Implementación de Machine Learning (ML) para clasificación de PQRSF con minería de texto en el área de atención al usuario de la Universidad Católica Luis Amigó

Autor (es)

Daniela Gómez Sepúlveda,

Lisbed Jiménez Villa,

Luis Andrés Rivera Delgado

Especialización en Big Data e Inteligencia de Negocios

Facultad de Ingenierías y Arquitectura

Universidad Católica Luis Amigó

Medellín, Colombia

2022

Implementación de Machine Learning (ML) para clasificación de PQRSF con minería de texto en el área de atención al usuario de la Universidad Católica Luis Amigó

Autor (es)

Daniela Gómez Sepúlveda
Lisbed Jiménez Villa
Luis Andrés Rivera Delgado

Trabajo de grado, presentado como requisito para optar al título de:

Especialista en Big Data e Inteligencia de Negocios

Asesor:
PhD. Juan Camilo Giraldo Mejía

Especialización en Big Data e Inteligencia de Negocios
Facultad de Ingenierías y Arquitectura
Universidad Católica Luis Amigó
Medellín, Colombia
2022

(Dedicatoria)

"Este trabajo se lo dedico a mi madre y hermanas que con su apoyo y amor contribuyeron a sacar adelante este proyecto".

Daniela Gómez Sepúlveda

Quiero agradecer a todas las personas que han puesto de su parte para ayudarme a crecer
como persona, superar dificultades y celebrar los éxitos.

Lisbed Jiménez Villa

En primer lugar, le doy gracias especialmente a Dios y a mi familia por darme el tiempo y las fuerzas necesarias para la consecución de este logro. A todas las personas que de algún modo nos han aportado y han hecho posible la realización con éxito de este trabajo de grado. A los profesores que compartieron sus conocimientos.

Luis Andrés Rivera Delgado

Agradecimientos

Agradecemos especialmente a nuestras familias por el tiempo y la comprensión durante todo el proceso que estuvimos en el desarrollo de la Especialización.

Agradecemos también a la Escuela de Posgrados y a todos los docentes que participaron en este proceso de formación, en especial al Dr. Juan Camilo Giraldo Mejía y a Juan Sebastián Parra Sánchez, por habernos brindado su apoyo y conocimiento para realizar el trabajo de grado.

Tabla de Contenido

Tabla de Figuras

1.INTRODUCCIÓN

El propósito del presente trabajo está orientado en la implementación de la clasificación de PQRSF de la universidad Católica Luis Amigo aplicando Machine Learning (ML) bajo el marco de referencia de la metodología CRISP-DM en el sistema de atención, para la gestión de las PQRSF que actualmente tiene la Universidad, con el fin de mejorar los tiempos de respuesta mediante la automatización.

La importancia de un Sistema de Peticiones, Quejas, Reclamos, Solicitudes y Felicitaciones (en adelante PQRSF) radica en el rol que desempeña en la gestión de calidad en las organizaciones, ya que es una herramienta que permite conocer, tramitar y resolver las necesidades del buen servicio que se presta a la comunidad, al punto de convertirse en canal de comunicación formal entre actores del negocio, aprovechando los sistemas de información que los implementan y administran. (Jiménez y Cadena, 2015, p.1) expresan que en la actualidad es necesario utilizar sistemas que permitan responder activamente a las necesidades de los usuarios, siendo competitivos al satisfacer las expectativas de servicio que ellos esperan.

De acuerdo con Silva (2020), citado por Romero (2021), un PQRSF "es un soporte técnico del registro de las solicitudes entrantes realizadas por los usuarios, que sean consultas o incidentes, sin importar el canal que hayan utilizado para comunicarse, ya sea por correo electrónico, vía telefónica" (p.47). Lo que significa que, un sistema de PQRSF permite registrar las solicitudes de todos los clientes en un sistema integrado que tiene varias opciones para registrar todo tipo de solicitudes.

La Universidad Católica Luis Amigó actualmente dispone de un mecanismo de soporte (Portal Institucional) desde el cual se realiza la creación de tickets, se reciben y se gestionan los PQRSF; pero ante el incremento de solicitudes recibidas por esta vía, se han identificado reiteradas inconformidades en el proceso de creación y asignación de los PQRSF de los usuarios; asunto que afecta el proceso de atención a las personas y los tiempos de respuesta.

Ante esta problemática, se advierte que el sistema actual de PQRSF de la Universidad Católica Luis Amigó no cuenta con una lista de opciones apropiadas para la selección de la unidad institucional que debe dar respuesta al servicio solicitado y en consecuencia, sus tickets son imprecisos o llegan a divisiones que no están relacionadas

con el caso; incluso sucede que algunas personas desistan de realizar los tickets por no encontrar la unidad institucional habilitada para atender la solicitud en forma correcta; ya que en la lista desplegable del campo que debe diligenciar el usuario no están todas las unidades de servicio al cliente. Por ello, todas las unidades institucionales deben ser incluidas en el campo de atención al usuario para que obtenga la atención y respuesta oportuna a sus PQRSF, con el fin de fortalecer un buen servicio institucional.

Este proyecto da respuesta a la problemática descrita que se vive en la Institución a partir de la utilización de las metodologías y técnicas empleadas en Big Data. Peñaloza y Vargas (2019) definen esta metodología como una forma evolucionada y digital en la que se procesan y gestionan grandes cantidades de datos que se convierten en conocimiento (p.13). De acuerdo con los conocimientos adquiridos en el desarrollo de la Especialización en Big Data e Inteligencia de Negocios, se aplicaron mejoras prácticas a la herramienta para permitir el análisis e interpretación de grandes volúmenes de datos. Aspectos que dan cuenta de la viabilidad de este proyecto.

Para el desarrollo de esta propuesta se ha diseñado el siguiente contenido. En el capítulo 1 el lector encuentra la caracterización del proceso de PQRSF de la Universidad Católica Luis Amigo; en el Capítulo 2 se explica la aplicación de la técnica de ML que permite definir la clasificación de PQRSF y en el capítulo 3 se prueban los resultados de la automatización de las PQRSF, aplicado en el contexto de estudio.

2.MOTIVACIÓN

En la actualidad es importante que todas las empresas y organizaciones tengan la mayor cantidad de datos cuantificados y organizados; esta es una tarea que les permite conocer los tiempos de respuesta en las PQRSF, en este caso de la Universidad Católica Luis Amigó, con el fin de ser eficiente al dar respuesta a los usuarios de los servicios educativos que presta a las comunidades. Sin olvidar que un buen servicio requiere en ocasiones aumentar el personal que se requiere para dar soluciones en un momento oportuno y eficaz.

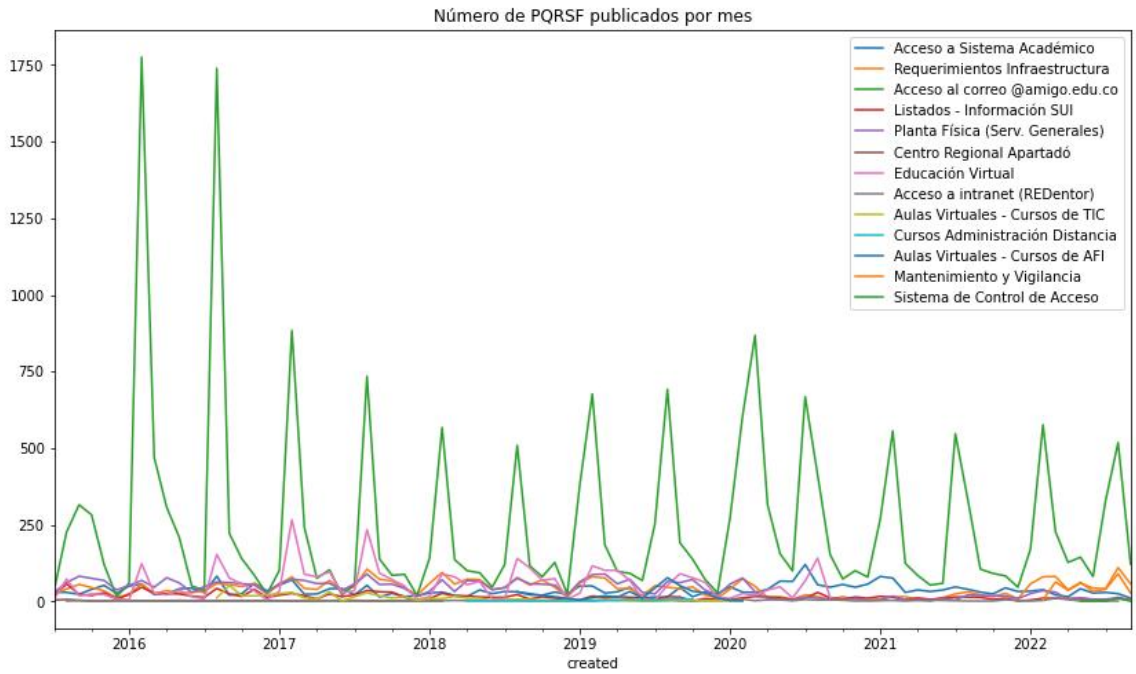
La Universidad está en el proceso de solicitar al Ministerio de Educación Nacional la Acreditación de Alta Calidad, lo que implica que todos los servicios que presta a la comunidad deben ser de calidad en referencia con todos los procesos administrativos que se articulan para posibilitar el desarrollo de las funciones sustantivas de la Educación Superior: docencia, investigación, extensión y servicios a la comunidad, bienestar e internacionalización. De ahí la importancia de ofrecer una excelente atención a los usuarios del servicio educativo que la Universidad ofrece a la comunidad; una de las formas de prestar un buen servicio consiste en tener un buen sistema de atención a los PQRSF, en el que queden documentados los tiempos de respuesta y los niveles de satisfacción de los usuarios.

