Facebook Messenger.

En las interacciones con el Chatbot existen algunas excepciones:

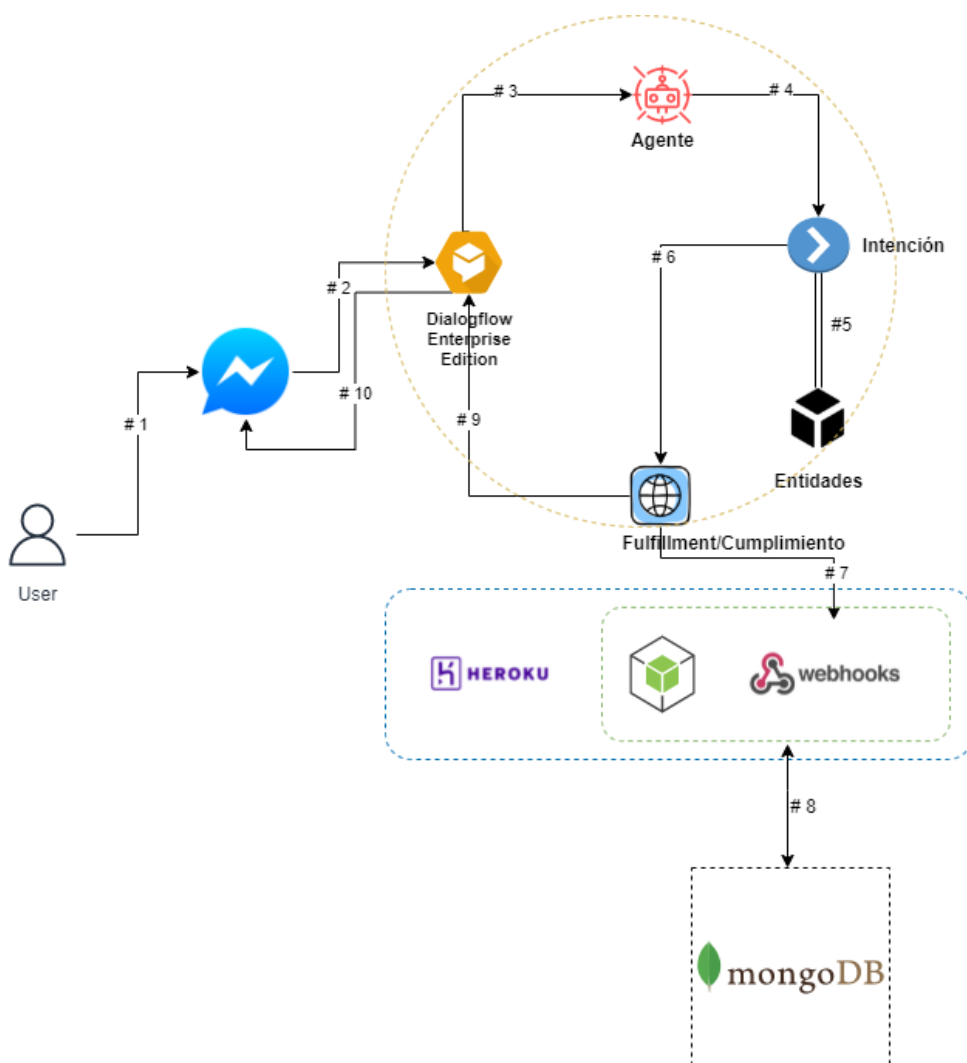
- Excepción 1:
 - Razón: Fallas en la conexión de internet.
 - Acción: Facebook Messenger no envía el mensaje (Mensaje no enviado).
- Excepción 2:
 - Razón: La Api de Facebook Messenger tiene problemas o falla.
 - Acción: Facebook Messenger no envía el mensaje (Mensaje no enviado).
- Excepción 3:
 - Razón: Dialogflow no activa ninguna intención, debido a que el mensaje del usuario no coincide con ninguna de las intenciones creadas.
 - Acción: Se activa la intención de “default fallback intent”, cuando sucede esto el Webhook captura la pregunta o mensaje del usuario y lo guarda en MongoDB para seguir alimentando con conocimiento al Chatbot, y al usuario se le mostrar en pantalla “Ups, no he entendido a que te refieres.”
 -

Arquitectura

La arquitectura se diferencia de la usada en la Figura 11 ya que esta vez se añade un Webhook externo que como se mencionó anteriormente se desarrolló en Node.js, este Webhook nos permitirá primeramente tener un control más exhaustivo de los mensajes a responder al usuario, y en segunda instancia guardar y consultar a la base de datos MongoDB.

Figura 13

Arquitectura Fulfillment con Webhook externo (Node.js) y MongoDB

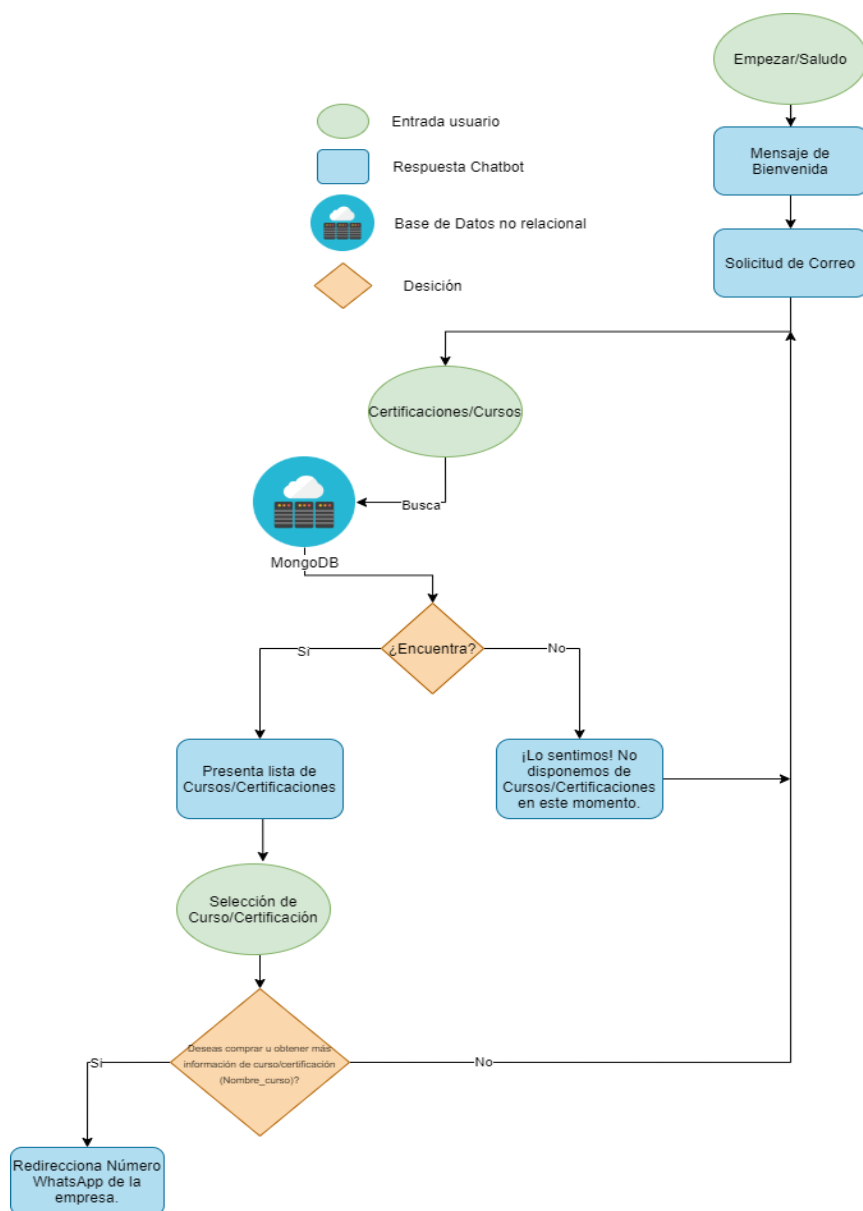


Nota. La figura representa la arquitectura de fulfillment con un Webhook externo desarrollado en Node.js y conexión a la base de datos no relacional MongoDB. Autor. (2022).

El diagrama de la Figura 14 representa el caso de uso RNF2 que consiste básicamente en la búsqueda de certificaciones en la base de datos MongoDB a través del Webhook en Node.js.

Figura 14

Proceso de la intención Certificaciones.



Nota. La figura representa la arquitectura del flujo del proceso que activa la intención de Certificaciones. Autor. (2022).

Descripción: Esta opción del menú que se le presentará al usuario “Certificaciones” le permite visualizar la lista de certificaciones que dispone la empresa, esta intención puede ser llamada desde el menú principal y también individualmente cuando el usuario escriba “certificaciones”.

El Chatbot devolverá una lista desplegable con las certificaciones disponibles y además un botón “Hacer compra” al pie de cada certificación; este botón al ser invocado retornará un mensaje al usuario donde se le especificará el WhatsApp de la empresa en forma de enlace para que así el usuario sea atendido por el personal de la empresa SICE directamente, tal y como se detalla en los siguientes ítems.

- El usuario selecciona la opción del menú principal “Certificaciones” o escribe la palabra “certificación o certificaciones”.
- Facebook Messenger toma ese mensaje y lo envía en formato texto.
- Dialogflow recibe ese mensaje en forma de texto y lo redirige al Chatbot “ChatbotSICE”.
- Se identifica el intent/intención del usuario y se activa la intención “CertificacionesSice” y se la pasa hacia el Webhook en la nube.
- El Webhook toma esa intención y activa la respuesta misma que se envía a Facebook Messenger.
 - Se presenta una lista desplegable con imagen de la certificación un título de la misma, una breve descripción, precio y un botón de “Hacer compra”.
 - El usuario al estar interesado en obtener una certificación selecciona una de la lista.
- Dialogflow recibe ese mensaje en forma de texto y lo redirige al Chatbot “ChatbotSICE”.

- Se identifica la intención y se la pasa al Webhook quien la recibe y se encarga de devolver la respuesta.
 - Se le muestra como respuesta un botón que dirige al WhatsApp de la empresa donde se le brindará más información de la certificación seleccionada y del caso concretar la compra.
- En caso de que el usuario no seleccione ninguna de las certificaciones y tenga otro tipo de preguntas el Chatbot responderá dependiendo de la intención que se active.

Algunas de las excepciones:

- Excepción 1:
 - Razón: Fallas en la conexión de internet.
 - Acción: Facebook Messenger no envía el mensaje (Mensaje no enviado).
- Excepción 2:
 - Razón: La Api de Facebook Messenger tiene problemas o falla.
 - Acción: Facebook Messenger no envía el mensaje (Mensaje no enviado).
- Excepción 3:
 - Razón: No se encuentra ningún documento en la colección "certifications" de MongoDB.
 - Acción: El Chatbot responde "Lo sentimos, pero de momento no disponemos de ninguna certificación activa."
- Excepción 4:
 - Razón: MongoDB falla.
 - Acción: El Chatbot responde "No se ha podido obtener información de las certificaciones."
 -

- Excepción 5:
 - Razón: Dialogflow no activa ninguna intención, debido a que el mensaje del usuario no coincide con ninguna de las intenciones creadas.
 - Acción: Se activa la intención de “default fallback intent”, cuando sucede esto el Webhook captura la pregunta o mensaje del usuario y lo guarda en MongoDB para seguir alimentando con conocimiento al Chatbot, y al usuario se le mostrar en pantalla “Ups, no he entendido a que te refieres.”

Implementación

En esta sección se hace hincapié en el trabajo realizado tanto en la plataforma Dialogflow como en el servidor en la nube que contiene el Webhook descritos en la sección anterior del planteamiento. En esta sección también se llevará a cabo pruebas de funcionalidad.

A continuación, se ejecuta la aplicación del proceso de la implementación de cada una de las funcionalidades especificadas en la sección anterior tanto en Node JS, MongoDB como en Dialogflow.

- Configuración MongoDB

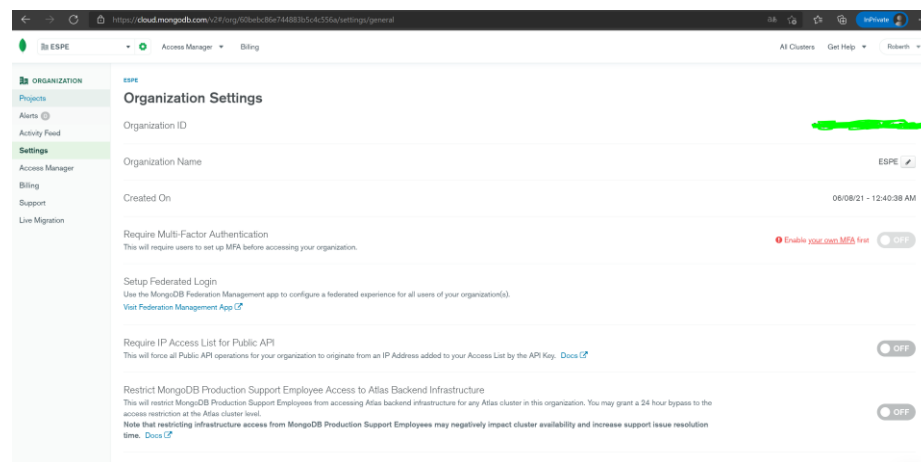
La generación de una base en MongoDB primeramente se debe crear una cuenta en MongoDB Atlas (Managed MongoDB Hosting | Database-as-a-Service | MongoDB) Figura 15.

Para efecto de este proyecto de investigación se ha seleccionado un proyecto “Free” que tiene algunas limitaciones por ser gratuito, pero se aprovechara al máximo en esta investigación.

La cuenta creada en MongoDB tiene la siguiente configuración, se procedió a ocultar información de Id por motivos de seguridad.

Figura 15

Proyecto en MongoDB.

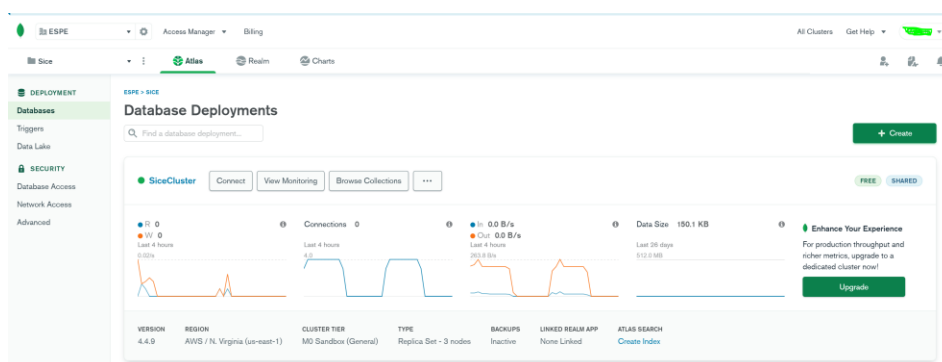


Nota. La figura representa la configuración general de la cuenta en MongoDB. Autor. (2022).

En la figura 16 se muestra la configuración del proyecto donde se creará y se administrará cada una de las colecciones.

Figura 16

Configuración proyecto MongoDB

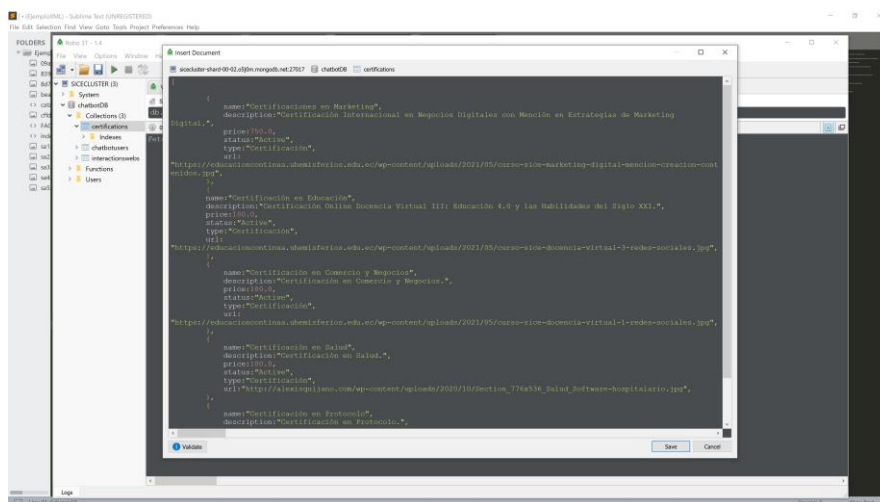


Nota. La figura representa la configuración general del proyecto en MongoDB en donde se crearán las dos colecciones. Autor. (2022).

Una vez normalizada la data que se encontraba en formato Excel se procede a la importación, los datos normalizados deben estar en formato JSON.

Figura 17

Insert documento en la colección courses



Nota. La figura representa la importación de los datos que antes estaban en formato Excel, se visualiza el array de objetos que contiene la información de los cursos y certificaciones.

En la figura 18, se muestra cada uno de los documentos insertados en la figura 17.

Figura 18

MongoDB documentos de la colección "courses"

Key	Value	Type
url	https://educacioncontinua.uhemisferios.edu.ec/wp-content/uploads/2021/05/c...	String
name	Certificación en Marketing	String
description	Certificación Internacional en Negocios Digitales con Menú en Estrategias de Marketing	String
price	180.0	Double
status	Activo	String
type	Certificación	String
url	https://educacioncontinua.uhemisferios.edu.ec/wp-content/uploads/2021/05/c...	String
name	Certificación en Educación	String
description	Certificación Online Docencia Virtual III: Educación 4.0 y las Habilidades del Siglo XXI	String
price	180.0	Double
status	Activo	String
type	Certificación	String
url	https://educacioncontinua.uhemisferios.edu.ec/wp-content/uploads/2021/05/c...	String
name	Certificación en Comercio y Negocios	String
description	Certificación en Comercio y Negocios	String
price	180.0	Double
status	Activo	String
type	Certificación	String
url	https://educacioncontinua.uhemisferios.edu.ec/wp-content/uploads/2021/05/c...	String
name	Certificación en Salud	String
description	Certificación en Salud	String
price	180.0	Double
status	Activo	String
type	Certificación	String
url	https://www.hidronor.cl/hidronor/wp-content/uploads/2014/03/certificaciones1...	String
name	Certificación en Protocolo	String
description	Certificación en Protocolo	String
price	180.0	Double
status	Activo	String
type	Certificación	String

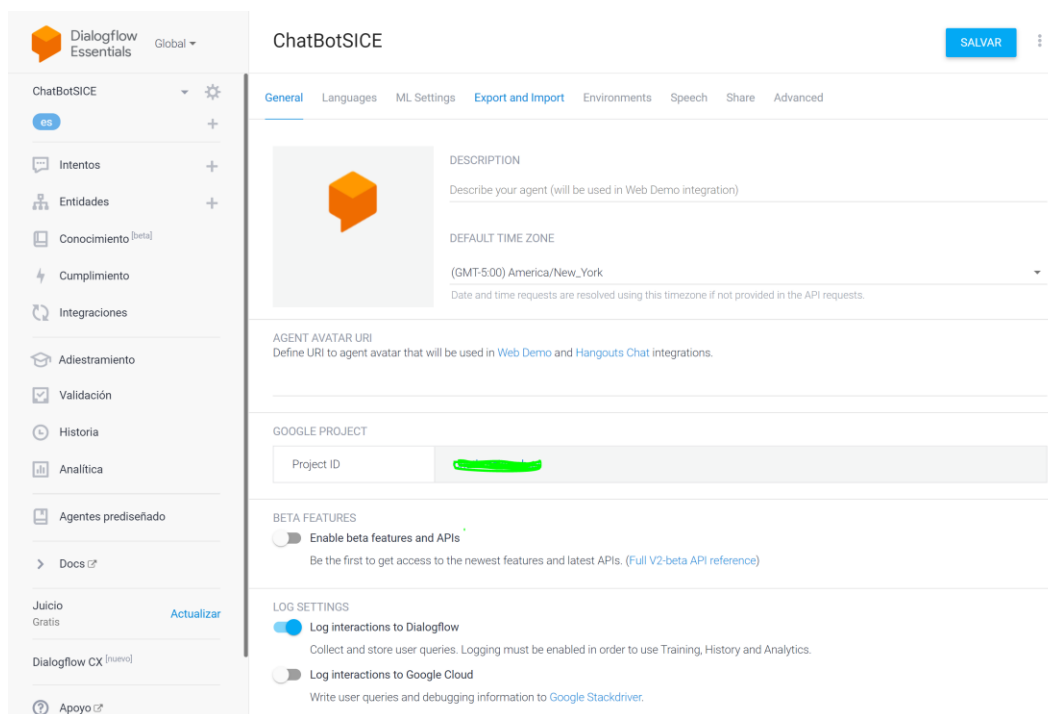
Nota. La figura muestra los documentos creados en la colección courses. Autor. (2022).

- Intents/Intenciones Dialogflow

Lo primero que se hace en Dialogflow es crear el agente conversacional (Chatbot) donde se le especifica el nombre, la zona horaria y si se desea una breve descripción.

Figura 19

Configuración general del Chatbot "ChatBotSICE"



Nota. La figura muestra la configuración que tendrá el Chatbot “ChatBotSICE” el que permitirá la interacción con los usuarios de la empresa SICE. Autor. (2021).

Intenciones:

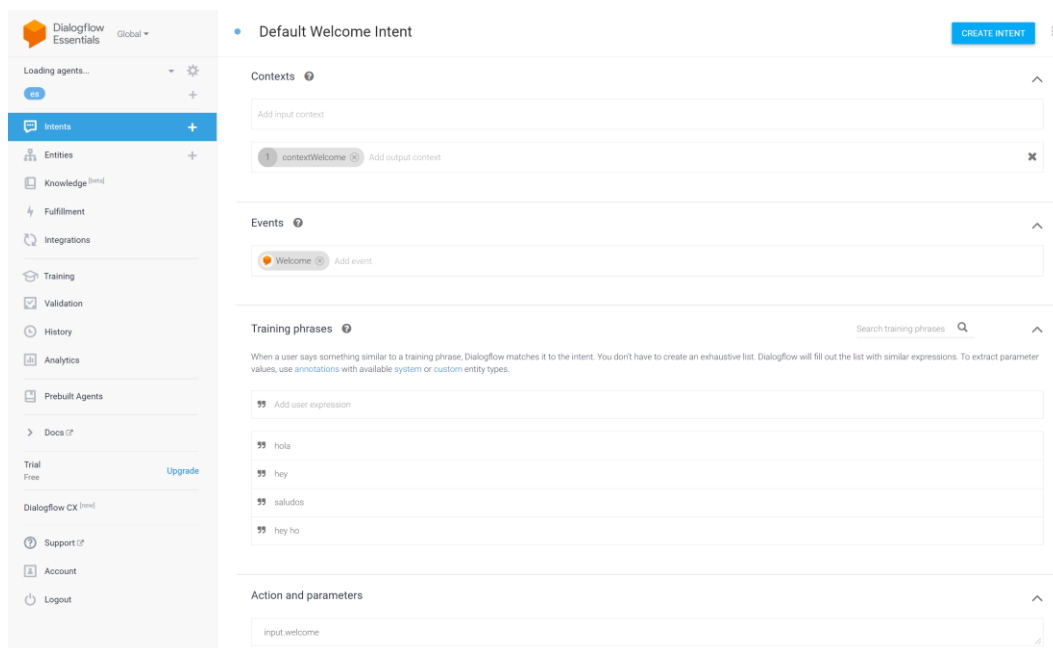
Se seleccionó las intenciones más importantes creadas en el agente conversacional y a continuación se procede a detallarlas.

- Intención de “Default Welcome Intent”: Esta intención tiene la configuración para que se active inmediatamente el usuario abra el chat en la página de Facebook o en la página web de la empresa, tal y como se muestra en la figura 20 esta

intención conta de freses de entrenamiento y un contexto y evento de “welcome” predefinido por Dialogflow.

Figura 20

Creación de la intención de bienvenida



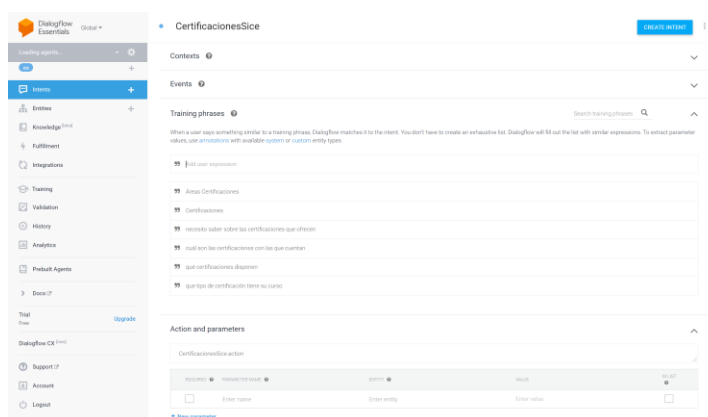
Nota. La figura muestra la configuración de la intención de bienvenida en Dialogflow. Autor.

(2022).

- Intención de “CertificacionesSice”: Para esta intención se configura únicamente las frases de entrenamiento y la acción que la activará, cabe recalcar que a todas las intenciones se les ha creado la acción, esto nos permitirá acceder a la intención desde el servidor de una manera más exacta.

Figura 21

Creación de la intención "CertificacionesSice" en Dialogflow

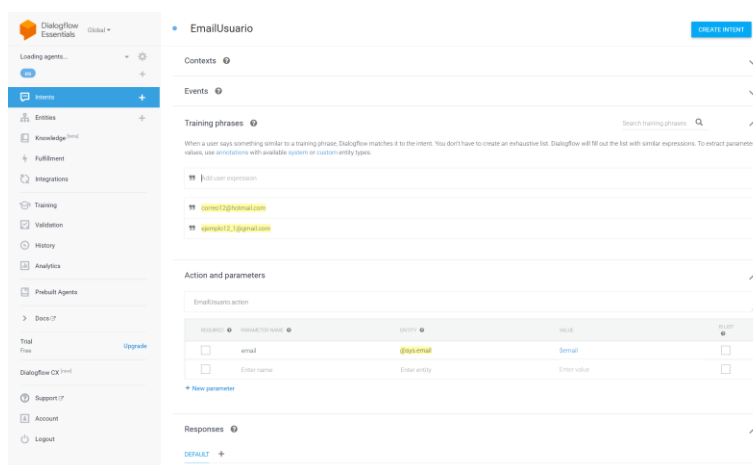


Nota. La figura muestra la configuración en Dialogflow de la intención que listará las certificaciones disponibles en la empresa SICE. Autor. (2022).

- Intención de "EmailUsuario": Esta intención contará con ejemplos de correo electrónico como frases de entrenamiento así como un parámetro propio de Dialogflow denominado "@sys.email" tal y como se muestra en la figura 22.

Figura 22

Creación de la intención en Dialogflow denominada "EmailUsuario"

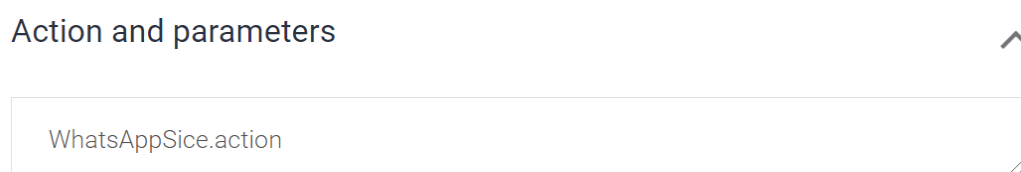


Nota. La figura muestra la configuración en Dialogflow de la intención que permitirá al Chatbot obtener el email del usuario. Autor. (2022).

En las figuras 19, 20 y 21, se detalla cómo es la creación y configuración de las intenciones, nuestro agente dispondrá de 20 intenciones para atender las frecuentes peticiones de los usuarios, la configuración para estas es la misma lo que cambia es las frases de entrenamiento y parámetros, y como se mencionó anteriormente lo único que contarán todas es con la acción configurada.

Figura 23

Configuración “action” en cada una de las intenciones



Nota. La figura muestra cómo se debe activar el “Action” en cada una de las intenciones del Chatbot, esto permite acceder desde el servidor a cada una de las intenciones, para nombrarlas en este proyecto se procedió a tomar el nombre de la intención más el punto “action”. Autor. (2022).

- Integración con API de Facebook

Para que el Chatbot funcione y se conecte tanto con el chat de Facebook Messenger y el chat de la página web primeramente se debe crear una App. En la figura 24 se detalla la configuración de nuestra app donde le hemos especificado un nombre “SiceBot”.

Figura 24

Configuración app de facebook developers

The screenshot shows the configuration interface for a Facebook app named 'VendedorBot'. The app ID is 500904677725893 and the type is 'Negocios'. The left sidebar contains navigation options: Panel, Configuración (expanded), Básica (selected), Avanzada, Roles, Alertas, Revisión de la app, Productos (with 'Agregar producto' link), Inicio de sesión con Facebo..., Webhooks, API de marketing, Messenger, Registro de actividad, and Registro de actividad.

The main configuration area includes the following fields and options:

- Identificador de la app:** 500904677725893
- Clave secreta de la app:** [Redacted]
- Nombre para mostrar:** VendedorBot
- Icono de la app:** A section for uploading a new icon (1024 x 1024 pixels). It shows the 'Icono actual' (a blue robot head) and a placeholder for the 'Icono nuevo'.
- Correo electrónico de contacto:** mauriciojumbocarrion@yahoo.es
- URL de la Política de privacidad:** https://sites.google.com/espe.edu.ec/politicas-chatbot/
- URL de Condiciones del servicio:** Condiciones del servicio del cuadro de diálogo de inicio de sesión ...
- Categoría:** Bots de Messenger para empres...
- Propósito de la app:** El propósito principal de esta app es acceder y usar datos de la plataforma de Facebook en nombre de...


Buttons for 'Descartar' and 'Guardar cambios' are located at the bottom right of the configuration area.

Nota. La figura muestra la configuración en Facebook developers de la app “SiceBot”. Autor. (2022).

Una vez realizadas las configuraciones básicas en la app de Facebook developers, se procede a generar el token de acceso que nos permitirá conectarnos a la página de Facebook de la empresa “SICE”, de igual manera en esta sección configuramos el o los Webhooks para el correcto funcionamiento de nuestro Chatbot.

Figura 25

Configuraciones de token y Webhooks para Facebook Messenger

Páginas ↑	Tokens
 BotTest 104197788551608	— Generar token
Agregar o eliminar páginas ⓘ	

Webhooks

Para recibir mensajes y otros eventos que envíen los usuarios de Messenger, la app debe tener habilitada la integración de webhooks.


URL de devolución de llamada:

Token de verificación:

Las solicitudes de validación y las notificaciones de webhook de este objeto se enviarán a esta URL.

Token que te enviará Facebook como parte de la verificación de la URL de devolución de llamada.