En ese sentido, este proyecto de mejoramiento del sistema de PQRSF de la Universidad Católica Luis Amigó es una oportunidad de mejora del servicio que ofrece a los usuarios. Proyecto que se deriva de la aplicación de los conocimientos que hemos adquirido en la Especialización de Big Data e Inteligencia de Negocios, tales como: proceso de aplicación de métodos de Clusterización (clasificación) y de análisis predictivo que permiten optimizar los tiempos de atención al usuario al mejorar la administración y clasificación de PQRS en el sistema al dar una correcta asignación de los mismos a cada área responsable.

En el primer semestre de 2022 la Universidad Católica Luis Amigo recibió 4214 PQRSF en el sistema de atención al usuario, se transfirieron 1302, de los cuales se vencieron 196 y se cerraron 4018, de ellos se reabrieron 28, de acuerdo con el nivel de servicio (SLA) que está definido en 24 horas para los casos de consultas de información existente en los medios de información de la Universidad y 48 horas para los que requieren

de mayor consulta con las diferentes áreas (información entregada por el área de atención al usuario luego de revisar sus cifras e indicadores). Ver anexo 1: Consulta SQL del sistema de PQRS.

Figura 1. Estadísticas del sistema de atención al usuario.



Fuente: Elaboración propia

Nota. El gráfico contiene las estadísticas del histórico del sistema de tickets.

Las PQRSF son un instrumento o insumo importante para las instituciones, en tanto, que sirven de referencia no sólo para estimar la satisfacción de los usuarios con respecto al servicio prestado por la Universidad, sino también que el sistema se convierte en una manera de clasificar las solicitudes y direccionarlas al personal encargado, con el fin de tener una respuesta ágil por una persona experta en el tema.

En este trabajo se emplearon diferentes técnicas de Big Data que permiten la clasificación de las PQRSF que se reciben en la institución, adicionalmente, se implementaron métodos de Clusterización y Minería de Texto, que facilitan al usuario registrar su solicitud de manera ágil para la asignación del ticket.

Por otra parte, las técnicas de Data Mining y Machine Learning son muy prácticas para la clasificación automatizada de PQRSF en sistemas como atención al usuario, estos

ayudan a la mejora de los tiempos de respuesta, lo que permite al usuario categorizar su requerimiento y mejorar la experiencia creando conocimiento por medio de datos utilizando métodos de análisis. (Paramesh & Shreedhara, 2019. p.344).

En la revisión de literatura se encontraron algunas instituciones que utilizan Help Desk, Mesa de Ayuda o Ticket IT para solucionar los requerimientos de TI (Tecnología e Información); entre ellos Zicari, P., Folino, G., Guarascio, M., & Pontieri, L. (2021) de la Universidad alemana de Jordania, quienes proponen un sistema de help desk basado en machine learning para la gestión de servicios de TI con el fin de mejorar el rendimiento del personal de TI para hacer frente a los problemas de los tickets técnicos. Por lo anterior, es importante implementar herramientas que permitan gestionar solicitudes de PQRSF, que permitan una mejor comunicación con el usuario en términos de eficiencia y efectividad, que sean funcionales y que permitan la unificación de criterios y la satisfacción de los usuarios en tiempos de respuesta.

Además, la minería de datos le permite depurar, hallar anomalías, patrones, ruido y redundancia en grandes conjuntos de datos, que nos permitieron entender cuáles son relevantes y cuáles no, lo que permite emplearlos en una amplia variedad de técnicas y hacer el aprovechamiento de la información para evaluar posibles resultados y acelerar la toma de decisiones. Sin embargo, hay muchos desafíos en la implementación de Minería de Datos para la clasificación de PQRSF, como es el manejo de los datos ruidosos no estructurados donde el rendimiento de clasificación varía directamente en relación con el algoritmo de clasificación y el conjunto de datos.(Paramesh & Shreedhara., 2019, p.6).

3.PLANTEAMIENTO DEL PROBLEMA

De acuerdo con el análisis realizado en la Universidad Católica Luis Amigó frente al servicio de PQRSF se identificó que existe una insatisfacción de los usuarios, específicamente en los tiempos de respuesta e imprecisiones en las devoluciones que obtienen, como consecuencia de equivocadas e imprecisas marcaciones de PQRSF dirigidas a unidades que no corresponden, algunas solicitudes sin suficiente claridad, etc., que ocasionan reprocesos, mala clasificación de requerimientos y finalmente mala atención.

De acuerdo con Becerril y Villa (2018), es importante reestructurar un sistema de quejas y reclamos, de tal forma que ayude a fortalecer el enfoque hacia el usuario, controlando, suministrando y asegurando la calidad y al mismo tiempo logrando la motivación del personal a través de la optimización de las habilidades laborales que permitan brindar un servicio acorde a las necesidades y expectativas de los usuarios (p.1).

De este modo se busca mejorar la clasificación de las PQRSF por medio de la minería de texto con el fin de facilitar la tarea del usuario, optimizando los tiempos de respuesta. La tecnología avanza cada vez más y por ello en la actualidad se plantea la necesidad de resolver situaciones como las que se han descrito a la hora de hacer uso de estos mecanismos de servicio en las instituciones.

De acuerdo con la revisión de la literatura, existen múltiples debates de cómo diseñar un sistema de clasificación de PQRSF, para lo cual nos enfocaremos en buscar una técnica de ML con minería de texto y algoritmos de aprendizaje supervisado, empleando un conjunto de datos etiquetados que permitan automatizar la clasificación; lo anterior, aborda varios temas: asignación, priorización, identificación de requerimientos duplicados, PQRSF mal descritos y cambios al interior de la organización; así como nuevos productos, servicios, reestructuraciones que pueden dar pie a una nueva categoría de clasificación, que no estaban establecidas en el entrenamiento previo.

Predominantemente, para la clasificación y etiquetado en el sistema de PQRSF, se utiliza la técnica de representación de texto de PQRSF TF-IDF (Term Frequency - Inverse Document Frequency) y modelos kNN, Naive Bayes, (hubness-aware classifiers), árboles de decisión, regresión logística, máquinas de vectores de soporte, redes neuronales. En algunos modelos emplea el procesamiento del lenguaje natural o PNL, que es la

conjunción de diferentes métodos mediante los que se realiza un análisis sintáctico y semántico (Riefer, Ternis & Thaler, 2016, p. 4).

El Sistema IT es necesario para la mejora de la interacción que tiene el usuario con la interfaz, ya que esta es poco amigable y en muchos casos se desiste de escalar la solicitud personal al encontrarse con demoras en los tiempos de respuesta, situación que se puede relacionar causalmente con la falta de personal y este a la vez puede ocasionar respuestas imprecisas y demoras en los tiempos de respuesta.

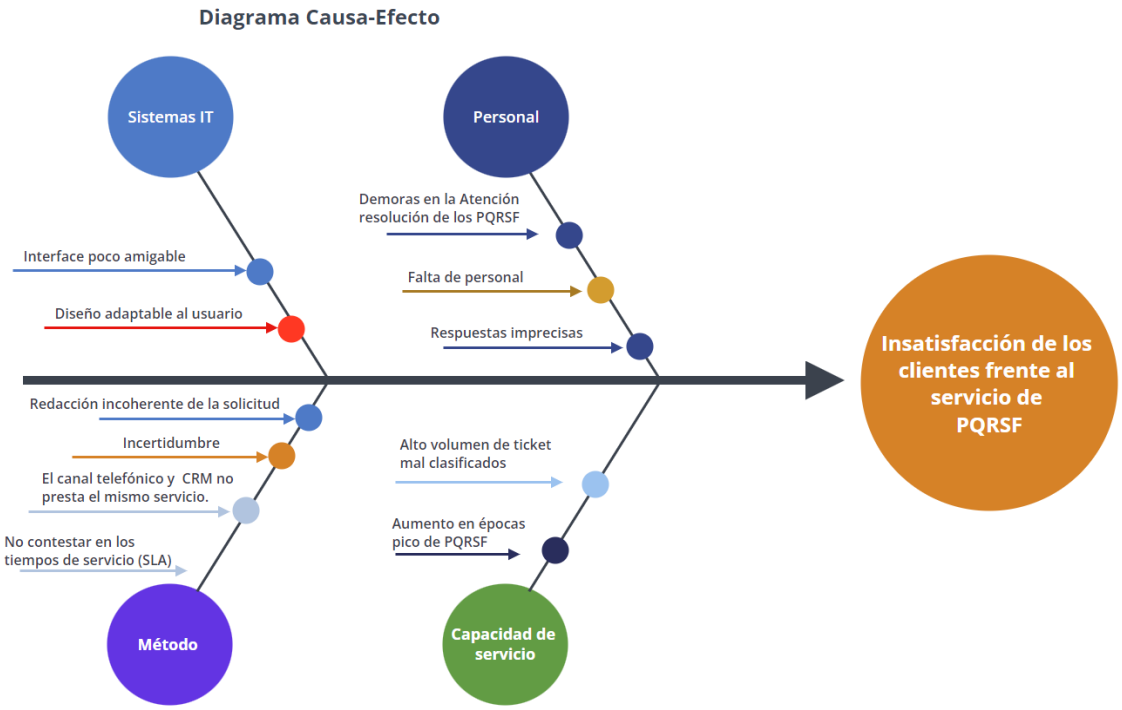
El procedimiento utilizado para el registro de las PQRSF, debe ser mejorado, ya que actualmente es un sistema manual en el que se presentan redacciones incoherentes, causandosan incertidumbre al usuario, debido a que se pueden dar situaciones en las que no queda bien clasificada la información, por lo tanto, es muy probable que no se cumpla con los tiempos de respuesta establecidos.

Además, el uso del sistema CRM Monitor sumado a una redacción incoherente de las PQRSF puede tener un grado de incertidumbre o impresiones que ocasionan que los tiempos de respuesta se incrementen y no se cumpla con la promesa de servicio de acuerdo con la solicitud, que pueden ser de 24 y 48 horas; por otro lado, la capacidad de servicio de los PQRSF en las temporadas altas puede generar clasificaciones erradas.