Editar URL de devolución de llamada
Mostrar errores recientes

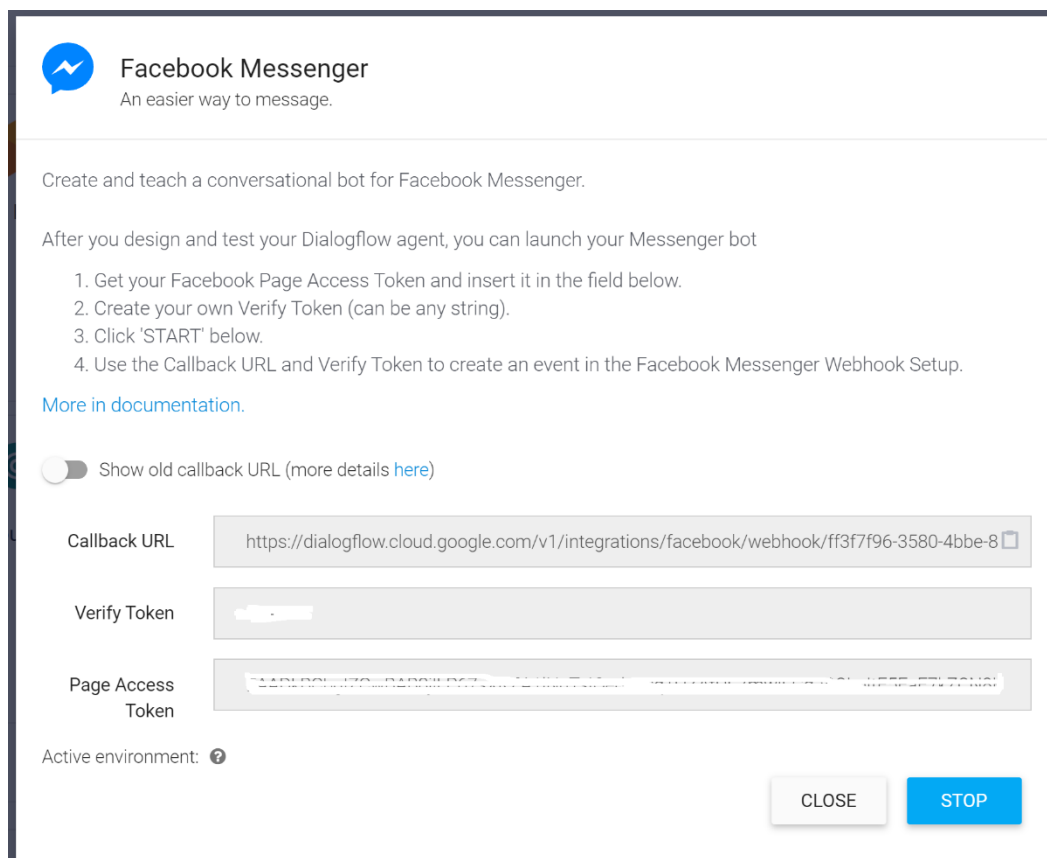
Páginas ↑	Webhooks
 BotTest 104197788551608	2 campos messages, messaging_postbacks Editar
Agregar o eliminar páginas ⓘ	

Nota. La figura muestra la configuración en Facebook developers de Messenger dentro de la app “SiceBot”. Autor. (2022).

Una vez generado el token en la app de Facebook developers se copia y se lo pega en la sección de integraciones de Dialogflow.

Figura 26

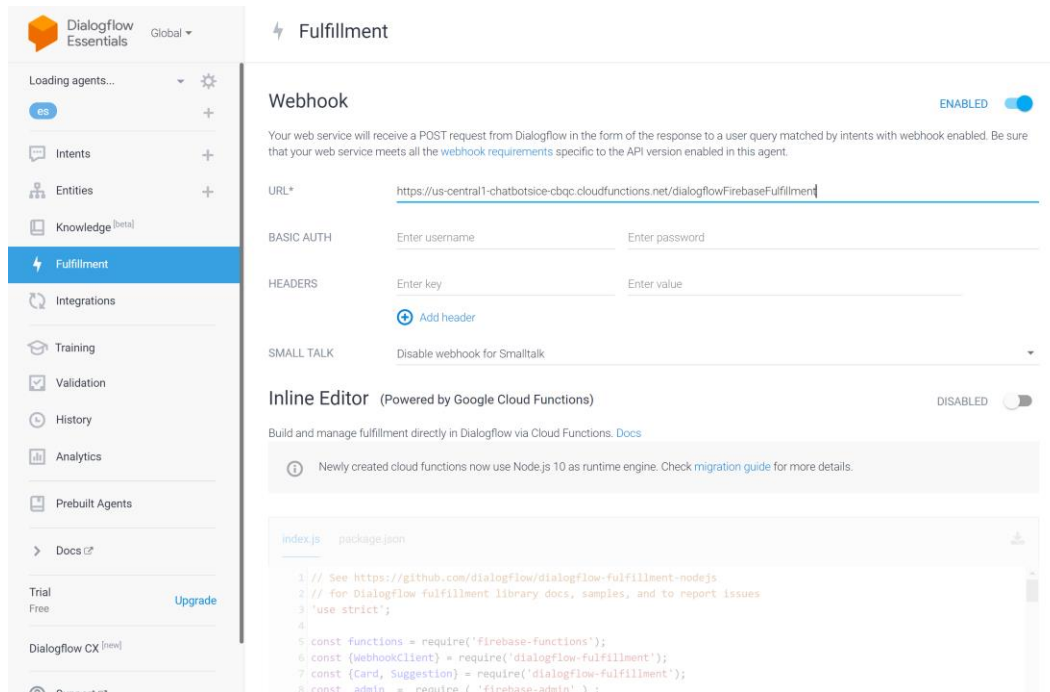
Integración Facebook en Dialogflow Integrations



The screenshot shows the 'Facebook Messenger' integration configuration interface. At the top, there is a Facebook Messenger logo and the text 'Facebook Messenger' and 'An easier way to message.' Below this, there is a section titled 'Create and teach a conversational bot for Facebook Messenger.' followed by instructions: 'After you design and test your Dialogflow agent, you can launch your Messenger bot'. A numbered list of steps follows: 1. Get your Facebook Page Access Token and insert it in the field below. 2. Create your own Verify Token (can be any string). 3. Click 'START' below. 4. Use the Callback URL and Verify Token to create an event in the Facebook Messenger Webhook Setup. A link 'More in documentation.' is provided. Below the instructions, there is a toggle switch for 'Show old callback URL (more details here)'. The 'Callback URL' field contains 'https://dialogflow.cloud.google.com/v1/integrations/facebook/webhook/ff3f7f96-3580-4bbe-8'. The 'Verify Token' field contains a masked string. The 'Page Access Token' field contains a masked string. At the bottom left, it says 'Active environment: ?'. At the bottom right, there are two buttons: 'CLOSE' and 'STOP'.

Nota. La figura muestra la configuración en Dialogflow integrations con Facebook a través del token generado en la figura 25. Autor. (2022).

La última configuración que se realiza en Dialogflow es la activación del Webhook de fulfillment tal y como se muestra en la figura 27, esto permite usar un Webhook externo el mismo como se especificó anteriormente esta desarrollado en Node.js y deployado en Heroku.

Figura 27*Activación Webhook en Dialogflow Fulfillment*

Nota. La figura muestra la activación del Webhook que se debe realizar en Dialogflow Fulfillment.

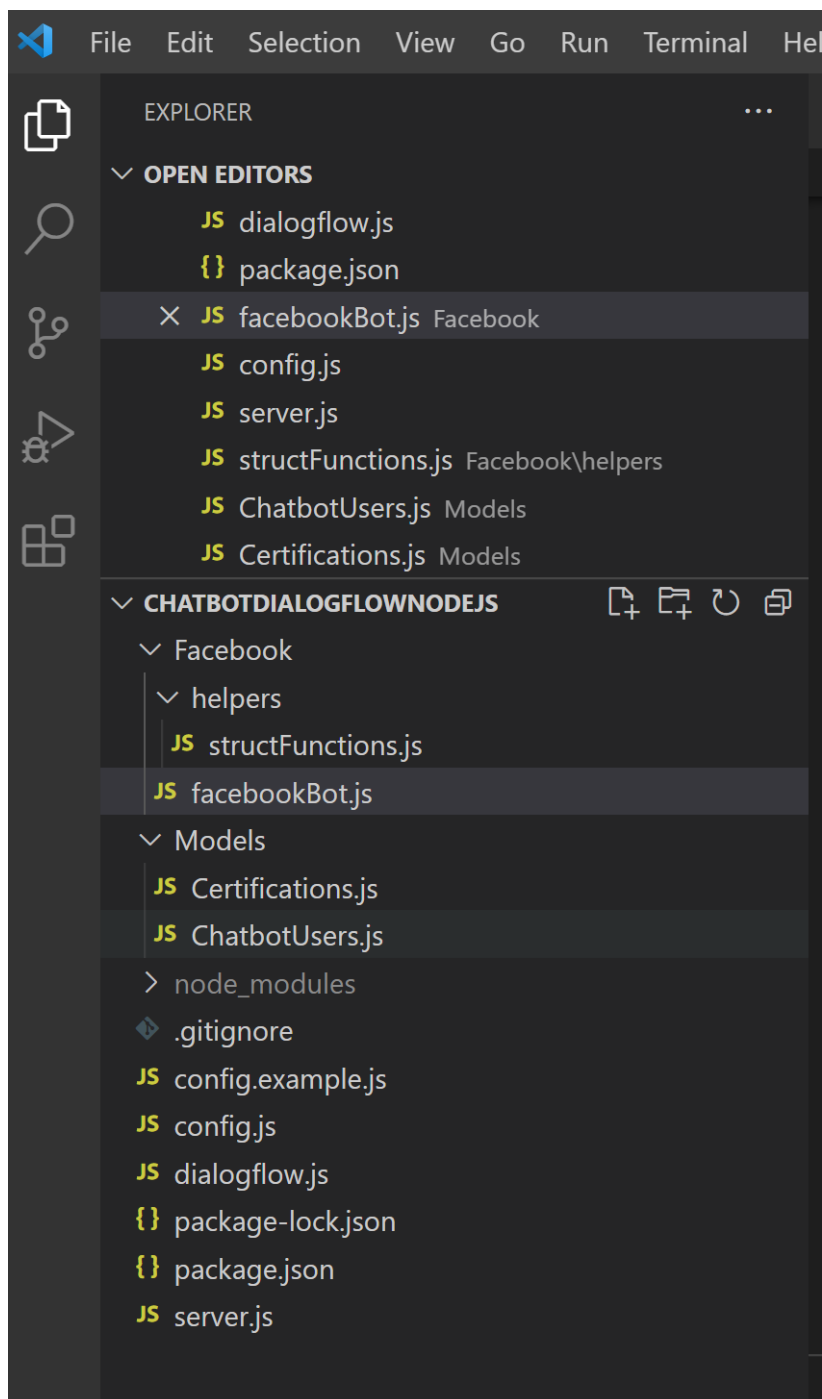
Autor. (2022).

- Webhook en Node.js

Creamos primeramente un proyecto vacío en Node.js, el editor que se usará para la creación del servidor será “Visual Studio Code”, el mismo que es de uso gratuito. Se ha denominado “ChatbotDialogflowNodeJs” mismo que tiene la siguiente estructura.

Figura 28

Estructura servidor en Node.js



Nota. La figura muestra la estructura del servidor en Visual Sudio Code. Autor. (2022).

Una vez estructurado el proyecto, lo siguiente es instalar las dependencias NPM.

Figura 29

Dependencias necesarias para ejecutar el proyecto en Node.js

```
"dependencies": {  
  "axios": "^0.20.0",  
  "body-parser": "^1.19.0",  
  "dialogflow": "^1.2.0",  
  "express": "^4.17.1",  
  "mongoose": "^5.12.13",  
  "node-telegram-bot-api": "^0.50.0",  
  "request": "^2.88.2",  
  "uuid": "^8.3.0"  
}
```

Nota. La figura muestra las dependencias necesarias para arrancar el proyecto en Node.js.

Autor. (2022).

El siguiente paso es configurar las credenciales tanto de Google api como las de del api de Facebook. Para esto primeramente creamos una interfaz la cual la denominamos config.js, donde definiremos cada una de las credenciales antes mencionadas.

Figura 30

Archivo de configuraciones de credenciales config.js

```
module.exports = {  
  ...  
  //Facebook App credentials  
  FB_PAGE_TOKEN: "123",  
  FB_VERIFY_TOKEN: "123",  
  FB_APP_SECRET: "123",  
  //Google project credentials  
  GOOGLE_PROJECT_ID: "123",  
  DF_LANGUAGE_CODE: "es",  
  GOOGLE_CLIENT_EMAIL: "123",  
  GOOGLE_PRIVATE_KEY: "123",  
};
```

Nota. La figura muestra la estructura de la interfaz que contendrá las credenciales de Google y Facebook. Autor. (2022).

En la figura 31, se establece la configuración del endpoint que se encargará de recibir los eventos de Facebook Messenger.

Figura 31

Configuración Servidor Express & Conexión MongoDB

```

server.js >
1  const express = require("express");
2  const bodyParser = require("body-parser");
3  const app = express();
4  const mongoose = require("mongoose");
5
6  const port = process.env.PORT || 3000;
7
8  // for parsing json
9  app.use(
10   bodyParser.json({
11     limit: "20mb",
12   })
13 );
14 // parse application/x-www-form-urlencoded
15 app.use(
16   bodyParser.urlencoded({
17     extended: false,
18     limit: "20mb",
19   })
20 );
21
22 mongoose.connect("mongodb://localhost:27010?retryWrites=true&majority=1",
23   {
24     useNewUrlParser: true,
25     useUnifiedTopology: true,
26     useFindAndModify: false,
27     useCreateIndex: true
28   },
29   (err, res) => {
30     if (err) return console.log("Hubo un error en la bd " + err);
31     console.log("BD online");
32   }
33 );
34 app.use("/messenger", require("./Facebook/FacebookBot"));
35
36 app.get("/", (req, res) => {
37   return res.send("👋👋👋");
38 });

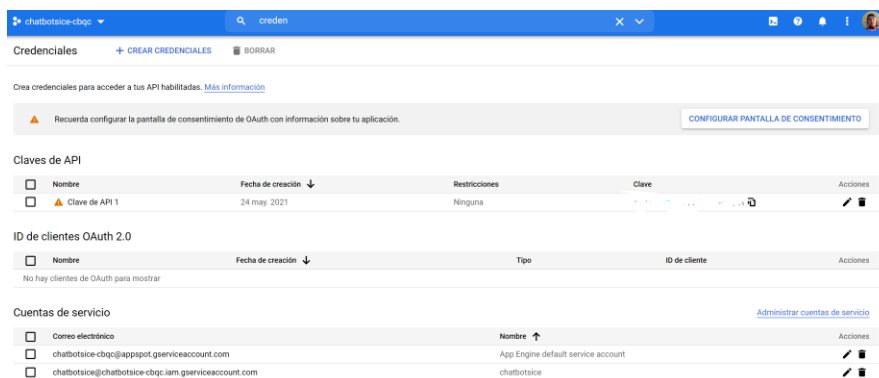
```

Nota. La figura detalla la conexión a los eventos de Facebook Messenger

Posterior a la configuración del servidor se debe realizar la configuración de la autenticación del proyecto en Google Project (Figura 32).

Figura 32

Configuración global en Google Cloud.



Nota. La figura muestra la configuración del proyecto en Google Cloud, en el cual estará montado el agente conversacional. Autor. (2022).

Dentro del archivo “facebookBot.js” se desarrolla la lógica, se hace el llamado a cada una de las librerías detalladas en la figura 29, dentro de este archivo “js” se evidencia todo el flujo; desde cuando Facebook Messenger envía a Dialogflow cada uno de los textos, estos son procesados y activan cada una de las intenciones en Dialogflow. También dentro de este archivo se interactúa con la base de datos.

Dentro del directorio “Models” se define cada una de los esquemas usados para la interacción con la base de datos MongoDB tal y como se muestra en la figura 33.

Figura 33

Definición esquemas (Esquema CertificationSchema)

```
const mongoose = require('mongoose');

const Schema = mongoose.Schema;
const ObjectId = Schema.ObjectId;

const CertificationSchema = new Schema({
  name: String,
  description: String,
  price: Number,
  status: String,
  type: String,
  url: String,
}, {
  versionKey: false
});

module.exports = mongoose.model('Certifications', CertificationSchema);
```

Nota. La figura se define es esquema de las certificaciones que se usa para la interacción con la base de datos. Autor. (2022).

Finalmente, en la figura 34 se observa cómo se define el método “handleDialogflowAction” que se encargará de procesar cada uno de los mensajes capturados en el Chatbot.