Por otro lado, la capacidad del servicio se ve afectada en épocas pico tales como los tiempos de matrículas, ingreso de nuevos estudiantes (con consultas sobre diferentes servicios de la universidad) que generan acumulacion de PQRSF y en muchos casos mal clasificados por el desconocimiento de usuarios nuevos en el funcionamiento de la institución, lo que impide que se identifiquen con claridad las áreas a las cuales se deben dirigir. Si bien, en la Universidad en el proceso de inducción se viene reforzando la información sobre los procesos institucionales para que los nuevos estudiantes conozcan las diferentes áreas y los servicios que prestan con el fin de reducir las solicitudes de PQRSF, la desatención obedece al desconocimiento de los usuarios para aclarar inquietudes lo que permitiría minimizar la insatisfacción de los clientes.

Este proyecto se desarrolla a partir de la siguiente pregunta orientadora: ¿Cómo implementar la clasificación de PQRSF de la Universidad Católica Luis Amigó aplicando *Machine Learning* (ML)?

Figura 2. Diagrama causa – efecto (espina de pescado)



Fuente: Elaboración propia

Nota. El gráfico representa la relación de diversos factores de las causas, sub-causas y un efecto o fenómeno.

4.JUSTIFICACIÓN

El trabajo se plantea bajo las directrices que emanan del sistema de calidad de la Universidad Católica Luis Amigo, busca generar diversos canales de comunicación intersubjetiva y no solo regular y controlar, sino también mejorar los canales de respuesta a los usuarios del servicio que sirvan de base y de interpretación del alto flujo valorativo de información que demanda la atención clara, ordenada y procedimental con la evidencia de la intervención.

Además, se pretende delimitar la composición de elementos específicos propios de la gestión empresarial, e inclinarse por un enfoque centrado en el usuario, a partir de establecimientos valorativos y propositivos, dicho de otra manera, lo que se pretende es determinar los grados de conformidad hacía respuestas que serán foco de una intervención, el cual favorece hallazgos importantes en el flujo comunicacional en una empresa, elementos que a su vez servirán de base para responder a los requerimientos de diversa índole en las instituciones.

Por otra parte, el mundo de hoy se patentiza interconectado por medio de un clic y tal asunto ha despertado el interés de propiciar bases informativas a gran escala que faciliten la reorganización de la información, así como la capacidad de conservar de manera deliberada los recursos que se establecen en los datos alfanuméricos, y en las fuentes Inter comunicativas que establecen el control y el manejo de información.

Una consecuencia del incremento de información en el sistema de PQRSF, es que en ocasiones se quedan solo en el papel o en su defecto en un buzón, puesto que tales cuestiones algunas veces no están clasificadas, ordenadas, categorizadas y sin un acompañamiento sistemático y ordenado; razón por la que se hace necesario crear espacios de interacciones mediáticas, que sirvan de base no solo para agrupar, ordenar, categorizar, socializar, responder y acompañar las PQRSF, sino que que además sean diligenciadas de manera óptima para así responder estratégicamente a las distintas demandas que surgen de las interacciones organizacionales con el usuario.

Lo anterior, sumado a la trazabilidad que se lleve de cada PQRSF en la institución, son instrumentos importantes que permiten mediante la categorización mejorar la toma de decisiones y determinar acciones que permitan contrarrestar el mejoramiento de los distintos procesos de atención, con el fin de resolver las inquietudes de la comunidad, pero

que, además, quede la evidencia de todo el proceso interno y los alcances que tuvo el requerimiento.

Por otra parte, se pretende que en virtud de la cantidad de PQRSF que se reciben en la institución, se puedan generar mecanismos bajo un modelo de minería de texto, que admite el manejo de información a gran escala, para atenderlas y darles solución de manera integral, basados en la Ley 1755 de 2015 que describe el objeto y modalidades del derecho de petición, garantizando las respuestas a las PQRSF cumpliendo con los términos de ley.

Por consiguiente, algunas técnicas de ML como la Minería de Texto ayudan a la automatización de las actividades y al entendimiento de las organizaciones, convirtiéndose en nueva serie de herramientas, enmarcadas bajo la Minería de Texto, permitiendo, no solamente, la investigación de la información, sino además, del planteamiento y hallazgo automático de hechos y conclusiones (patrones, normas, conjuntos, funcionalidades, modelos, secuencias, interacciones, correlaciones...) que tienen la posibilidad de desembocar en importantes decisiones y aumento en la satisfacción de los tiempos de respuesta.

Con la implementación que se hará al sistema en clasificación de PQRSF, se busca que la Universidad tome iniciativas en los procesos de atención y satisfacción del usuario, no solo desde la gestión de PQRSF, sino desde la implementación de soluciones tecnológicas que permitan la optimización en el funcionamiento de los servicios y mejorando los tiempos de respuesta, entre otros indicadores de manera que la eficacia y la disponibilidad mejoren.

Este proyecto permite aplicar, conceptualizar y profundizar en los conocimientos adquiridos en la especialización, aplicando la metodología Crisp-Dm y las técnicas de ML en el proceso de Servicio y Atención al usuario.

5.OBJETIVOS

5.1.Objetivo General

Implementar la clasificación de PQRSF de la Universidad Católica Luis Amigó aplicando Machine Learning (ML).

5.2.Objetivos Específicos

- Caracterizar el proceso de PQRSF de la Universidad Católica Luis Amigó
- Aplicar ML para generar modelos que permitan, a partir de su evaluación, seleccionar el modelo más eficiente para la clasificación de PQRSF.
- Probar el modelo seleccionado con un caso simulado de PQRSF aplicado en la Universidad Católica Luis Amigó.

6.MARCO METODOLÓGICO

Tabla 1.

Matriz metodológica

OBJETIVO	ACTIVIDADES	ENTREGABLE
Caracterizar el proceso de PQRSF de la Universidad Católica Luis Amigó FASE1-2-3	● Determinar el propósito del análisis y los indicadores que permitirán establecer el estado de rendimiento y funcionalidad del modelo a proponer ● Caracterizar el proceso relacionado con PQRSF. ● Acceder y caracterizar la data de las PQRSF. ● Preparar la data	● Documentar los propósitos de las PQRSF ● Documento de los procesos de PQRSF. ● Documento con la caracterización de la BD. ● Documento donde se evidencie el proceso de limpieza
Aplicar ML para generar modelos que permitan a partir de su evaluación seleccionar el modelo más eficiente para la clasificación de PQRSF. FASE 4	● Aplicar las técnicas de ML y para construir el modelo ● Evaluar los modelos generados para elegir el más eficiente. ● Seleccionar el modelo	● Documento con la descripción del resultado de la ejecución de cada una de las técnicas. ● Documento con el comparativo de los modelos aplicados y los criterios de selección. ● Documento en el que se describe el modelo elegido, con los resultados, justificación
Probar el modelo seleccionado con un caso simulado de PQRSF	● Generar un plan de prueba que permita evaluar los resultados	● Documento que explica la aplicación del plan

aplicado en la Universidad Católica Luis Amigó. FASE 5-6	● Planear la implementación	de prueba sobre el modelo seleccionado y los resultados. ● Informe con el resumen de los puntos más importantes y la experiencia adquirida

Nota: Matriz metodología Crisp- DM; elaboración propia, datos tomados de (IBM, 2021)

Los objetivos previamente descritos con sus actividades y entregables se desarrollarán a partir de la metodología Crisp-Dm (Cross-Industry Standard Process for Data Mining). Es un método validado, de los más utilizados; incluye un modelo y guía. Para la aplicación de esta metodología, se tienen en cuenta seis fases que dependen entre sí y en cada fase se encuentran los siguientes elementos:

La fase I tiene como propósito alinear los objetivos del proyecto con los del negocio, razón por la que se le ha dado el nombre de “compresión del negocio”. En la fase dos se recopilan los datos para comprenderlos y detectar posibles problemas de calidad; posteriormente, en la fase tres se realizará la preparación de los datos seleccionando los campos requeridos para hacer la limpieza de estos y posteriormente se realizará el modelado de los mismos.

Continuando en la cuarta fase, por medio del modelado se seleccionan las técnicas y se aplicarán los algoritmos para la construcción del modelo. Pasando a la fase cinco se evalúa el modelo para saber qué tan acertados son los resultados con respecto a los objetivos que se trazaron en etapas anteriores, para revisar si se deben de realizar ajustes en etapas anteriores o si debe ajustar el modelo y finalmente en la fase seis se realiza el despliegue e implementan los resultados obtenidos, facilitando la obtención de conocimiento a partir de los datos.

Para alcanzar el objetivo número uno, se aplicaron las fases uno, dos y tres de la metodología Crisp -DM con el propósito de caracterizar el proceso de PQRSF de la Universidad Católica Luis Amigó, lo primero que se hizo fue identificar y conocer todo el flujo del proceso actual, posteriormente se determino el propósito del análisis y los indicadores que permitieron establecer el estado de rendimiento y la funcionalidad del

modelo a proponer, para luego acceder a la data, dándole finalmente tratamiento a la misma.

Por otra parte, para dar respuesta al objetivo número dos, se obtendrá apoyo de la fase cuatro de Crisp-Dm, donde se aplicará ML para generar modelos que permitan, a partir de la evaluación, seleccionar el más eficiente para la clasificación de PQRSF.

Por último, y no menos importante, el logro del objetivo número tres estarán fundamentadas en las fases cinco y seis de Crisp-DM con el fin de probar el modelo seleccionado con un caso simulado de PQRSF aplicado en la Universidad Católica Luis Amigó, en el que a partir de un plan de prueba se evaluará los resultados obtenidos y se planteará la implementación.

7. MARCO REFERENCIAL

7.1.Marco Teórico

En la actualidad se percibe un gran interés en implementar sistemas de clasificación automática de ticket (solicitudes, PQRS) de atención a usuarios, debido al crecimiento de las empresas, a la conectividad y a la complejidad de brindar una atención oportuna a los diferentes requerimientos que tienen las personas.

Para ello se trabajará con Minería de Texto según Cristian Hugo MORALES Alarcón 1; Ciro Diego RADICELLI García 2; María Fernanda JARAMILLO Pinos 3; y Elba María BODERO Poveda4, definida como la extracción automática de información a partir del texto, la cual se encuentra previamente desconocida y que es potencialmente útil. Es así que a partir de la utilización de la técnica de minería de texto se busca dar respuesta a la clasificación de las PQRSF que le realizan a la Universidad.