Figura 34

facebookBot.js en Node.js

```
async function handleDialogFlowAction(  
  sender,  
  action,  
  messages,  
  contexts,  
  parameters,  
  messageUser  
) {  
  console.log("USER MSN" + messageUser);  
  switch (action) {  
    case "RespuestaRapidaQuick.action":  
      sendQuickReply(sender, "Ejemplo de Quick",  
        [{  
          content_type: "text",  
          title: "Red",  
          payload: "<POSTBACK_PAYLOAD>",  
          image_url: "https://image.flaticon.com/icons/png/24/610/610333.png"  
        }, {  
          content_type: "text",  
          title: "Green",  
          payload: "<POSTBACK_PAYLOAD>",  
          image_url: "https://image.flaticon.com/icons/png/24/609/609780.png"  
        }]  
      );  
      break;  
    case "CertificacionesSice.action":  
      //Function from save the questions of users
```

Nota. La figura se muestra el método principal dentro del archivo facebookBot.js que se encarga de procesar los mensajes de los usuarios de la empresa SICE. Autor. (2022).

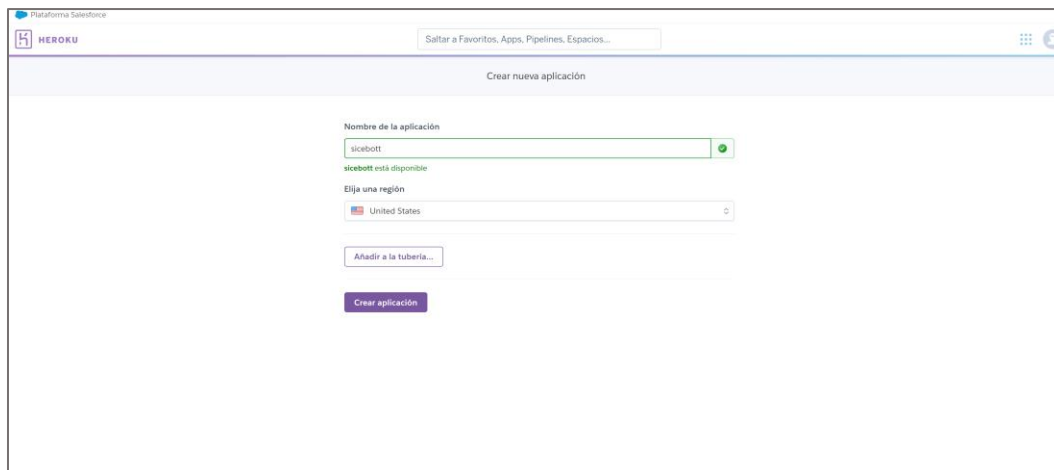
Las líneas de código usadas en este proyecto en Node.js se detallan completamente en la siguiente ruta de GitHub:

- Heroku

Una vez concluido con el desarrollo en nuestro proyecto, y dando cumplimiento a la arquitectura se procede a crear un proyecto en la consola de Heroku especificando el nombre y la región de nuestro servidor en la nube.

Figura 35

Creación y configuración servidor en la nube en Heroku.



Nota. Creación de la instancia en Heroku que servirá para alojar nuestro servidor Node.js.

Autor. (2022).

Para el despliegue en Heroku del servidor creado en Node.js se utilizó Github, esta herramienta para el control de versiones del código nos permitirá una integración continua del servidor alojado en la instancia.

Desarrollo del modelo en base a la metodología CRISP-DM

1. Comprensión del negocio

Esta etapa es la primera etapa del enfoque CRISP-DM, durante la cual se llevó a cabo una conversación con los gerentes de la empresa SICE para conocer más sobre el negocio y los objetivos a alcanzar con las técnicas de minería de datos de la aplicación.

2. Comprensión de los datos

Esta etapa abarca varios ítems, recopilación, descripción, exploración y finalmente la verificación de los datos.

2.1. Recopilación de los datos

Para iniciar con este proceso se procede con la adquisición de los datos necesarios para el desarrollo del estudio, los datos que son proporcionados por la empresa SICE son registros de clientes y ventas realizadas desde el año 2015 hasta el año 2022. Los datos recopilados se procedieron a categorizar de la siguiente manera:

Ventas: Son aquellos registros que contienen datos de las facturas en venta en el periodo mencionado anteriormente.

Clientes: Son datos personales de cada uno de los clientes de la empresa SICE.

2.2. Descripción de los datos

En este paso se procede a describir cada una de las colecciones en Mongo DB usadas para la aplicación de las técnicas de minería de datos.

Colección Invoice: Esta colección contiene los datos de las facturas realizadas por la empresa SICE entre los años 2017 – 2022, esta colección cuenta con un total de 3000 documentos.

Colección Customers: Esta colección contiene información necesaria de cada uno de los clientes de la empresa. Cuenta con un total de 800 documentos.

2.3. Exploración de los datos

Considerando el objetivo de la investigación, la segmentación y los patrones de compra de los clientes de la empresa SICE, se realiza una evaluación de los datos de todas las transacciones para tener más contexto de cómo se encuentra la actividad de ventas de la empresa.

2.4. Verificación de calidad de los datos

Dentro del análisis realizado se notó que la colección Customers contiene errores, específicamente en el campo, id ya que existen valores duplicados, así como también existen varias faltas de ortografía en los campos nombre y apellidos.

3. Preparación y muestreo de los datos

Dentro de esta etapa se llevan a cabo varias actividades como son la selección, limpieza, construcción, integración y el formato de los datos.

3.1. Selección de datos

Para seleccionar los datos a analizar se tomó en cuenta principalmente los objetivos del negocio.

Dentro de la colección “invoice” se seleccionaron los atributos invoiceNo, invoiceId, courseCode, description, customerId, name, lastname, invoiceDate, unitPrice, y quantity quedando un total de 2980 documentos.

En la colección “customers” se seleccionaron los atributos más sobresalientes y necesarios como son el customerId, name, lastName. La selección de documentos de clientes finales, ya que estos nos permitirán realizar un correcto análisis; de todos estos documentos se seleccionaron aquellos que tienen transacciones que van desde el año 2019 hasta el presente año 2022, quedando un total de 800 documentos.

3.2. Limpieza de la data

Dentro de esta etapa se llevó a cabo algunas tareas de limpieza de datos para ello se utilizó RStudio, primero se importó los datos de las colecciones customers y invoice, se descartaron algunos datos inconsistentes como por ejemplo los datos duplicados de los clientes.

Así mismo se descartaron datos irrelevantes para este estudio como por ejemplo estado civil y fecha de nacimiento de los usuarios de la empresa.

En cuanto a los registros de la colección invoice se procedió a realizar el análisis y limpieza, quedando finalmente con los siguientes campos: invoiceNo, invoiceId, courseCode, description, customerId, name, lastname, invoiceDate, unitPrice, y quantity.

Cabe recalcar que existen algunos datos atípicos mismo que corresponde a los datos que nos permiten aplicar la normalización RFM, en la siguiente sección se realizará los cálculos para obtenerlos. Una vez finalizada la limpieza de los datos se lograron obtener los datos suficientes para empezar con la aplicación de técnicas de minería de datos.

3.3. Creación de la nueva data

Dentro del análisis RFM, nos dice que se necesita de tres datos necesarios para llevar a cabo el proceso de clasificación de los clientes en base a sus preferencias de compra. Estos datos se detallan a continuación:

- Recencia: Para obtener este dato primero se obtiene la fecha de cuando el cliente realizo su última compra dentro de la empresa a este dato se le resta la fecha actual dando como resultado el número de días transcurridos.

Figura 36

Creación campo Recencia RFM

```
dplyr::group_by(CustomerID) %>%
  dplyr::summarise(Recency = Sys.Date() - max(InvoiceDate),
                  Recency = as.numeric(Recency),
```

Nota. La figura muestra el cálculo que se llevó a cabo en RStudio para obtener el dato de la recencia. Autor. (2022).

- Frecuencia: Para obtener este campo simplemente obtenemos el total de transacciones que cada cliente ha hecho dentro de la empresa.

Figura 37

Cálculo de la frecuencia RFM

```
frecuenci = dplyr::n_distinct(InvoiceNo),
```

Nota. La figura muestra el cálculo de la frecuencia en base a los datos de las facturas/transacciones por cliente. Autor. (2022).

Monto: Para obtener este dato simplemente se calculó el total del dinero que cada cliente ha gastado en la compra de cursos o certificaciones dentro de la empresa.

Figura 38

Cálculo del monto RFM

```
monitery = sum(total_dolar)) %>%
```

Nota. La figura nos detalla cómo se obtiene el valor del monto dentro de RStudio. Autor. (2022).

3.4. Integrar la Data

En esta etapa y una vez obtenido los campos de la recencia, frecuencia y monto se procedió a unirlos dentro de un mismo dataframe usando RStudio, este dataframe contiene los datos agrupados y con la información optima resultado de los comandos ejecutados anteriormente en RStudio y con los cuales empezaremos el análisis ya que cuenta con todos los datos necesarios.

Tabla 9

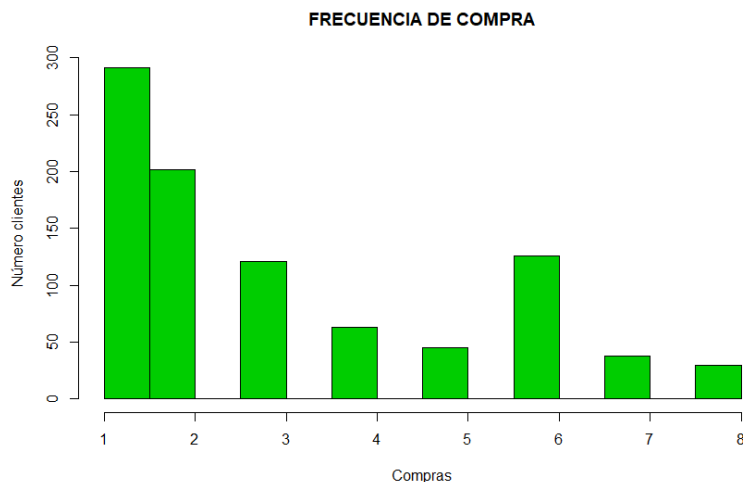
Estructura dataframe df_RFM

Atributo	Descripción
----------	-------------

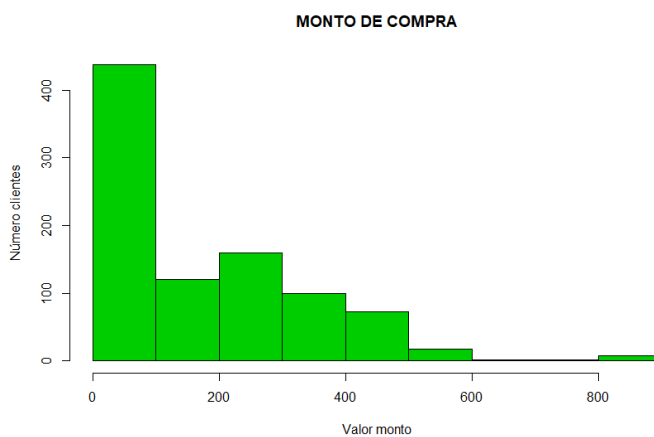
CustomerID	Atributo que identifica a cada cliente
Recency	Atributo que contiene los días que han transcurrido desde la última compra en la empresa SICE
Frequenci	Atributo que contiene la suma de las veces que un cliente ha comprado dentro de la empresa
Monitery	Atributo que contiene el total de dinero que un cliente ha gastado en la empresa.

3.5. Normalizar data en base al modelo RFM

Primeramente, se analizaron las tres variables del modelo RFM, esto para tener una idea de cómo se distribuyen estos datos, para ello se hizo uso nuevamente de la herramienta RStudio tal y como se muestra en las figuras 39, 40 y 41, esto nos permitirá crear histogramas y ver el comportamiento de la recencia, frecuencia y monto.

Figura 39*Histograma de frecuencia de compra*

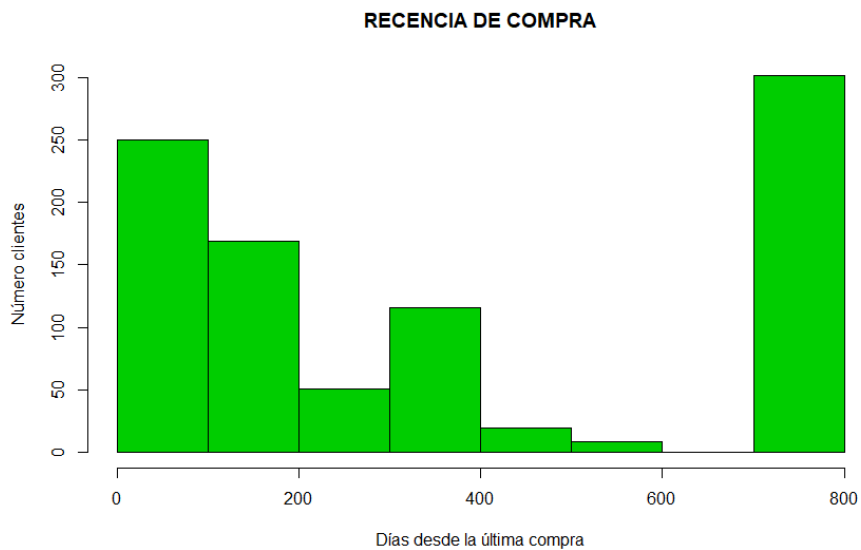
Nota. En este gráfico podemos observar que la mayoría de clientes ha comprado un curso o certificación, por otro lado, podemos ver que pocos clientes han comprado más de un curso.

Figura 40*Histograma de monto de compra*

Nota. En este histograma podemos visualizar que la mayor cantidad de clientes de la empresa ha consumido un monto bajo, mientras que son poco los clientes que han consumido un monto mayor.

Figura 41

Histograma recencia de compra

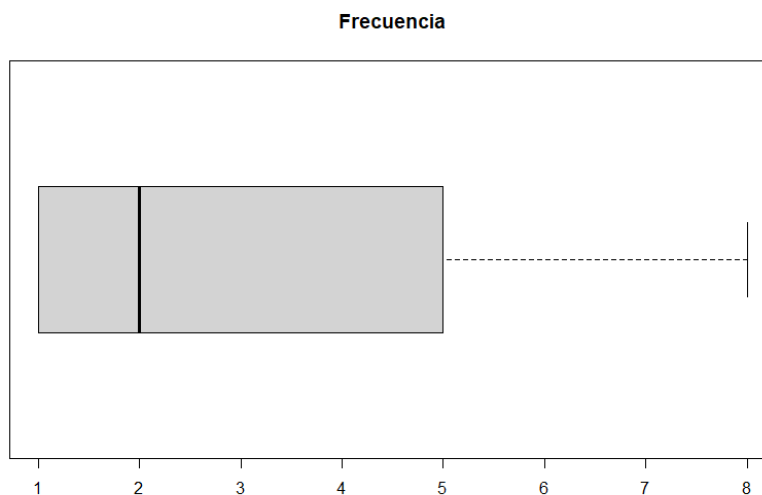


Nota. Podemos ver que en los últimos meses el consumo por parte de los clientes ha sido bajo con relación a años atrás. Autor. (2022).

A continuación, se muestra en las figuras 42, 43, 44 cada uno de los diagramas de caja que nos dan cuenta de los valores atípicos de cada una de las variables RFM usadas para este presente proyecto de investigación.

Figura 42

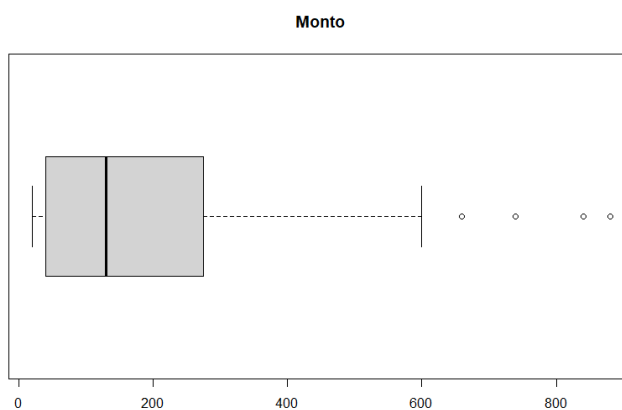
Diagrama variable frecuencia RFM



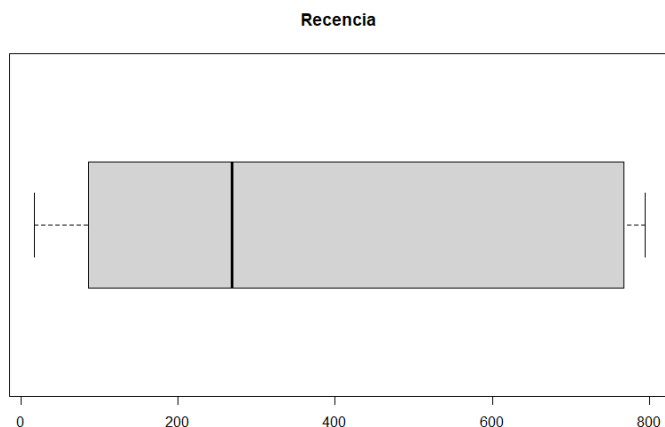
Nota. Este diagrama nos muestra ver que existen pocos valores atípicos dentro de la frecuencia de compra de los clientes de la empresa. Autor. (2022).

Figura 43

Diagrama variable monto RFM



Nota. En cuanto a la variable monto podemos visualizar que también existen ciertos valores atípicos. Autor. (2022).

Figura 44*Diagrama variable recencia RFM*

Nota. En esta figura podemos ver que la variable recencia no presenta valores en los sesgos de la distribución. Autor. (2022).

Una vez identificado los valores atípicos en cada una de las variables RFM se procedió a definir cada una de las escalas, para poder definir estas mismas se hizo uso del método de la codificación dura (J. McCarty y M. Hastack, 2007), el cual nos dice que para llevar a cabo este cálculo se toma en consideración al personal de la empresa que tenga conocimiento y experiencia. Una vez tomado en cuenta los criterios del personal de la empresa se ha definido las escalas que se detallan en la tabla 10.

Tabla 10*Puntuaciones en función de cuartiles*

Recencia	Frecuencia	Monto	Escala	Ponderación escala
(0, 150) días	5 - 6	(>450) \$	4	Alto
(151, 350) días	3 - 4	(150- 449) \$	3	Medio
(351, 500) días	1 - 2	(50 - 149) \$	2	Bajo

Recencia	Frecuencia	Monto	Escala	Ponderación escala
(501, +) días	<1	(0 - 49) \$	1	Muy Bajo

4. Modelado

En esta fase se debe o deben seleccionar los métodos de minería de datos que permitan dar una mejor solución al problema planteado y por consecuente la toma de decisiones por parte de la empresa SICE. Cabe recalcar que en algunos de los casos será necesario volver a la fase de Análisis desde esta fase.

Respecto a la segmentación de los clientes de la empresa SICE en base a los hábitos y comportamientos de compra, se tomarán en consideración las variables del modelo RFM.

Teniendo en consideración que en la actualidad hay varios algoritmos para realizar clustering, y que en la sección del estado del arte varios autores recomiendan algunos algoritmos que nos permitirán obtener los mejores resultados en base a la problemática de la empresa SICE se seleccionó dos algoritmos el A priori y el algoritmo K-means.

Una vez seleccionado los métodos a aplicar se procedió a generar cada uno de los algoritmos sobre las variables de recencia, frecuencia y monto.

Algoritmo K-means

Para empezar con la segmentación usando el algoritmo K-means es necesario y recomendable normalizar nuestro data frame esto con la finalidad de no tener valores muy grandes y otros muy pequeños, es decir acortar esas distancias y que se acerquen mucho más al cero, para eso ejecutamos uno de los tantos comandos que nos permite RStudio tal y como se muestra en la figura 45 y que nos devolverá un nuevo data frame con los mismo atributos pero con puntuaciones más fáciles de procesar, ver figura 46.

Figura 45

Normalización de las puntuaciones

```
#Normalizar las puntuaciones
df_RFM2 <-
  df_RFM %>%
  tibble::column_to_rownames(var = 'CustomerID') %>%
  scale %>%
  tibble::as_tibble()

summary(df_RFM2)
```

Nota. Comando para normalizar cada una de las puntuaciones de nuestro data frame original para así tener y trabajar con puntuaciones típicas. Autor. (2022).

Figura 46

Nuevo Data frame con datos normalizados

	Recency	frequenci	monitery
83	0.2239203	4.3256632	2.72307156
84	0.2239203	4.3256632	2.72307156
85	0.2239203	4.3256632	2.72307156
86	0.2239203	4.3256632	2.72307156
87	0.2239203	4.3256632	2.72307156
88	0.2239203	4.3256632	2.72307156
89	0.2239203	4.3256632	2.72307156
90	0.2239203	4.3256632	2.72307156
91	0.2239203	4.3256632	2.72307156
147	0.3449931	3.1441615	0.20896472
96	0.2844567	3.1441615	0.20896472
97	0.2844567	3.1441615	0.20896472
9	-2.5434574	3.1441615	4.23153567
10	-2.5434574	3.1441615	4.23153567
11	-2.5434574	3.1441615	4.23153567
133	0.3449931	1.9626598	-0.07836177
134	0.3449931	1.9626598	-0.07836177
135	0.3449931	1.9626598	-0.07836177

Nota. En la figura se puede apreciar cómo se ha generado el nuevo data frame una vez ejecutado el comando de la figura 45 este nos servirá para los cálculos subsiguientes. Autor. (2022).

Teniendo nuestro data frame listo lo siguiente es encontrar el número de clústeres ideal en base a los datos que se tiene, para ello ejecutamos una serie de comandos que nos permitirán obtener el mejor resultado, un paso anterior al cálculo del número de clústeres es recomendable visualizar la matriz de distancias para ver el patrón que toman los datos y si estos merecen un análisis.

Figura 47

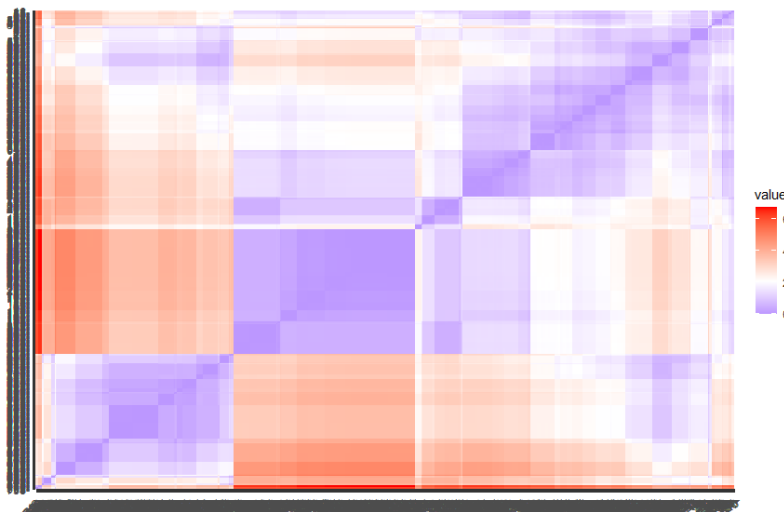
Código cálculo de la matriz de distancias RStudio

```
#Calcular la matriz de distancias
m.distancia <- get_dist(df_RFM2, method = "euclidean") #el método aceptado también puede ser: "maximum", "manhattan",
fviz_dist(m.distancia, gradient = list(low = "blue", mid = "white", high = "red"))
```

Nota. La figura muestra el fragmento de comandos que permiten primeramente calcular la matriz de distancias, y finalmente ver si los datos son susceptibles de aplicar un análisis tipo clúster.

Figura 48

Visualización matriz de distancias



Nota. La figura nos muestra la matriz de distancias, y podemos ver que en ciertos puntos (color rojo) existe ciertos clientes que comparten tendencias, y podemos concluir que si podemos seguir con el proceso de clústeres. Autor. (2022).

Lo siguiente como se menciona anteriormente es llevar a cabo el cálculo de número de clústeres óptimo, para ello dentro de RStudio usaremos tres métodos muy conocidos para la obtención del número de segmentos.

Figura 49

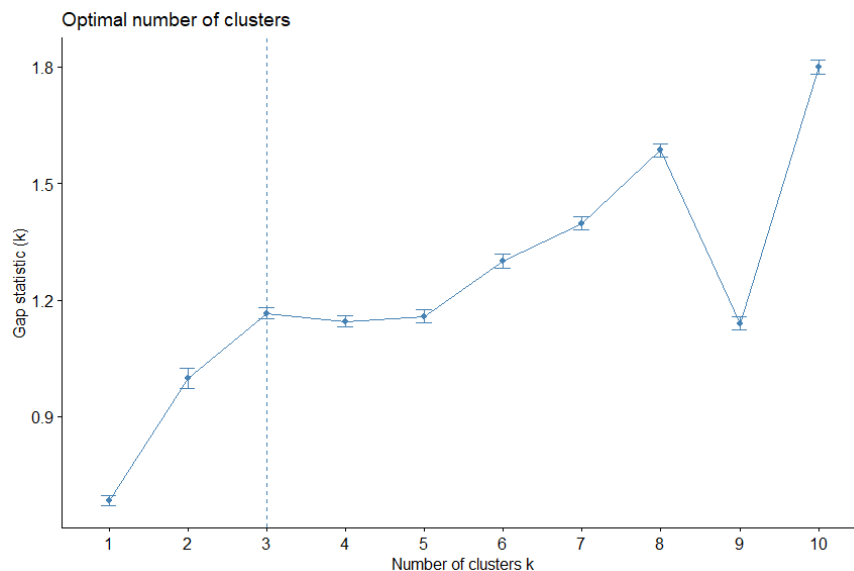
Cálculo de número de clústeres RStudio

```
#Estimar el número de clústers
fviz_nbclust(df_RFM2, kmeans, method = "wss")
fviz_nbclust(df_RFM2, kmeans, method = "silhouette")
fviz_nbclust(df_RFM2, kmeans, method = "gap_stat")
```

Nota. En la figura se detalla los comandos usados para el cálculo de número de clústeres, estas funciones reciben como parámetro nuestro data frame normalizado el algoritmo (k-means) y finalmente el método. Autor. (2022).

Figura 50

Número de clústeres óptimo RStudio



Nota. Una vez ejecutado los comandos para el cálculo de clústeres óptimo, podemos ver que el número sugerido es 3. Autor. (2022).

Para agrupar cada uno de los clientes en uno de los tres clústeres se usó la función de RStudio para el algoritmo K-means, dicha función viene integrada en la librería stats misma que debe ser instalada previamente. A continuación, en la figura 51 se detalla cómo se hizo uso de la función.

Figura 51

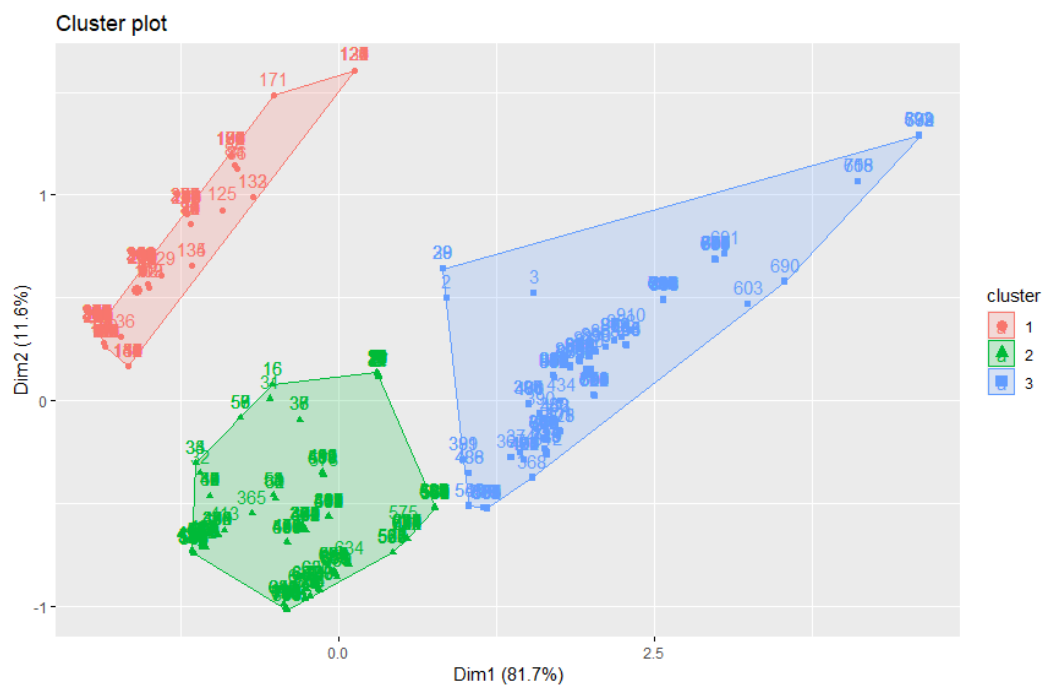
Aplicación algoritmo K-means

```
#Calculamos los tres clústers K-means|
clusterSICE <- kmeans(df_RFM2, centers = 3, nstart = 25)
clusterSICE
str(clusterSICE)
```

Nota. La función k-means de RStudio nos pide tres parámetros, en primer lugar, nuestro data frame, segundo el número de clústeres calculado anteriormente y finalmente nstart que es un valor por defecto y que se ha colocado el valor de 25 ya que este valor arroja la menor suma de error al cuadrado. Autor. (2022).

Cabe recalcar que la función K-means en RStudio también acepta dentro de sus parámetros el número máximo de iteraciones, pero para el presente trabajo de investigación hemos dejado el valor por defecto que es 10.

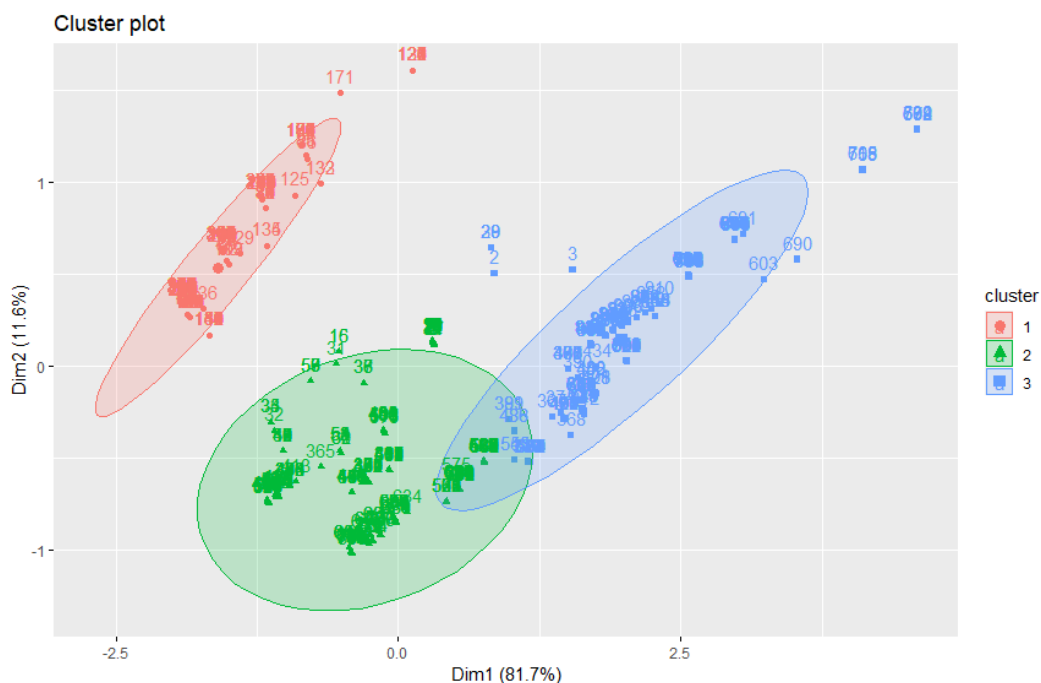
A continuación, en las figuras 52 y 53 se visualiza el resultado de ejecutar la función detallada en la figura 51:

Figura 52*Gráfico clústeres agrupados*

Nota. En la figura se muestra cada uno de los clústeres con sus respectivos datos, para ejecutar la gráfica se hizo uso de la función “fviz_cluster” de RStudio. Autor. (2022).