Por tanto, al efectuar la revisión de literatura, se encontraron múltiples debates de cómo diseñar un sistema de clasificación de ticket, por lo que se direccionó a la búsqueda de un diseño de un sistema práctico con técnicas de representación del texto y algoritmos de aprendizaje supervisado, empleando un conjunto de datos etiquetados que permitieran obtener una asertividad en la clasificación, abordando varios problemas como su asignación, priorización, identificación de duplicados, tickets mal descritos y cambios al interior de la organización, como nuevos productos, servicios, reestructuraciones, que pueden dar pie a una nueva categoría de clasificación, que no estaban establecidas en el entrenamiento previo.

Autores como Revina, Buza y Meister (2020), muestran en un estudio como la representación lingüística que no solo demuestra ser altamente explicable, sino que también demuestra un aumento sustancial en la calidad de la predicción en comparación con TF-IDF. Asimismo, para la clasificación y etiquetado de sistema de tickets se utiliza la técnica de representación de texto de ticket TF-IDF (Term Frequency - Inverse Document Frequency (frecuencia de ocurrencia del término en la colección de documentos) y modelos kNN, Naive Bayes, (hubness-aware classifiers), árboles de decisión, regresión logística, máquinas de vectores de soporte, redes neuronales; además algunos modelos emplean el procesamiento del lenguaje natural o PNL, que es la conjunción de diferentes

métodos mediante los que se realiza un análisis sintáctico y semántico. El artículo muestra que existe una creciente complejidad con relación a la clasificación y las técnicas que se deben emplear en las empresas para no seguir realizando procesos manuales, sino que por el contrario buscan implementar soluciones tecnológicas como la clasificación de texto. (Revina et al., 2020, p. 1)

Por otra parte, Becerril & Villa, (2018) afirman que "Para poder brindar un servicio acorde a las necesidades y expectativas de los clientes, es importante la existencia de un sistema que permita controlar, administrar y asegurar la calidad". (p.2). para ello es importante realizar la reestructuración de un sistema de quejas y reclamos, que ayude a fortalecer el enfoque hacia el cliente, controlando, administrando y asegurando la calidad y al mismo tiempo logrando la motivación del personal al ver mejoras en habilidades en el trabajo, brindando un servicio acorde a las necesidades y expectativas de los clientes.

Además, Arvinder & Chopra (2016) afirman que, "La Minería de Textos se utiliza en todos los campos, ya sea para inteligencia comercial, análisis de redes sociales, análisis de sentimientos, análisis biomédico, análisis de procesos de software e incluso para análisis de seguridad" (p.1). Lo que quiere decir que la minería de datos es un campo de investigación bastante predeterminado, y que existen herramientas accesibles para aplicarla. Sin embargo, la mayor parte de estas tecnologías no soportan datos no estructurados.

Para Zicari et al., (2021) "La preparación de datos es un paso crucial para mejorar el rendimiento de la tarea de clasificación, el preprocesamiento incluye las operaciones de limpieza de los textos con errores (faltas de ortografía), así como la eliminación de espacios y palabras vacías" (p.2). Buscando de esta manera aumentar el porcentaje de clasificación, incrementando beneficios para las empresas y los usuarios, brindando respuestas más rápidas y precisas.

La información no estructurada de la revisión es importante, rápida y diversa, y supone un reto para la información empresarial y en particular, la decisión de comprensión de las características dinámicas de un marco para los usuarios empresariales que desean comprender el modelo del proceso.

La aplicación de Metodologías como: la minería de procesos aplicada para extraer el proceso existente y evaluar si un proceso comercial sigue las pautas de ITIL; la Minería de Datos de dominio enfocado para optimizar los servicios de TIS; las técnicas de Minería de

Procesos como desafío para analizar el abuso de los usuarios en espera y las técnicas estadísticas para analizar los datos de los tickets de soporte de TI para identificar anomalías; permite el diseño de modelos preventivos automatizados fundamentados en el aprendizaje automático, descubriendo procesos con el fin de recuperar información más útil para el proceso de gestión de incidentes.

Por otra parte, Zhong et al., (2012) afirma que “La minería de texto, es el descubrimiento de conocimientos interesantes en documentos de texto. Es un tema desafiante encontrar conocimiento preciso (o características) en documentos de texto para ayudar a los usuarios a encontrar lo que buscan” (p.1), ya que permite realizar clasificaciones para interpretar los datos y, aunque obtener su rendimiento es algo complejo, se recomienda hacerlo creando patrones en un tiempo razonable como lo son las bolsas de palabras.

Ademas, Tolciu et al (2021) menciona que, “el Procesamiento del Lenguaje Natural (PLN), ha demostrado ser un tema difícil, en el que los humanos siguen superando a las máquinas”. (p.1). Sin embargo, técnicas como las redes neuronales clasifican los documentos de texto dada la capacidad que tiene para trabajar datos en función del tiempo; actualmente, los sistemas de gestión de tickets son utilizados para mejoras de la organización. Los tickets son casos que contienen nombre y una solicitud por parte del cliente.

Según Zicari et al., (2021) “Una tendencia bastante reciente es el uso de Aprendizaje profundo de arquitecturas para la clasificación de tickets, que demostraron ser un medio poderoso y conveniente para inducir modelos precisos a partir de datos sin procesar de bajo nivel”. (p.2). El desarrollo de modelos de aprendizaje automático en las empresas ha ayudado a dar mejores clasificaciones de las PQRSF mediante la asignación automática y la clasificación de textos para la obtención de características y a su vez las categorías correspondientes.

Por otra parte, Yan et al., (2017) exponen que un algoritmo de clasificación clásico, KNN se usa ampliamente en la clasificación de texto. Pero cuando se enfrenta a datos de texto masivos en Internet, el algoritmo de clasificación de texto KNN en serie parece ser inadecuado” (p.3). Lo que significa que cuando se quiere obtener información útil de los datos se deben adoptar algoritmos eficientes; y en los últimos años, se ha tenido un gran avance en cuanto a los clasificadores de texto, los más comunes para cumplir dicho

propósito son: los vecinos K más cercanos (KNN), máquina de vectores de soporte (SMV), algoritmo de bayes, red neuronal, algoritmo de refuerzo y árbol de decisión.

En el caso del Big Data, con base en él, se aplican varias metodologías para el análisis de los datos como funciones matemáticas, análisis estadístico, funciones lógicas lo que permite manipular, organizar, limpiar los datos para clasificar y etiquetar de forma predictiva los tickets; para ello, se utilizan métodos supervisados para el análisis predictivo útiles para la asignación de una etiqueta al ticket; este tipo de algoritmos son los modelos de regresión lineal y logística, los árboles de decisión, las redes neuronales y K-NN (k - nearest neighbor) (Vallalta, 2022) y métodos no supervisados para realizar la Clusterización (clasificación de ticket) a un departamento o dependencia para su atención utilizando los métodos de agrupamiento jerárquico y k-means.

También, es importante tener aproximaciones sobre algunos conceptos clave de Big Data que según Hassani et al., (2020), "el Big Data está emergiendo como una poderosa herramienta para aprovechar el poder de los datos textuales no estructurados utilizándolos para extraer nuevos conocimientos e identificar patrones significativos y correlaciones ocultas en los datos". (p.1). Este se caracteriza por una automatización y gestión del conocimiento en todas las dimensiones que forman la sociedad actual. Con estas maneras se procura que ese conocimiento pueda ser potenciado, difundido e intercambiado; es decir, es una acción que debe conducir a un bienestar económico y social.

De otra parte, en el escrito Técnicas de aprendizaje de máquina utilizadas para la minería de texto Berry y Kogan (2010), afirman que "los mayores temas estudiados en la minería de texto son la extracción de palabras clave, clasificación, agrupamiento, extracción de nombres y entidades, detección de anomalías y tendencias y flujos de texto. Cada uno de esos temas forma parte de una subárea de la minería de texto". (p.13). Que son algoritmos que permiten identificar patrones en grandes cantidades de datos, creando sistemas automatizados capaces de realizar predicciones.

Por otra parte, el lenguaje que se va a implementar es ***Python***, ya que es un lenguaje de programación creado en 1989 empleado en la "administración de sistemas, ciencia de datos, computación científica (donde domina con diferencia), inteligencia artificial, internet de las cosas" (García, 2017, p. 151).

También se encuentra la **minería de texto**, cuya tecnología se remonta a la década de los noventa, esta sería una aplicación de la lingüística computacional y del

procesamiento de textos que pretende facilitar la identificación y extracción de nuevo conocimiento a partir de colecciones de documentos o corpus textuales (Brun y Senso, 2004, p. 11).

De acuerdo con lo anteriormente expuesto, este proyecto se orienta principalmente a la aplicación de *Machine Learning en Minería de Texto* que es un subcampo de la inteligencia artificial que utiliza códigos suaves en vez del enfoque convencional del código fijo; se refiere al mecanismo en el que, a pesar de la ausencia de instrucciones explícitamente programadas, la máquina puede seguir aprendiendo desde la experiencia"'; que permita mejorar el sistema de PQRSF de la Universidad Católica Luis Amigó.

7.2.Marco Conceptual

A continuación, se relacionan algunos de los conceptos que permiten una mejor comprensión del desarrollo de los objetivos del trabajo de grado:

Como punto de partida está **CRISP-DM** (Cross-Industry Standard Process for Data Mining) que es un método de Minería de Datos que propone la aplicación de un ciclo de vida con un procedimiento gradual a un proyecto de exploración de datos siendo la más utilizada en la actualidad. Esta metodología optimiza el proceso de planeación de cada uno de los objetivos planteados anteriormente; ya que está diseñada para ejecutarse por fases, donde cada una permite identificar las actividades que se realizarán y cómo serán desarrolladas para dar cumplimento a los objetivos.

Big Data. Vargas y Peñaloza (2019) expresan que es una forma evolucionada y digital en la que se procesan y gestionan grandes cantidades datos que se convierten en conocimiento; esta caracterizados por una "automatización y gestión del conocimiento en todas las dimensiones que forman la sociedad actual"; su propósito radica en que ese conocimiento pueda ser "potenciado, difundido e intercambiado". (pp. 13-14); es decir, es una acción que debe conducir a un bienestar económico y social. Su aporte es que manejamos grandes cantidades de registros de las PQRSF por parte de los usuarios para aplicarles clasificación, estos datos son obtenidos por una base de datos.

Minería de texto, cuya tecnología se remonta a la década de los noventa. Esta sería una aplicación de la lingüística computacional y del procesamiento de textos que pretende

facilitar la identificación y extracción de nuevo conocimiento a partir de colecciones de documentos o corpus textuales (Brun y Senso, 2004, p. 11). Su aporte es que dado lo estudiado a lo largo de la especialización, la minería es ideal para la realización de clasificación o predicciones.

Python, es un lenguaje de programación que su creación se remonta a 1989, y el cual es empleado en la "administración de sistemas, ciencia de datos, computación científica (donde domina con diferencia), inteligencia artificial, internet de las cosas" (García, 2017, p. 151). Este lenguaje es en el que se soporta el código para la clasificación de las PQRSF.

A continuación, se describen las librerías que fueron aplicadas en este proyecto:

- **Numpy:** Es una librería de Python que realiza cálculos numéricos. Adicional permite trabajar con matrices y vectores realizando cálculos de una manera muy eficiente (Numpy.org, s.f).
- **Pandas:** Esta librería se usa para tareas de ciencia de datos, análisis y aprendizaje automático. Es capaz de leer y escribir archivos con extensión CSV, SQL, HDF5, entre otros. Cargar, seleccionar, leer y filtrar de manera sencilla dataset y/o tablas y realizar tareas como unir, limpieza, relleno, normalización, visualización, análisis estadístico, inspección, cargando y guardando de datos. (pandas.pydata.org,s.f).
- **Matplotlib:**Librería de python especializada en la creación de gráficos bidimensionales, permite crear y personalizar los tipos de gráficos más comunes, entre ellos: Diagramas de barras e Histograma (Matplotlib.org, s.f).
- **Seaborn:** librería de python especializada para dibujar gráficos estadísticos atractivos e informativos (seaborn.pydata.org, s.f.).
- **string:** las cadenas en ython o string son un tipo inmutable que permite almacenar secuencias de caracteres. Para crear una, es necesario incluir el texto entre comillas doble " ``. (El libro de Python, s.f.).

Por otra parte, están las **PQRSF**, sigla que significa peticiones, quejas, reclamos y sugerencias. La petición es aquella en la que la persona solicita de forma respetable una información o consulta para conseguir una solución; las quejas hacen alusión a un descontento por las acciones de alguien más, estas deben ser resueltas en un máximo de 15 días; el reclamo consiste en el descontento por la presentación de un servicio, al igual que las quejas estas deben ser resueltas en un máximo de 15 días; finalmente están las

sugerencias, definidas como aquellas recomendaciones por parte del usuario con el fin de mejorar los servicios que se prestan, estos tienen un plazo de 10 días (Asamblea de Antioquia, s.f.). Es a través de las peticiones, quejas, reclamos, sugerencias, solicitudes y felicitaciones, que los usuarios expresan sus necesidades a la universidad y con las cuales se trabajarán las técnicas de ML para la clasificación.

Mientras que **data mining** es el encargado de hallar patrones en grandes volúmenes de datos, donde trata en mayor parte por medio de tecnologías como las redes neuronales y la lógica difusa (fuzzy logic) (Weber, 2000). Con este proceso se pretende descubrir patrones existentes en las PQRSF, apoyados en métodos de aprendizaje de clasificación.

En cuanto al **Aprendizaje automático/Machine Learning** es un subcampo de la inteligencia artificial que utiliza códigos suaves en vez del enfoque convencional del código fijo; se refiere al mecanismo en el que, a pesar de la ausencia de instrucciones explícitamente programadas, la máquina puede seguir aprendiendo desde la experiencia (Subasi, 2020). Con este aprendizaje se implementará el desarrollo de una técnica que permita que el modelo aprenda por medio del entrenamiento de los datos.

Además, el **Árbol de decisión** es una técnica enmarcada dentro del desarrollo de procedimientos y sistemas de argumento usados en indagaciones de IA (inteligencia artificial) y programación de aplicaciones, por su composición son sencillos de entender y examinar; su implementación diaria se puede ofrecer en diagnósticos doctores, predicciones meteorológicas, controles de calidad, y otros inconvenientes que necesiten de estudio de datos y toma de elecciones (Calancha,2011). Para este proyecto se utilizará esta técnica como apoyo para determinar qué tan idónea es para la clasificación de la PQRSF.

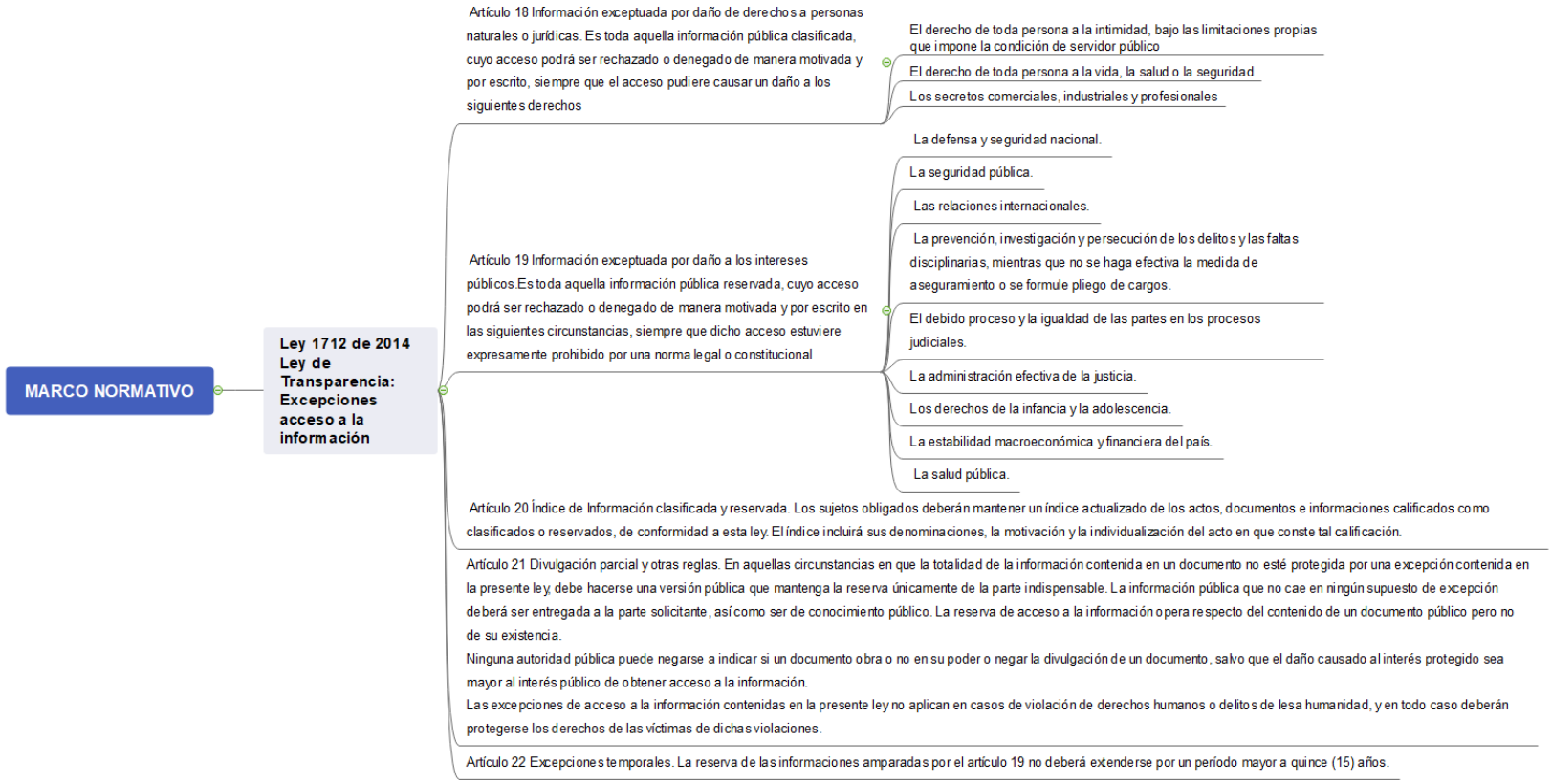
A la vez, se empleó la **regresión logística que es una:** "herramienta de modelización muy potente, es una generalización de la regresión lineal y se utiliza para evaluar la probabilidad" (Amine et al., 2021). Para este proyecto se buscó el apoyo en esta técnica para determinar qué tan idónea es para la clasificación de la PQRSF.

Finalmente, **Máquinas de sector vectorial:** las máquinas de sector vectorial (SVM) son un clasificador supervisado adecuado para problemas de clasificación de texto, dado que puede manejar grandes características (Paramesh & Shreedhara, 2019. p.338). Esta es una de las técnicas que se implementará para ser evaluada y determinar qué tan idónea es para la clasificación de la PQRSF.