Figura 53

Gráfica 2 clústeres agrupados



Nota. Con la ayuda de la misma función “fviz_cluster” se logró obtener la gráfica que nos muestra la agrupación de cada uno de los clientes según su clúster. Autor.

(2022).

A si mismo se procedió a graficar el dendrograma con los resultados que nos arrojó K-means haciendo uso de la función de RStudio “fviz_dend” como se visualiza en la figura 54.

Figura 54

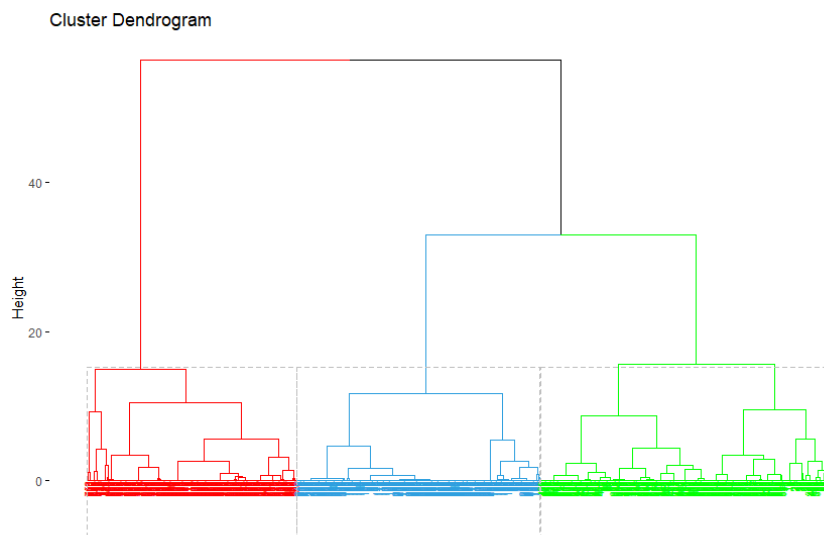
Función para la visualización del dendrograma resultante

```
#Graficamos el dendrograma de los clústeres
dendogramaSice <- hcut(df_RFM2, k = 3, stand = TRUE)
fviz_dend(dendogramaSice, rect = TRUE, cex = 0.5,
          k_colors = c("red", "#2E9FDF", "green", "black"))
```


Nota. En la figura se puede observar que en base a los datos calculados anteriormente podemos hacer uso de la función de RStudio “fviz_dend” misma que nos arroja el dendograma que se muestra en la figura 55. Autor. (2022).

Figura 55

Dendograma Sice



Nota. La figura nos muestra de una manera más detallada como es que se ubican cada uno de los clientes de la empresa SICE según el clúster al que pertenecen. Autor. (2022).

Para determinar los clientes más leales de la empresa SICE se procedió a calcular la distancia al punto cero de cada uno de los clústeres, los clientes con distancia cero más alejada son aquellos clientes que son más leales, mientras que los clientes con menos distancia al cero son los menos leales.

Cada clúster tiene un centroide para cada una de las variables RFM, existe una fórmula para el cálculo de los centros (Ching-Hsue & You-Shyang, 2009), tal y como podemos observar en la figura 56.

Tabla 11*Resumen clústeres K-means*

Número clúster	Recencia	Frecuencia	Monto	Distancia cero	Fidelidad
1	1.321	-0.622	-0.840	1.684	Bajo
2	-0.324	-0.547	-0.268	0.689	Medio
3	-1.006	1.253	1.167	1.986	Alto

En la tabla 10 podemos observar que los clientes que se ubiquen en el clúster 3 son los más fieles y que mejores resultados muestran para la empresa, por otro lado, el clúster 2 si bien no compran con frecuencia dentro de la empresa gastan una cantidad considerable de dinero, lo que permite que sean los clientes que están próximos a escalar al clúster ideal, finalmente tenemos el clúster 1 que son los que menos compran y tienen a su vez una frecuencia de compra muy baja.

Finalmente se procedió a crear un nuevo data frame mismo que contiene los datos personales de cada uno de los clientes junto con las variables RFM y a que clúster pertenecen con el objetivo de guardarlos en una nueva colección de mongoDB que denominamos "knowledge", en la figura 58 se muestra el nuevo data frame creado.

Figura 58*Data frame final k-means*

	CustomerID	recency	frequency	monetary	R_Score	F_Score	M_Score	RFM_Score	clus
1	CUS00056	521	1	200	1	2	3	123	1
2	CUS00057	521	1	200	1	2	3	123	1
3	CUS00058	521	1	200	1	2	3	123	1
4	CUS00059	521	1	200	1	2	3	123	1
5	CUS00132	737	3	185	1	3	3	133	1
6	CUS00133	737	3	185	1	3	3	133	1
7	CUS00134	737	2	130	1	3	2	132	1
8	CUS00135	737	2	130	1	3	2	132	1
9	CUS00125	744	3	125	1	3	2	132	1
10	CUS00173	770	5	100	1	4	2	142	1
11	CUS00032	476	1	90	2	2	2	222	1
12	CUS00033	494	1	90	2	2	2	222	1
13	CUS00034	494	1	90	2	2	2	222	1
14	CUS00035	494	1	90	2	2	2	222	1
15	CUS00039	435	1	90	2	2	2	222	1

Algoritmo Apriori

El segundo algoritmo empleado en el presente estudio investigativo es el algoritmo Apriori, este nos permitirá generar un conjunto de reglas de asociación, una regla de asociación tiene la forma $X \Rightarrow Y$, en donde X & Y son agrupaciones de elementos en nuestro caso elementos de curso y certificaciones de la empresa SICE. El nombre comúnmente con el que se denomina a X es antecedente mientras que a Y se la denomina consecuente, dando así el significado de que la presencia de X es una transacción implica la presencia de Y (Sunitha, Adilakshmi, & Swathi, 2014) (Liu & Shih, 2005) (Fathian & Reza Gholamian, 2010).

Para generar las reglas de asociación se aplicó el algoritmo Apriori que es el algoritmo que más se usa para este tipo de estudios. Este algoritmo nos permite descubrir conjuntos de elementos frecuentes y genera un grupo de reglas en base a un gran número de transacciones (Prasad, Raison, & Malik, 2011). Primero, cada elemento frecuente se identifica a través de todas las transacciones a analizar, luego se expande a un conjunto de elementos cada vez más

grande hasta que el conjunto de elementos resultante alcanza un umbral de frecuencia específico llamado soporte (Tang D. , 2014).

Dentro del algoritmo Apriori podemos identificar tres métricas importantes las mismas que se detallan a continuación:

- Soporte: Permite calcular el número de transacciones que ocurren juntas en los datos de cada elemento presente en la regla versus el número total de transacciones (Orozco, 2017).

$$\text{Soporte}(X \Rightarrow Y) = \frac{X \cup Y}{D}$$

- Confianza: la confianza es la probabilidad de que una transacción contenga X también contenga Y.

$$\text{Confianza}(X \Rightarrow Z) = \frac{\text{Soporte}(X \cup Z)}{\text{Soporte}(X)}$$

- Lift: Evalúa el nivel de dependencia entre los términos que componen la regla de asociación. Es decir $X \Rightarrow Z$, el lift viene a representar en que grado Z tiene a ser frecuente cuando X ocurra, o, al contrario.

$$\text{Lift}(A|B) = \frac{\text{Confianza}(A|B)}{\text{Soporte}(B)}$$

Para la aplicación del algoritmo Apriori se usaron los datos de las colecciones invoice, customers y courses, de esta forma la colección armada a partir de las colecciones mencionadas anteriormente quedó conformada de la siguiente manera: la columna InvoiceNo, InvoiceId, CourseCode, Description, Quantity, InvoiceDate, UnitPrice y CustomerId.

Figura 59

Data frame inicial Apriori

InvoiceNo	InvoiceId	CourseCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID
1	F000000001	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00001
2	F000000002	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00002
3	F000000003	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00003
4	F000000004	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00004
5	F000000005	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00005
6	F000000006	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00006
7	F000000007	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00007
8	F000000008	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00008
9	F000000009	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00009
10	F000000010	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00010
11	F000000011	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00011
12	F000000012	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00012
13	F000000013	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00013
14	F000000014	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00014
15	F000000015	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00015
16	F000000016	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00016
17	F000000017	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00017
18	F000000018	CC000001	Docencia Virtual I	1	2021-01-09	90	CUS00018

Nota. Data frame con el cual iniciamos el análisis aplicando el algoritmo Apriori. Autor. (2022).

Para la aplicación del algoritmo Apriori también se hace uso de la herramienta RStudio, empezamos cargando los datos a la R, una vez cargado todos los datos procedemos primeramente a remover los negativos del dataframe y dar formato de transacción al dataframe “sice_apriori” usando los comandos propios de RStudio.

Figura 60

Limpieza del dataframe *sice_apriori*

```
# Remoción de negativos y NA's
sice_apriori <-
  df_apriori %>%
  dplyr::filter(Quantity > 0 ,
               UnitPrice > 0) %>%
  tidyr::drop_na()

head(sice_apriori, n = 10)
str(sice_apriori)

# Damos formato de transacción a sice_apriori
# -----
library(plyr)

# Por cada transacción se crea una fila conteniendo todos los productos separandolos por una coma
# -----
Lista_cursos <- ddply(sice_apriori, c("CustomerID"), function(df3)paste(df3$Description, collapse = ","))

# Quitamos la columna transacción
# -----
Lista_cursos$InvoiceId <- NULL
```

Una vez realizada la limpieza del dataframe, se obtiene un nuevo dataframe el mismo que nos servirá para aplicar el algoritmo, este dataframe queda estructurado de la siguiente manera.

Figura 61

Data frame *Lista_cursos*

	CustomerID	V1
1	CUS00001	Docencia Virtual I,Docencia Virtual II
2	CUS00002	Docencia Virtual I,Docencia Virtual II
3	CUS00003	Docencia Virtual I,Docencia Virtual II
4	CUS00004	Docencia Virtual I,Docencia Virtual II
5	CUS00005	Docencia Virtual I,Docencia Virtual II
6	CUS00006	Docencia Virtual I,Docencia Virtual II
7	CUS00007	Docencia Virtual I,Docencia Virtual II
8	CUS00008	Docencia Virtual I,Docencia Virtual II
9	CUS00009	Docencia Virtual I,Docencia Virtual II
10	CUS00010	Docencia Virtual I,Docencia Virtual II
11	CUS00011	Docencia Virtual I,Docencia Virtual II
12	CUS00012	Docencia Virtual I,Docencia Virtual II
13	CUS00013	Docencia Virtual I,Docencia Virtual II
14	CUS00014	Docencia Virtual I,Docencia Virtual II
15	CUS00015	Docencia Virtual I,Docencia Virtual II
16	CUS00016	Docencia Virtual I,Docencia Virtual II

Nota. En esta figura podemos observar la estructura y datos del dataframe “Lista_cursos”, se visualiza que se agrupo los cursos que aparecen en las transacciones por cada uno de los clientes. Autor. (2022).

Con este dataframe podemos obtener los cursos más vendidos en la empresa, para ello usamos el comando “itemFrequencyPlot” de RStudio.

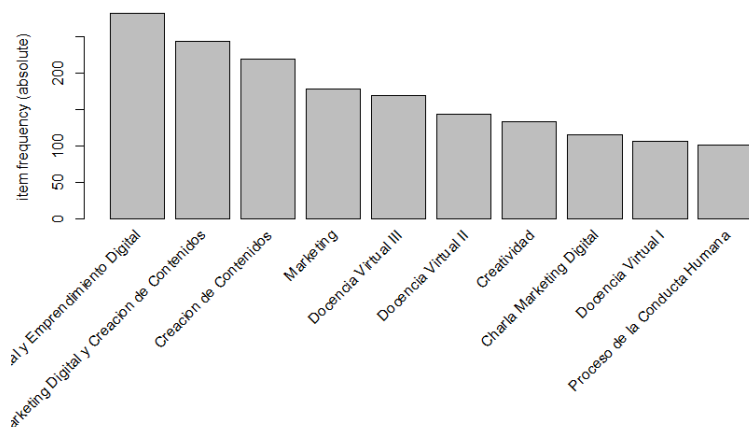
Figura 62

Comando para la creación del histograma de cursos más vendidos

```
# graficamos los cursos que más se han vendido en la empresa SICE
# -----
itemFrequencyPlot(transacciones, topN = 10, type = "absolute")
```

Figura 63

Histograma cursos más vendidos SICE



Lo siguiente es obtener cada una de las reglas de asociación para ellos hacemos uso del comando en RStudio “apriori” este recibe por parámetro el listado de transacciones y una lista con los valores tanto para el soporte como para la confianza.

Figura 64

Obtención reglas de asociación

```
# obtenemos las reglas de asociación por medio del algoritmo apriori
# -----
reglas <- apriori(transacciones, parameter = list(supp = 0.9, conf = 0.7, minlen = 2))
```

Una vez obtenidas las reglas, se procede a ejecutar una serie de comandos con el fin de obtener las reglas sin ningún tipo de duplicidad y redundancia como se muestra en la siguiente figura.