7.3.Marco Normativo

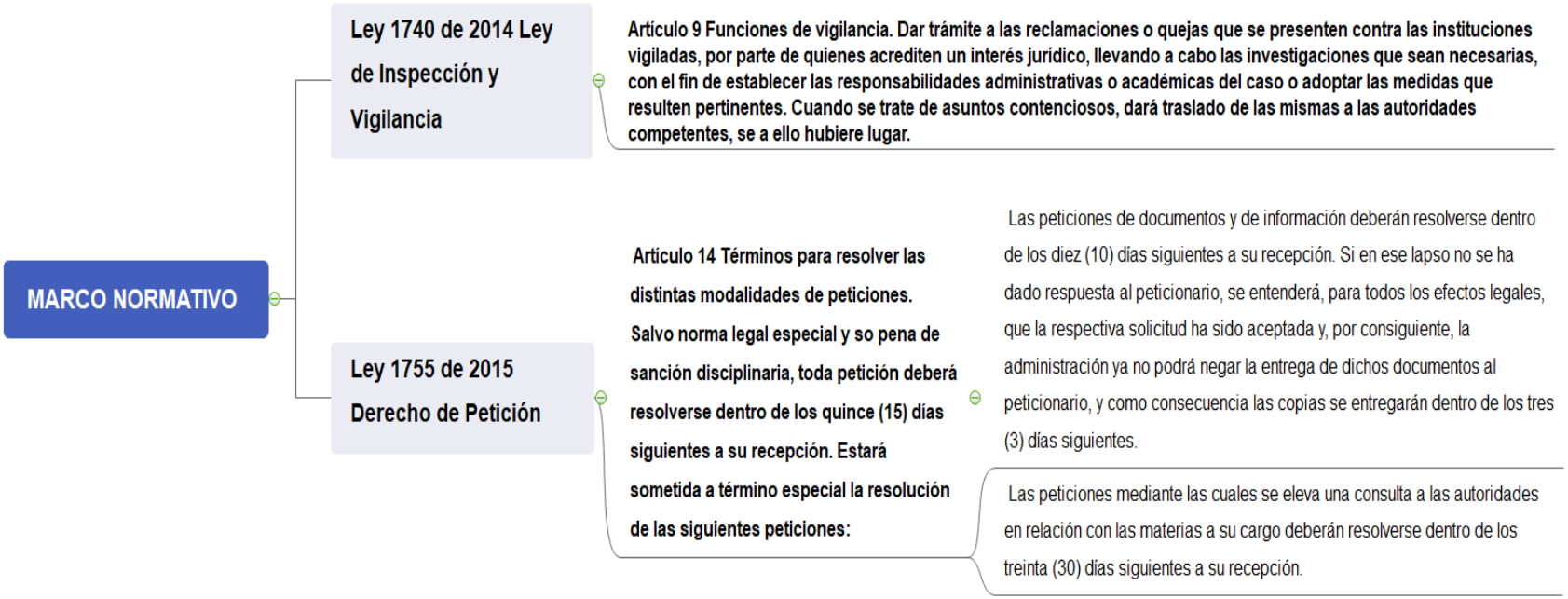
La Universidad Católica Luis Amigó en cumplimiento de lo establecido en el artículo Art 23 de la Constitución Política de Colombia, que dice: “Toda persona tiene derecho a presentar peticiones respetuosas a las autoridades por motivos de interés general o particular y a obtener pronta resolución”. Y las demás normas concordantes, entre ellas la Ley 1712 de 2014 Ley de Transparencia título III Excepciones acceso a la información en los artículos del 18 al 22. Ley 1740 de 2014 Ley de Inspección y Vigilancia numeral 4 del artículo 9. Ley 1755 de 2015 por la cual se regula el derecho de petición a través de la Secretaría General y los procedimientos internos de atención al ciudadano; en cuanto a la recepción y trámite de las peticiones, quejas, reclamos, solicitudes y felicitaciones, ha dispuesto mediante Resolución rectoral No 50 de 2018. A continuación, se describe el marco normativo relacionado:

Figura 3. Ley 1712 de 2014



Nota: la figura muestra los artículos de la Ley 1712 de 2014, que están relacionados con la transparencia en el manejo de la información, los datos fueron tomados de (Ley 1712, 2014)

Figura 4. Ley 1740 de 2014 y Ley 1755 de 2015



Nota: la figura muestra los artículos de la Ley 1740 de 2014 y 1755 de 2015, con algunos factores que se deben tener en cuenta en las reclamaciones y sol ución de peticiones, los datos fueron tomados de (Ley 1712, 2014) y (Ley 1755 de 2015).

8.DESARROLLO DEL PROYECTO

8.1 Caracterización del proceso de PQRS de la Universidad Católica Luis Amigo (fase 1,2,3 del Crisp – Dm)

En este apartado se describen cada una de las actividades realizadas para lograr la caracterización del proceso de PQRS de la Universidad Católica Luis Amigo: desarrollo del primer objetivo.

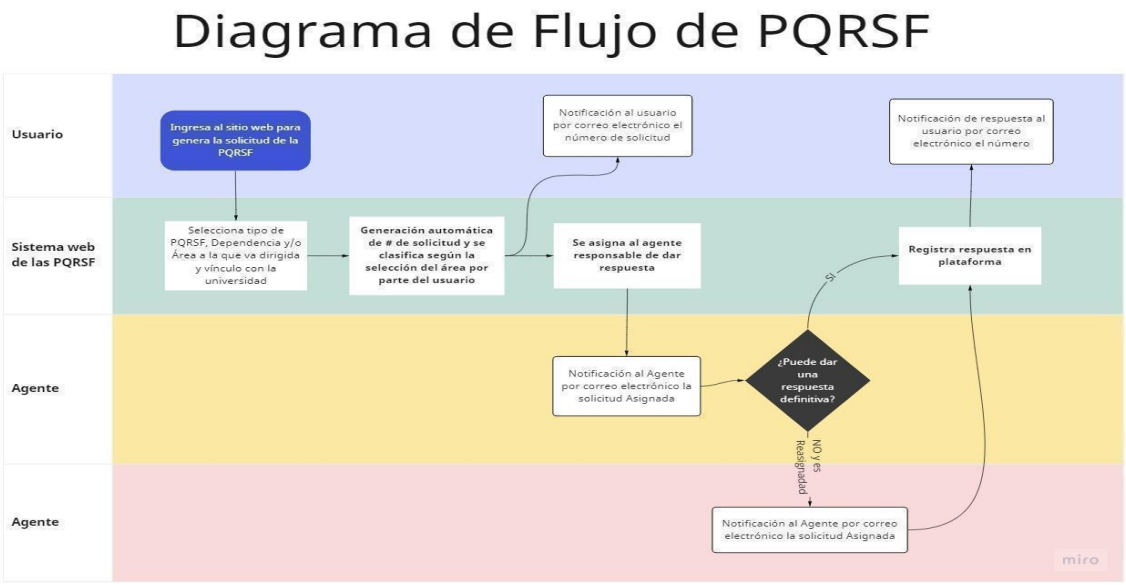
8.1.1 *Determinar el propósito del análisis y los indicadores que permitirán establecer el estado de rendimiento y funcionalidad del modelo a proponer (FASE 1).*

El propósito del presente trabajo está motivado en la necesidad de mejorar el sistema de peticiones, quejas, reclamos, solicitudes y felicitaciones (PQRSF). A partir de la aplicación de una técnica de ML que permite realizar la clasificación de estas, con el fin de mejorar los tiempos de respuesta en relación con la promesa de servicio y a su vez la satisfacción del usuario. Estos propósitos fueron:

- (Porcentaje (nivel de clasificación de los PQRSF)) Reducir el porcentaje de clasificación de las PQRSF que fueron etiquetadas erróneamente en un 15%, para el semestre 1-2022 se reclasificaron 1.302 solicitudes de 4.214 equivalentes al 31%.
- Mejorar el nivel de satisfacción del cliente, reduciendo en un 10% del promedio los tiempos de respuesta en los PQRSF mal clasificados.

8.1.2 Caracterizar el proceso relacionado con PQRSF (FASE 1).

Figura 5. Diagrama de flujo con el proceso que se realiza en la Universidad Católica Luis Amigó a las PQRSF Fuente (Elaboración propia)



En la figura 5 se muestra el proceso actual del funcionamiento de una PQRSF en la Universidad Católica Luis Amigó, permitiendo tener claridad sobre el proceso que se va a intervenir. Iniciando con el ingreso a sitio web para registrar la solicitud, seguidamente se puede seleccionar el tipo de PQRSF, dependencia o área a la que va dirigida, al igual que el vínculo que se tiene con la Universidad, el siguiente paso es la generación automática de la asignación de la solicitud, que es clasificado una vez se selecciona el área, posteriormente es notificado mediante correo electrónico el número de la solicitud y en el proceso interno se asigna el agente encargado de dar la respectiva respuesta, donde se pueden dar por terminada la respuesta o se puede generar una reasignación según sea el caso.

8.1.3 Acceder y caracterizar la data de las PQRSF (FASE 2)

8.1.3.1. Recolección de los datos

La recolección de datos se realizó mediante la búsqueda de la fuente de datos, en este caso el sistema de atención a usuarios que soporta las PQRSF, se hizo la gestión con el encargado de los datos y los permisos necesarios y se obtuvo los datos en formato .xls el cual nos permitió visualizarlos y manipularlos para su preparación.

Una vez se accedió a la base de datos y se conocieron los atributos, el siguiente paso consistió en el análisis y estructura de los mismos para determinar qué atributo ayudaba a realizar la clasificación de las PQRSF.

Tabla 2.

Descripción de los campos de la base de datos

Variable	Tipo	Descripción
ticket_id	Numérico	Código del PQRSF registrado para hacer los seguimientos.
topic_id	Numérico	Código del área a la que fue asignada la PQRSF
Topic	Texto	Área a la que se le asigna la PQRSF
Subject	Texto	Asunto de PQRSF
Body	Texto	Cuerpo del mensaje
Created	Fecha	Fecha de creación del mensaje
Closed	fecha	Fecha de cierre del mensaje
thread_type	Texto	Tratamiento que se le dio a la solicitud (M, N, R), mensaje (M), Transferido(N), Respondido(R)
dept_id	Numérico	Código del departamento o unidad
dept_name	Texto	Nombre del departamento o unidad al que se remite la solicitud
group_name	Texto	nombre del grupo de trabajo
Source	Texto	canal del PQRSF
State	Texto	creado, transferido, cerrado
Staff	Texto	quien lo atendió
Priority	Texto	prioridad del PQRSF (bajo, normal, alto, emergencia)
Poster	Texto	Usuario de la PQRSF

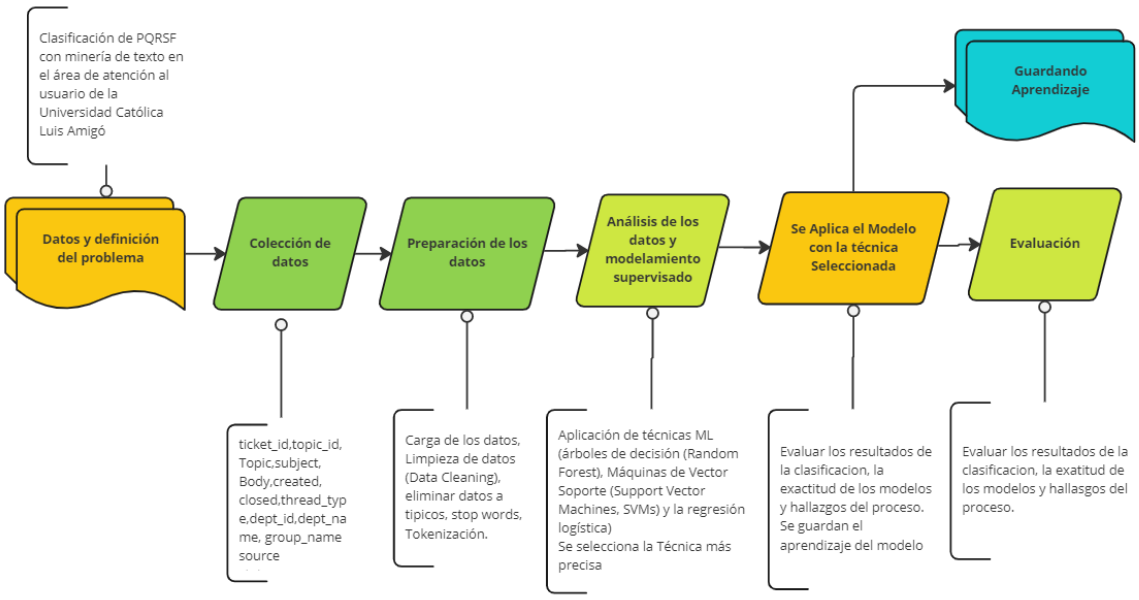
Además, se realizó un análisis acerca de los atributos que aportan valor para cumplir con los indicadores descritos anteriormente, que son relacionados en la tabla 3:

Tabla 3.

Campos de respuesta

Indicador	Cantidad de Atributos	Atributos
(Porcentaje (nivel de clasificación de los PQRSF)) Reducir el porcentaje de clasificación de las PQRSF que fueron etiquetadas erróneamente en un 15%, para el semestre 1-2022 se reclasificaron 1.302 solicitudes de 4.214 equivalentes al 31%.	9	topic, subject, body, created, closed, thread_type, dept_name, topic_id, dept_id
Mejorar el nivel de satisfacción del cliente reduciendo en un 10% del promedio los tiempos de respuesta.	8	topic, body, priority, state, created, closed, topic_id, dept_id

Figura 6. Diagrama de Arquitectura



Fuente: Elaboración propia

En el diagrama anterior se puede evidenciar la arquitectura de cómo se procesan los datos. Donde se inicia con la clasificación de las PQRSF, la aplicación de la minería de texto para construir con colección de los mismos, luego se pasa a la preparación, que permite realizar una limpieza, eliminando las palabras vacías (Stopwords), información

atípica y se aplicar la seguridad de datos a través de la tokenzación. Posteriormente, se aplican varias técnicas de ML como árboles de decisiones, Máquinas de soporte vectorial y regresión logística para así escoger la más eficiente. Finalmente, se procede a la evaluación de los resultados de la mejor técnica y se guardan el aprendizaje del modelo.

8.1.4 Preparar la data (FASE 3)

- **Preparación de los datos**

En este apartado, se identifican las variables, utilizar y analizar los datos en busca de información que no aporta valor, como redundancia de los datos duplicados, registros nulos, datos erróneos que no aportan información relevante para el desarrollo del trabajo, como son: nombre, correos electrónicos, números de celular, caracteres especiales, fechas, entre otros.

- **Seleccionar los datos**

En esta fase se seleccionó el atributo que aportará a la variable objetivo que es el campo **topic** que contiene la clasificación de las PQRSF, para posteriormente se realizó una limpieza de los datos que permitiera realizar la aplicación de las técnicas de minería de texto orientado a la clasificación de PQRSF; en este proceso de depuración o limpieza de datos se verificó la calidad de los mismos y se establecieron los campos requeridos para el desarrollo de la siguiente fase.

Estos fueron los atributos seleccionados, ya que son los más importantes con relación al objetivo definido en la fase 1.

- **Topic:** área a la que se asigna el PQRSF
- **Subject:** asunto de la PQRS
- **body:** cuerpo del mensaje
- **created:** fecha de creación del mensaje
- **closed:** fecha de cierre del mensaje
- **thread_type:** tratamiento que se le dio a la solicitud (M, N, R) Mensaje (M), Transferido(N) Respondido(R)

- **dept_name:** nombre del Departamento o Unidad al que se remite la solicitud
- **dept_id:** código del Departamento o Unidad
- **topic_id:** código del área a que fue asignado la PQRSF
- **ticket_id:** código del PQRSF registrado para hacer los seguimientos
- **State:** creado, transferido, cerrado.
- **priority:** prioridad del PQRSF (bajo, normal, alto, emergencia)

Tabla 4.

Campos de respuesta clasificados

VARIABLE	TIPO	NATURALEZA	ESCALA	OBSERVACIÓN
Topic	Categórico	Cualitativo	Nominal	Politómicas
Subject	Categórico	Cualitativo	Nominal	Politómicas
Body	Categórico	Cualitativo	Nominal	Politómicas
Created	Numérica	Cuantitativa	Intervalo	Discretas
Closed	Numérica	Cuantitativa	Intervalo	Discretas
thread_type	Categórico	Cualitativa	Nominal	Politómicas
dept_name	Categórico	Cualitativa	Nominal	Politómicas
dept_id	Numérico	Cuantitativo	De razón	Discreta
topic_id	Numérico	Cuantitativo	De razón	Discreta
ticket_id	Numérico	Cuantitativo	De razón	Discreta
State	Categórico	Cualitativo	Nominal	Politómicas
Priority	Categórico	Cualitativo	Nominal	Politómicas

- **Limpieza de datos (Data Cleaning)**

En este punto, se realizó la conversión de texto a minúsculas, se eliminaron las URL, los signos de puntuación o caracteres especiales (//*-°>< {^`), eliminación de los números y la eliminación de espacios en blanco

- **Tokenización**

En esta etapa se dividieron las cadenas de texto en partes más pequeñas o tokens para este caso de un párrafo se pasó a tener las palabras separadas para facilitar su tratamiento con las técnicas.

Figura 7. Tokenización

	body	texto_tokenizado
0	Buenas noches...Solicito recuperar mi contrase...	[buenas, noches, solicito, recuperar, contrase...
1	Cordial saludo. Desde el pasado lunes 13 de j...	[cordial, saludo, desde, pasado, lunes, julio,...
2	Al momento de ingresar al sistema académico, s...	[momento, ingresar, sistema, académico, solici...
3	BUENOS DÍASAl momento de llevar los papeles no...	[buenos, díasal, momento, llevar, los, papeles...
4	(Código + Nombre)Código : V_11_PROY_DOCENTE...	[código, nombre, código, proy, docentescontrat...

Nota: En esta imagen se puede observar que a la columna body se le aplicó una tokenización, que consiste en separar las palabras por seguridad de datos para obtener un mejor procesamiento cuando se aplique la técnica.

- **Stop words**

En este paso se hizo un filtro para quitar las palabras que no aportan información relevante para la clasificación, como artículos, preposiciones, pronombres.

Figura 8. Stop words

	texto_tokenizado	texto_stopwords
0	[buenas, noches, solicito, recuperar, contrase...	[recuperar, contraseña, sistema, académico, pu...
1	[cordial, saludo, desde, pasado, lunes, julio,...	[pasado, lunes, julio, fué, asignado, cubículo...
2	[momento, ingresar, sistema, académico, solici...	[momento, ingresar, sistema, académico, record...
3	[buenos, díasal, momento, llevar, los, papeles...	[díasal, momento, llevar, papeles, informaron,...
4	[código, nombre, código, proy, docentescontrat...	[código, nombre, código, docentescontratonombr...
5	[buenos, dias, necesito, ayuda, para, ingresar...	[necesito, ayuda, ingresar, sistema, académico...
6	[cordial, saludo, anteriormente, había, report...	[anteriormente, reportado, dificultad, correo,...
7	[cordial, saludo, muy, amablemente, solicito, ...	[amablemente, arreglo, chapa, puerta, economat...
8	[hola, estimados, amigos, servicios, generales...	[estimados, amigos, servicios, generales, inme...
9	[cordial, saludo, mes, pasado, habian, hecho, ...	[mes, pasado, habian, hecho, mantenimiento, si...

Nota: En esta imagen, se puede visualizar que a la columna de texto tokenizado se le aplicó Stopwords, que no es más que la limpieza de texto eliminando las palabras comunes, palabras vacías sin significado por sí sola, preposiciones, pronombres entre otros que no aportan valor al implementar la técnica de clasificación.

8.2 Aplicación dr ML para generar modelos que permitan a partir de su evaluación seleccionar el modelo más eficiente para la clasificación de PQRSF. (Fase 4 Crisp – Dm)

En esta etapa y para dar respuesta al objetivo número dos se aplicaron las técnicas de árboles de decisión (Random Forest), Máquinas de Vector Soporte (Support Vector Machines, SVMs) y la regresión logística, por medio de la herramienta Google Collaboratory también llamado Colab que permite escribir y ejecutar código desde una interfaz web para realizar la implementación con el lenguaje de programación Python (versión 3) donde se utilizaron las siguientes librerías: numpy, pandas, string joblib, matplotlib, seaborn, sklearn, nltk. Para el modelamiento se usaron técnicas supervisadas de clasificación,con el fin de elegir el modelo de MML más eficiente para la clasificación por medio de la minería de texto.