Figura 65

Limpieza de las reglas de asociación

```
# Revisamos si hay reglas duplicadas
# -----
duplicated(reglas)

# verificamos reglas redundantes
# -----
redundantes <- is.redundant(reglas)
redundantes

# which para ver cuales son las reglas redundantes
# -----
which(is.redundant(reglas))

# quitamos las reglas que son redundantes
# -----
reglas_podadas <- reglas[!redundantes]
```

Se despliega las reglas finales

Figura 66

Conjunto de reglas de asociación finales

```

> # Desplegamos las reglas finales
> # -----
> inspect(reglas_podadas)

```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Creación de Contenidos}	=> {Charla Marketing Digital}	0.06547619	1.0000000	0.06547619	15.272727	11
[2]	{Charla Marketing Digital}	=> {Creación de Contenidos}	0.06547619	1.0000000	0.06547619	15.272727	11
[3]	{Marketing Digital y Creación de Contenidos}	=> {Marketing Digital y Emprendimiento Digital}	0.07738095	1.0000000	0.07738095	12.000000	13
[4]	{Proceso de la Conducta Humana}	=> {Manejo del Estrés}	0.10119048	1.0000000	0.10119048	9.882353	17
[5]	{Manejo del Estrés}	=> {Proceso de la Conducta Humana}	0.10119048	1.0000000	0.10119048	9.882353	17
[6]	{Proceso de la Conducta Humana}	=> {Manejo de Estrés en Situaciones Complejas}	0.10119048	1.0000000	0.10119048	9.882353	17
[7]	{Manejo de Estrés en Situaciones Complejas}	=> {Proceso de la Conducta Humana}	0.10119048	1.0000000	0.10119048	9.882353	17
[8]	{Manejo del Estrés}	=> {Manejo de Estrés en Situaciones Complejas}	0.10119048	1.0000000	0.10119048	9.882353	17
[9]	{Manejo de Estrés en Situaciones Complejas}	=> {Manejo del Estrés}	0.10119048	1.0000000	0.10119048	9.882353	17
[10]	{Docencia Virtual II}	=> {Docencia Virtual I}	0.17857143	1.0000000	0.17857143	5.600000	30
[11]	{Docencia Virtual I}	=> {Docencia Virtual II}	0.17857143	1.0000000	0.17857143	5.600000	30
[12]	{Creación de Contenidos, Marketing Digital y Creación de Contenidos}	=> {Marketing}	0.05357143	1.0000000	0.05357143	12.923077	9
[13]	{Creación de Contenidos, Marketing Digital y Emprendimiento Digital}	=> {Marketing Digital y Creación de Contenidos}	0.05357143	1.0000000	0.05357143	12.923077	9
[14]	{Creación de Contenidos, Marketing Digital y Emprendimiento Digital}	=> {Marketing}	0.05357143	1.0000000	0.05357143	12.923077	9
[15]	{Charla Marketing Digital, Marketing Digital y Creación de Contenidos}	=> {Marketing}	0.05357143	1.0000000	0.05357143	12.923077	9
[16]	{Charla Marketing Digital, Marketing Digital y Emprendimiento Digital}	=> {Marketing Digital y Creación de Contenidos}	0.05357143	1.0000000	0.05357143	12.923077	9
[17]	{Charla Marketing Digital, Marketing Digital y Emprendimiento Digital}	=> {Marketing}	0.05357143	1.0000000	0.05357143	12.923077	9
[18]	{Marketing, Marketing Digital y Emprendimiento Digital}	=> {Marketing Digital y Creación de Contenidos}	0.06547619	1.0000000	0.06547619	12.923077	11
[19]	{Marketing Digital y Emprendimiento Digital}	=> {Marketing Digital y Creación de Contenidos}	0.07738095	0.9285714	0.08323333	12.000000	13
[20]	{Creación de Contenidos}	=> {Marketing}	0.05952381	0.9090909	0.06547619	11.748252	10
[21]	{Charla Marketing Digital}	=> {Marketing}	0.05952381	0.9090909	0.06547619	11.748252	10
[22]	{Creación de Contenidos, Marketing}	=> {Marketing Digital y Creación de Contenidos}	0.05357143	0.9000000	0.05952381	11.630769	9
[23]	{Creación de Contenidos, Marketing}	=> {Marketing Digital y Emprendimiento Digital}	0.05357143	0.9000000	0.05952381	10.800000	9
[24]	{Charla Marketing Digital, Marketing}	=> {Marketing Digital y Creación de Contenidos}	0.05357143	0.9000000	0.05952381	11.630769	9
[25]	{Charla Marketing Digital, Marketing}	=> {Marketing Digital y Emprendimiento Digital}	0.05357143	0.9000000	0.05952381	10.800000	9
[26]	{Marketing Digital y Creación de Contenidos}	=> {Marketing}	0.06547619	0.8461538	0.07738095	10.934911	11
[27]	{Marketing}	=> {Marketing Digital y Creación de Contenidos}	0.06547619	0.8461538	0.07738095	10.934911	11
[28]	{Marketing}	=> {Marketing Digital y Emprendimiento Digital}	0.06547619	0.8461538	0.07738095	10.153846	11
[29]	{Creación de Contenidos}	=> {Marketing Digital y Creación de Contenidos}	0.05357143	0.8181818	0.06547619	10.573427	9

Las reglas obtenidas nos servirán para hacer recomendaciones a cada uno de los clientes sobre un curso en específico mediante nuestro el uso del Chatbot implementado en la empresa SICE.

Capítulo IX

Implementación y análisis de resultados

Una vez finalizado el proceso del modelado en cuanto a la segmentación y la obtención de reglas de asociación para la sugerencia de cursos a los clientes de la empresa SICE, se procede a implementar los resultados obtenidos en el agente conversacional que se encuentra en los canales conversacionales de la empresa mismo que se encuentra detallado en el capítulo III de este documento. Para llevar a cabo la implementación inicialmente se procede a guardar la información obtenida en el modelado en nuestra base de datos no relacional MongoDB con la finalidad de que nuestro Chatbot obtenga la información al instante y pueda responder correctamente en base a las necesidades de cada uno de los clientes de la empresa. Cuando el agente conversacional ya se encuentre entrenado correctamente se realizará el contraste en algunos indicadores previos a la implementación de la propuesta.

En las tablas detalladas a continuación se muestra como quedó estructurada cada una de las colecciones producto del modelado.

Tabla 12

Estructura de la colección Knowledges

Atributo	Descripción
customerID	Atributo que identifica a cada documento dentro de la colección
recency	Atributo que contiene los días que han transcurrido desde la última compra en la empresa SICE

Atributo	Descripción
frequency	Atributo que contiene la suma de las veces que un cliente ha comprado dentro de la empresa
monetary	Atributo que contiene cuanto al dinero que un cliente ha gastado en la empresa
R_Score	Atributo que contiene la puntuación de recencia de compras
F_Score	Atributo que contiene la puntuación en cuartiles de la frecuencia de compras
M_Score	Atributo que contiene la puntuación en cuartiles del monto
cluster	Atributo que contiene a que cluster pertenece cada cliente.

Tabla 13

Estructura de la colección frequents

Atributo	Descripción
id	Atributo que identifica a cada documento dentro de la colección
rules	Atributo que contiene la regla obtenida en el algoritmo Apriori

Atributo	Descripción
support	Valor del soporte de la regla
confidence	Valor de la confianza de la regla
coverage	Valor del coverage
lift	Valor del lift de la regla
count	Valor de cuantas veces se repite la regla
leverage	Valor del leverage de la regla
conviction	Valor de la convicción de la regla

Ejecución

Una vez que se finalizó con la clasificación de los clientes en base a su comportamiento de compra usando el algoritmo de K-means, se propone a la empresa SICE llevar a cabo unas acciones en relación al marketing para cada uno de los segmentos descubierto en el modelado de los datos tales como los siguientes:

- Para los clientes que se encuentran en lo más alto es decir los que tiene una recencia, frecuencia y monto alto, sería importante lo siguiente:
 - Ofrecer programas de fidelización
 - Ofrecer nuevos cursos/certificaciones
 - Ofrecer descuentos especiales
 - Ejecutar encuestas de satisfacción
 - Ejecutar campañas exclusivas

Todo esto para de alguna u otra manera hacerle sentir a los clientes que se encuentran en este rango que la empresa los valora por toda la confianza y fidelidad con la marca.

- Para aquellos clientes que tengan una recencia, frecuencia y monto medio, tener en consideración lo siguiente:
 - Enfocar en ofrecerles cursos/certificaciones con mayor valor económico
 - Ejecutar acciones que permitan a estos clientes formar parte del grupo superior
 - Ofrecer descuentos atractivos

Ejecutar estas acciones con el fin de incentivar mucho más las compras dentro de la empresa, es decir lograr una mayor fidelidad.

- Finalmente, los que se encuentran en la parte inferior se recomienda ejecutar cada una de las siguientes acciones:
 - Ejecutar una campaña de precios agresiva
 - Ofrecer nuevos cursos/certificaciones
 - Ejecutar encuestas para determinar u identificar los problemas
 - Ofrecer descuentos

Estas acciones van encaminadas para que los clientes de este segmento realicen alguna compra y con el tiempo poder escalarlos al nivel siguiente.

Los datos almacenados del modelado en nuestra base de datos, se procedió a crear dentro de Dialogflow una nueva entidad que se la denominó “TipoCertificacion” la cual contiene todos los cursos y certificaciones que la empresa ofrece a sus clientes tal y como se muestra en la figura 67. Así mismo se creó un nuevo “intent” “CertificacionesSicelInfo” dentro de

Dialogflow se la entrenó con algunas de las frases más comunes identificadas y mediante las cuales se activará al momento que el cliente envíe su mensaje a través del chat, esta entidad recibe por parámetro la nueva entidad creada, esta configuración se detalla en la figura 68.

Figura 67

Entidad TipoCertificación

The screenshot displays the Dialogflow console interface for configuring an entity named 'TipoCertificación'. The left sidebar shows the navigation menu with 'Entities' selected. The main content area shows the configuration for this entity, including a table of values and their display names, and a set of configuration checkboxes.

Configuration options:

- Define synonyms
- Regex entity
- Allow automated expansion
- Fuzzy matching

Docencia Virtual I	Docencia Virtual I
Docencia Virtual II	Docencia Virtual II
Docencia Virtual III	Docencia Virtual III
Docencia Virtual IV	Docencia Virtual IV
Branding	Branding
Plan de Negocios	Plan de Negocios
Marketing Digital	Marketing Digital
Bioseguridad en Atención Odontológica y Manejo de Desechos	Bioseguridad en Atención Odontológica y Manejo de Desechos
Marketing Digital Fuera del País	Marketing Digital Fuera del País
Análisis de Datos	Análisis de Datos
Manejo del Estrés	Manejo del Estrés
Proceso de la Comunicación Humana	Proceso de la Comunicación Humana
Finanzas en Tiempo de Crisis	Finanzas en Tiempo de Crisis
Generalidades del Coronavirus	Generalidades del Coronavirus
Aula Virtual	Aula Virtual
Manejo de Software	Manejo de Software
Charla Marketing Digital	Charla Marketing Digital
Marketing Digital y Creación de Contenidos	Marketing Digital y Creación de Contenidos
Marketing Digital y Emprendimiento Digital	Marketing Digital y Emprendimiento Digital
Creación de Contenidos	Creación de Contenidos
Marketing	Marketing
Manejo de Estrés en Situaciones Complejas	Manejo de Estrés en Situaciones Complejas

Figura 68

Intención CertificacionesSiceInfo

The screenshot shows the Dialogflow console for the 'CertificacionesSiceInfo' intent. The 'Training phrases' section contains a warning: 'Template phrases are deprecated and will be ignored in training time. More details here.' Below the warning, there are five user expressions: 'me ayudas con información del curso de certificationName', 'información del curso de certificationName', 'me ayudas con información de certificationName', 'información de la certificación de certificationName', and 'por favor me ayudas con información de la certificationName'. The 'Action and parameters' section shows the 'CertificacionesSiceInfo action' with a table of parameters:

REQUIRED	PARAMETER NAME	ENTITY	VALUE	IS LIST
<input type="checkbox"/>	certificationName	@TipoCertificacion	Certificationname	<input type="checkbox"/>
<input type="checkbox"/>	Enter name	Enter entity	Enter value	<input type="checkbox"/>

A continuación, se desarrolló la lógica en nuestro servidor, esta lógica permite al Chatbot recibir los mensajes por parte del usuario este a su vez se conecta con el Webhook, se realiza la consulta a nuestras colecciones “frequents” y “knowledges” que contienen los patrones de compra y el segmento al cual pertenecen cada uno de los clientes de la empresa, con esta información el Chatbot puede responder adecuadamente según el requerimiento del cliente. En la figura 69 se muestra una interacción a través de la Fan page de la empresa SICE la cuál recibe una consulta por parte de un cliente acerca de un curso, el Chatbot recibe ese mensaje lo procesa e identifica la intención configurada en Dialogflow se le responde al cliente de acuerdo a como se encuentra configurado en nuestro servidor.

Figura 69

Chatbot SICE



En este caso Figura 69, el cliente solicitó información de un curso “Docencia Virtual I” el mismo que mediante el modelado con el algoritmo Apriori nos arrojó una regla que nos especificaba que: los clientes que compran “Docencia Virtual I” siempre compran también el curso de “Docencia Virtual II”, y viceversa, esta regla tiene un alto valor en la confianza.

Figura 70

Regla de asociación activada en el mensaje del cliente

```
[10] {Docencia Virtual II} => {Docencia Virtual I} 0.17857143 1.0000000 0.17857143 5.600000 30
[11] {Docencia Virtual I} => {Docencia Virtual II} 0.17857143 1.0000000 0.17857143 5.600000 30
```

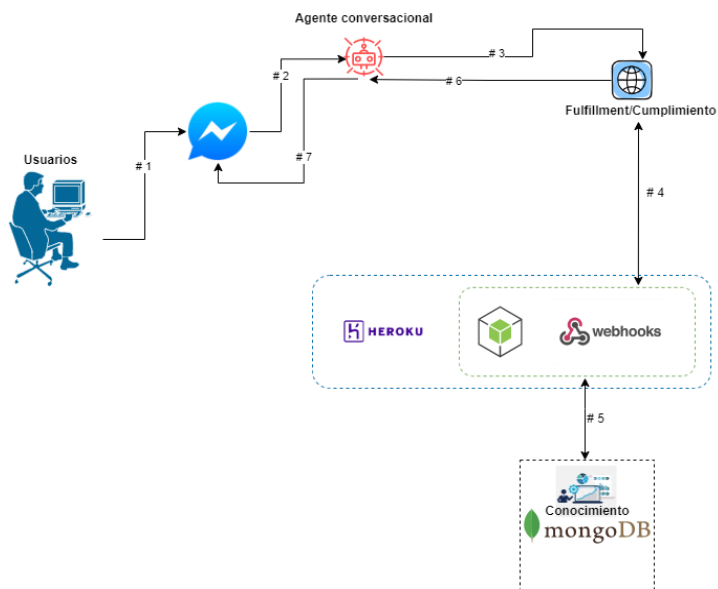
Como la información de cada una de las reglas de asociación se encuentra en nuestras colecciones de MongoDB, el Chatbot identifica el interés del cliente por parte de un curso o certificación en este caso en el curso de “Docencia Virtual I” y le recomienda el curso de “Docencia Virtual II” luego de un minuto de inactividad en el chat tal y como se ve en la figura 71. De esta manera el Chatbot administra cada uno de los mensajes que se recibe de una

manera correcta y por supuesto libera de esta actividad a los administradores tanto de la Fanpage y de la página web.

Figura 71

Sugerencia de compra al cliente en base al modelado con el algoritmo Apriori



Figura 72*Flujo final usuarios/Chatbot*

Nota. La figura muestra el flujo final para la sugerencia de compra de cursos y demás mensajes a cada uno de los clientes tanto a través de la Fan page como de la página web de la empresa.

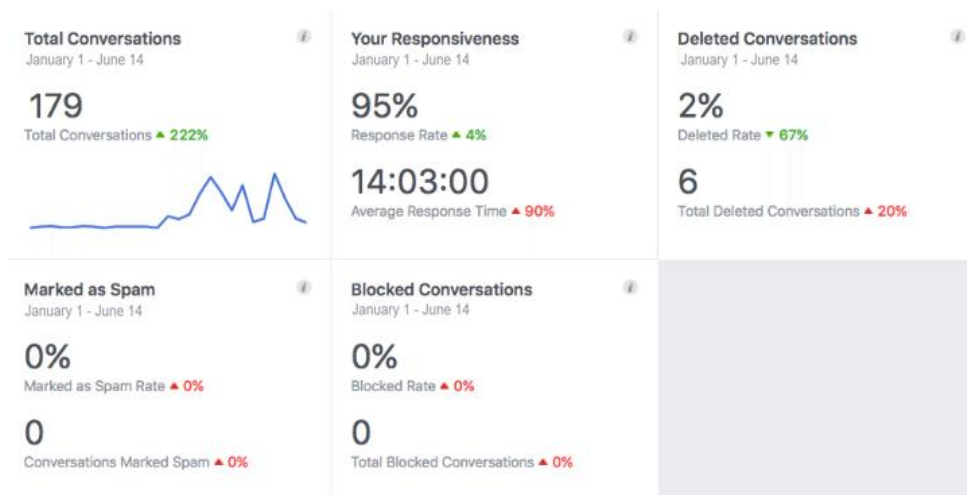
Autor. (2022).

Análisis de los resultados

Inicialmente se tenía un tiempo de espera muy largo en relación a cada uno de los mensajes que llegaban a la Fan Page de la empresa SICE, debido a que muchas de las veces los administradores de la empresa no contaban con el tiempo para responder inmediatamente a los mensajes de los clientes, el tiempo promedio de espera era de aproximadamente 14 horas.

Figura 73

Estadística Mensajes Fan Page SICE antes de aplicar los resultados



Nota. La figura muestra las estadísticas con respecto a los mensajes que llegan a la página de Facebook de la empresa SICE, en el periodo del 1 de enero al 15 de junio del 2021 en donde se comenzó a desarrollar el presente trabajo de investigación. Autor. (2022).

Una vez entrenado e implementado el Chatbot con los datos que nos arrojó cada uno de los modelos se puede apreciar en las mismas estadísticas de la página de Facebook de la empresa que el tiempo primeramente se redujo considerablemente, además las estadísticas en cuanto a número de seguidores, y alcance de la página aumentaron tal y como se muestra en las figuras 74, 75, y 76. Esto nos demuestra que la ejecución de minería de datos junto con la implementación del Chatbot cumple con los requisitos viéndose optimizado cada uno de los canales conversacionales de la empresa y reducido el tiempo de atención a cada uno de los clientes.

Figura 74

Estadística alcance de la página SICE

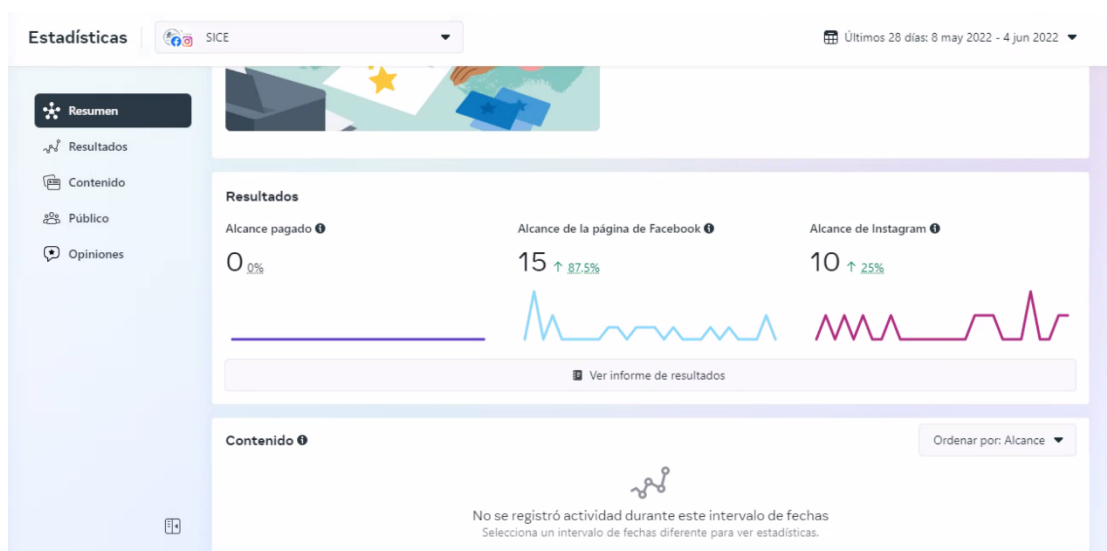


Figura 75

Estadística interacciones en la página SICE

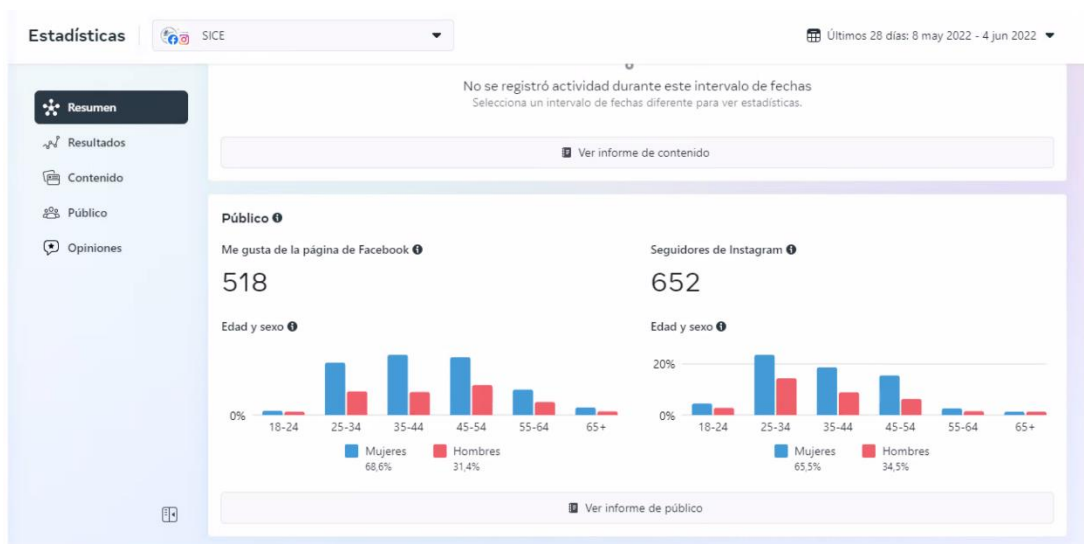
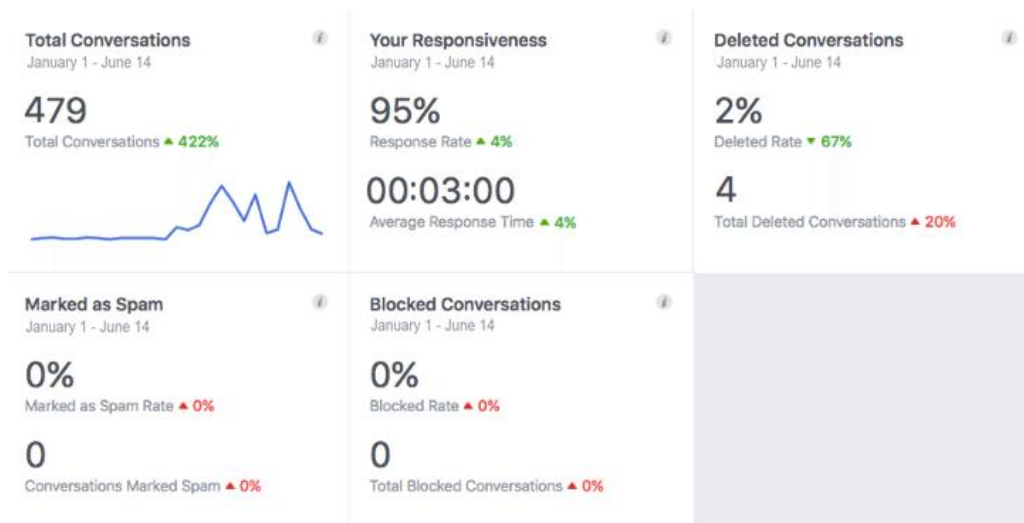


Figura 76

Estadística Mensajes Fan page SICE después de aplicar el modelo



Nota. La figura muestra las estadísticas con respecto a los mensajes que llegan a la página de Facebook de la empresa SICE, en el periodo del 1 de enero al 15 de junio del 2022 en donde se puede observar que se por un lado se ha aumentado el número de conversaciones, así mismo el tiempo de respuesta es muchísimo menor que cuando se inició con el presente trabajo investigativo, viendo así resultados positivos para la empresa. Autor. (2022).

Capítulo V

Conclusiones y recomendaciones

Conclusiones

El uso de la metodología CRISP-DM se ajustó perfectamente a cada una de las características del presente trabajo de investigación, ya que se obtuvo un modelo funcional mismo que se adapta a cada una de las exigencias que se habían planteado por parte de la empresa SICE.

Tomando en cuenta la revisión del estado del arte para la segmentación de clientes en base a técnicas de Minería de Datos, se procedió a construir con todos los datos proporcionados por la empresa el modelo denominado RFM (Recencia, Frecuencia y Monto), sobre este se aplicó el algoritmo de agrupación denominado k-means y el algoritmo de asociación Apriori.

Cada uno de los grupos de clientes de la empresa SICE que se obtuvieron luego de aplicar el algoritmo k-means en el presente trabajo de investigación, permitieron revelar tres niveles de lealtad: Alto, Medio y Bajo, con estos resultados la empresa puede elaborar estrategias de marketing focalizadas para cada grupo con el fin de retener la mayor cantidad de clientes.

Cada una de las recomendaciones de cursos encontradas luego de aplicar el algoritmo de asociación Apriori a cada una de las transacciones de la empresa, permitieron elaborar reglas de asociación mismas que le servirán a la empresa SICE para promocionar y recomendar un curso o certificación dependiendo los deseos de cada cliente.

La implementación del Chatbot en los canales conversacionales de la empresa SICE permitió a la empresa primeramente mejorar la comunicación con cada uno de sus clientes así

mismo gestionar de manera particular cada conversación en base a los resultados obtenidos en el modelado de los datos, es decir permitirá a la empresa recomendar cursos o certificaciones según los gustos o deseos de cada cliente.

El uso de la herramienta RStudio permitió una exploración, procesamiento, modelado y evaluación de los datos, esta herramienta proporciona una buena experiencia al llevar a cabo el proceso de Minería de datos ya que contiene muchas librerías y funciones ideales para este tipo de trabajos.

Recomendaciones

Para la ejecución del modelo obtenido se recomienda analizar primeramente cada una de las características técnicas con las que se cuenta en la actualidad en la empresa SICE, debido a que la aplicación del modelo demanda algunos cambios que pueden afectar a los sistemas que se encuentran funcionando, una vez analizado esto se puede determinar las mejores alternativas para la puesta en producción.

Se recomienda a la empresa SICE aplicar diferentes estrategias de marketing focalizadas a cada uno de los grupos de clientes identificados en el modelado, dando prioridad aquellos clientes que tengan un valor alto para la empresa, ya que según expertos es más difícil adquirir nuevos clientes que mantener a los buenos clientes.

Cuando se aplica minería de datos es recomendable usar al menos dos técnicas distintas ya que esto permitirá constatar y complementar los resultados obtenidos, aplicando este mecanismo se puede realizar una comparación mediante la cual se pueda seleccionar la técnica que más se adapte a las necesidades de la empresa u organización.

En relación con el Chatbot se recomienda que para versiones a futuro se pueda integrar hacia más asistentes virtuales como por ejemplo “Google Assistant”, esto permitirá hacer uso de

más funcionalidades dentro de la plataforma de Dialogflow, como por ejemplo la interpretación de mensajes de voz y texto. Además, se recomienda estar muy pendiente de visualizar el historial dentro de Dialogflow con la finalidad de observar conversaciones que hayan fallado y así poder corregirlas a tiempo.

Bibliografía

- Alpaydin, E. (2014). *Introduction to Machine Learning*.
- Andritsos, P. (2002). Data clustering techniques.
- Aplextm. (2015). *¿Por qué tener un sistema para atención de peticiones, quejas, y reclamos?* Obtenido de ¿Por qué tener un sistema para atención de peticiones, quejas, y reclamos?: <https://arandasoft.com/blog/por-que-tener-un-sistema-para-atencion-de-peticiones-quejas-y-reclamos/>
- Bellini Saibene, Y. N., & Volpacchio, M. (2014). *Fases del proceso de CRISP-DM Adaptado de 10_fig*. Obtenido de https://www.researchgate.net/figure/Fases-del-proceso-de-CRISP-DM-Adaptado-de-10_fig2_306959832
- Birant, D. (2011). Data Mining Using RFM Analysis.
- Brynjolfsson, E. (2017). What's Driving the Machine Learning Explosion?
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., . . . Trench, M. (2017). ARTIFICIAL INTELLIGENCE THE NEXT DIGITAL FRONTIER?
- Ching-Hsue, C., & You-Shyang, C. (2009). Classifying the segmentation of customer value via RFM model and RS theory.
- COGNIAPPS. (2016). *Historia de los Chatbots*. Obtenido de Historia de los Chatbots.: <https://medium.com/@cogniapps/historia-de-los-chatbots-bd71f3fd914a>
- Fathian, M., & Reza Gholamian, M. (2010). Mining important association rules based . *Data Analysis Techniques and Strategies*.
- Gordon S, L., & Berry, M. (2011). *Data Mining Techniques*.
- Guojun, G., Chaoqun, M., & Jianhong, W. (2007). *Data Clustering: Theory, Algorithms, and Applications*.
- Halkidi, M., Batistakis, Y., & Vazirg, M. (2001). On Clustering Validation Techniques.
- Hansen, A., & Mouritsen, J. (1999). Managerial Technology and Netted Networks. 'Competitiveness' in Action: The Work of Translating Performance in a High-Tech Firm.
- Hernández Orallo, J., Ramírez Quintana , M. J., & Ferri Ramírez, C. (2004). *Introducción a la minería de datos*.
- Huimin, L., Yujie, L., Min, C., Kim, H., & Seiichi, S. (2018). Brain Intelligence: Go beyond Artificial Intelligence.
- Jin, X., & Han, J. (2011). K-Medoids Clustering.
- Jindal, M., & Kharb, N. (2013). K-means Clustering Technique on Search Engine Dataset using Data Mining Tool.
- Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects.

- Kaufman, L., & Kaufman, L. (1990). *Finding Groups in Data: An Introduction To Cluster Analysis*.
- Kaufman, L., & Rousseeuw, R. (1990). *Finding Groups in Data: An Introduction To Cluster Analysis*.
- Liu, D.-R., & Shih, Y. (2005). «A CLV-Based Method for Product Recommendation.
- Manaure, A. (2017). *¿Cuáles son los tipos de chatbots que tu empresa necesita?* Obtenido de ¿Cuáles son los tipos de chatbots que tu empresa necesita?:
<https://thestandardcio.com/2017/10/19/cuales-los-tipos-Chatbots-empresa-necesita/>
- Miguel, G., Cuadrado, J. J., Sicilia, M., Rodríguez, D., & Rejas, R. (2007). Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software.
- Mitra, S., & Acharya, T. (2003). *Data Mining*.
- Molina López, J. M., & García Herrero, J. (2006). TÉCNICAS DE ANÁLISIS DE DATOS.
- MORELO TAPIAS, K. A. (2014). *Sistema para la caracterización de perfiles de clientes de la empresa zona T*.
- Mythili, S., & Madhiya, E. (2014). An Analysis on Clustering Algorithms in Data Mining.
- Node.js. (2022). *Acerca de Node.js*. Obtenido de <https://nodejs.org/es/about/>
- Orozco, M. (2017). Método de reglas de asociación para el análisis de afinidad entre objetos de tipo texto.
- Ospina Quintero, J. C. (2015). Introducción a los sistemas de recomendaciones.
- Pajares Martinsanz, G., & Santos Peñas, M. (2005). *Inteligencia artificial e ingeniería del conocimiento. (Spanish Edition)*.
- Pandey, A., & Shukla, M. (2015). Analysis And Implementation Of K-Mean And K-Medoids Algorithm For Large Dataset To Increase Scalability And Efficiency.
- Pandey, S., & Dubey, S. (2013). A Comparative Analysis of Partitioning Based Clustering Algorithms and Applications.
- Pareto, V. (1896). *Cours d'Economie Politique. Tome Premier, The Economic Journal*.
- Philip C, J. (2019). *Introduction to artificial intelligence*.
- Prakash, S., & Aarohi, S. (2015). PERFORMANCE ANALYSIS OF CLUSTERING ALGORITHMS IN DATA MINING IN WEKA.
- Prasad, P., Raison, G., & Malik, L. (2011). Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions.
- Sudipto, G., Rajeev, R., & Kyuseok, S. (2001). ROCK: A Robust Clustering Algorithm for Categorical Attributes.
- Sunitha, R., Adilakshmi, T., & Swathi, V. (2014). A NOVEL ASSOCIATION RULE MINING AND CLUSTERING BASED HYBRID METHOD FOR MUSIC RECOMMENDATION SYSTEM.

Tang, D. (2014). Essays on retail analytics and material information modeling.

Tang, Z., & MacLennan, J. (2005). *Data mining with SQL Server 2005*.

TECHTARGET. (2017). *Aprendizaje automático (machine learning)*. Obtenido de Aprendizaje automático (machine learning): <https://searchdatacenter.techtarget.com/es/definicion/Aprendizaje-automaticomachine-learning>

Tsai, C.-F., Hu, Y.-H., & Lu, Y.-H. (2013). Customer segmentation issues and strategies for an automobile dealership with two clustering techniques.