Los principales objetivos de la implementación fueron:

- Identificar las palabras empleadas por cada área de atención de PQRSF.

- Crear un modelo de machine learning capaz de clasificar los PQRSF por área de atención que envían los usuarios con base en las palabras empleadas.
- Para el desarrollo de estos objetivos, se realizó la importación de las librerías en Python de acuerdo a las funciones que van a desempeñar en el código:
    - Tratamiento de los datos: NumPy, Pandas, String, Joblib.
    - Graficar los resultados: Matplotlib y Seaborn.
    - Pre-procesado y modelado: sklearn y nltk

Posteriormente se cargó el DataSet desde Google Drive, ya que este facilita tenerlo siempre a la mano:

Figura 9. Importación de archivo

Nota: Se agrega el parámetro de separación en el siguiente método para indicar el carácter por el cual se va a realizar la carga de Set de datos sin la necesidad de realizar mayores cambios en el archivo original, ya que la idea es que dicha limpieza se realice directamente en Python.

Con el comando *dataset=pd.read_csv('/ruta del drive')* se carga el dataset, y con el siguiente comando *dataset. head (10)* se visualizan los datos del dataset cargado.

Figura 10.Carga Dataset

	ticket_id	topic	subject	body	created	closed	thread_type
0	5	Acceso a Sistema Académico	Contraseña	Buenas noches...Solicito recuperar mi contrase...	2015-07-15 21:26:47	2015-07-21 09:01:13	M
1	6	Requerimientos Infraestructura	Dificultades en la conexión a internet	Cordial saludo. Desde el pasado lunes 13 de j...	2015-07-16 08:20:56	2015-07-21 07:58:28	M
2	7	Acceso a Sistema Académico	No puedo ingresar al sistema académico.	Al momento de ingresar al sistema académico, s...	2015-07-16 10:10:33	2015-07-21 08:58:18	M
3	8	Acceso a Sistema Académico	NO PUEDO IMPRIMIR LIQUIDACIÓN MATRICULA	BUENOS DÍASAl momento de llevar los papeles no...	2015-07-16 10:25:37	2015-07-21 09:03:00	M
4	9	Acceso a Sistema Académico	informe docentes activos	(Código + Nombre)Código : V_11_PROY_DOCENTE...	2015-07-16 10:34:41	2015-07-21 07:56:10	M
5	11	Acceso a Sistema Académico	Ayuda acceso al Sistema Academico	Buenos Dias, necesito ayuda para ingresar al s...	2015-07-16 11:46:50	2015-07-21 09:00:51	M
6	12	Acceso al correo @amigo.edu.co	Solicitud para habilitar el correo institucional	Cordial saludo,Anteriormente había reportado m...	2015-07-16 12:07:33	2015-07-16 15:16:16	M
7	13	Requerimientos Infraestructura	Arrego de puerta y lamparas	Cordial saludo, muy amablemente solicito el ar...	2015-07-16 13:17:42	2019-10-03 17:21:49	M
8	15	Requerimientos Infraestructura	Arreglo del aparato telefónico	Hola estimados amigos de servicios generales, ...	2015-07-16 14:12:14	2019-10-03 17:20:58	M

Nota: Se cargó dataset con el comando dataset. head (10), posterior a ellos podemos visualizar que los datos cargaron correctamente

Figura 11.Descripción de variables

```
[ ]  dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 38452 entries, 0 to 38451
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   ticket_id    38452 non-null  int64 
 1   topic        38452 non-null  object
 2   subject      38452 non-null  object
 3   body         38452 non-null  object
 4   created      38452 non-null  object
 5   closed       38316 non-null  object
 6   thread_type  38452 non-null  object
dtypes: int64(1), object(6)
memory usage: 2.1+ MB
```

Nota: Descripción de las variables, el comando dataset.info () nos describe las columnas, sus etiquetas, el tipo de datos y el uso de la memoria

Figura 12. Descripción estadística

```
dataset.describe()
```

	ticket_id
count	38452.000000
mean	20995.163347
std	11235.917938
min	5.000000
25%	11839.750000
50%	21336.500000
75%	30683.250000
max	39929.000000

Nota: Descripción Estadística, resumen del conjunto de datos, tendencia central, dispersión

Primero se realiza un análisis exploratorio con el fin de entender cómo están los registros del dataset de las PQRSF y se revisa la distribución en el tiempo.

8.2.1 Aplicación de las técnicas de ML y construcción del modelo

En esta etapa, se aplicaron las técnicas de ML seleccionadas para construir el modelo de clasificación de las PQRSF, que fueron los árboles de decisión (Random Forest), Máquinas de Vector Soporte (Support Vector Machines, SVMs) y la regresión logística para posteriormente evaluar los resultados de los modelos generados, permitiéndonos seleccionar la más eficiente.

Los datos anteriormente mencionados se dividieron de la siguiente manera: 80 % entrenamiento y 20% prueba. Se realizó la transformación a vectores del conjunto de datos de la variable “texto_en_str” y en este caso se usó el método TFIDF (frecuencia de término – frecuencia inversa de documento) que permitió obtener un índice de relevancia de cada palabra en el documento o corpus. lo que facilita al algoritmo realizar la predicción con mayor efectividad.

Posteriormente, se construyó el sistema de métricas para la evolución del modelo donde se implementaron los siguientes parámetros Acurracy (Exactitud), Precision (precisión), Recall (sensibilidad), F1score. la comparación del rendimiento combinado de la precisión y la sensibilidad, como se muestra a continuación:

Figura 13.Métricas de Máquinas de Soporte Vertical (SVM)

Informe de clasificación de LinearSVC (1-gram):

	precision	recall	f1-score	support
Acceso a Sistema Académico	0.67	0.55	0.61	648
Requerimientos Infraestructura	0.92	0.97	0.94	4439
Acceso al correo @amigo.edu.co	0.86	0.82	0.84	693
Listados - Información SUI	0.73	0.76	0.75	765
Planta Física (Serv. Generales)	1.00	0.43	0.60	7
Centro Regional Apartadó	0.60	0.17	0.26	36
Educación Virtual	0.76	0.76	0.76	668
Acceso a intranet (REDentor)	0.93	0.92	0.93	264
Aulas Virtuales - Cursos de TIC	0.33	0.27	0.30	49
Cursos Administración Distancia	0.00	0.00	0.00	9
Aulas Virtuales - Cursos de AFI	0.62	0.42	0.50	99
Mantenimiento y Vigilancia	0.50	0.21	0.30	108
Sistema de Control de Acceso	0.00	0.00	0.00	5
accuracy			0.85	7790
macro avg	0.61	0.48	0.52	7790
weighted avg	0.84	0.85	0.85	7790

Matriz de confusión para LinearSVC:

Nota: En esta imagen se muestran las etiquetas de clasificación del sistema de PQRSfy Métricas con las que se evalúa la técnica de Máquinas de soporte vectorial (SVM)

La Matriz de confusión, permitió la evaluación de la precisión de la clasificación y una visualización matricial de los resultados de las predicciones, el desempeño del modelo de clasificación sobre el dataset etiquetado. Por otra parte, la diagonal representa los puntos donde los datos etiquetados coinciden con los datos etiquetados a través de la predicción del modelo y mientras más altos sean estos valores, más predicciones correctas tiene el modelo.

Figura 14. Matriz de Confusión

Nota: Matriz de confusión para idénticas cual o cuales son las variables que tiene mayor peso

A continuación, se describe el resultado de la aplicación de cada una de las técnicas y se analizan los resultados para definir de acuerdo a la evaluación cuál es el modelo que más eficiente para la clasificación de PQRSF.

- **Métricas macro promediadas con dos conjuntos de datos:**

$$\text{Macro - Precisión} = \frac{\text{Precisión1} + \text{Precisión2}}{2}$$

$$\text{Macro - Recall (sensibilidad)} = \frac{\text{Recall1} + \text{Recall2}}{2}$$

$$\text{Macro - F -Score} = 2.\frac{\text{Macro - Precisión. Macro - Recall}}{\text{Macro - Precisión} + \text{Macro - Recall}}$$

- **Técnica de regresión logística:** esta técnica mostró los siguientes resultados:
    - **Accuracy:** esta métrica mostró que la exactitud de la clasificación de los casos acertados fue del 85.33%.
    - **Macro Precisión:** esta métrica arrojó un porcentaje de 53.77% en la calidad de las clasificaciones correctas.
    - **Macro Recall:** esta métrica arrojó un 40.7% en relación a la cantidad de clasificaciones que el modelo fue capaz de identificar.
    - **Macro F1score**: La comparación del rendimiento combinado de la precisión y la sensibilidad es del 0.4311.
- **Técnica de árboles de decisión:** esta técnica devolvió los siguientes resultados: **Accuracy:** esta métrica mostró que la exactitud de la clasificación de los casos acertados fue del 57.29%.
    - **Macro Precisión:** esta métrica nos arrojó un porcentaje de 13.36% en la calidad de las clasificaciones correctas.
    - **Macro Recall:** esta métrica nos arrojó un 7.93% en relación a la cantidad de clasificaciones que el modelo fue capaz de identificar.
    - **Macro F1score**: La comparación del rendimiento combinado de la precisión y la sensibilidad es del 0.06.
- **Técnica de Máquinas de Soporte vectorial:** esta técnica ha devuelto los siguientes resultados:
    - **Accuracy:** esta métrica mostró que la exactitud de la clasificación de los casos acertados fue del 85%.
    - **Macro Precisión:** esta métrica arrojó un porcentaje de 62.16% en la calidad de las clasificaciones correctas.
    - **Macro Recall:** esta métrica nos arrojó un 47.16% en relación a la cantidad de clasificaciones que el modelo fue capaz de identificar.
    - **Macro F1score**: La comparación del rendimiento combinado de la precisión y la sensibilidad es del 0.5119.

2.2.2 Evaluación de los modelos generados para elegir el más eficiente

A continuación, se muestran los resultados obtenidos de las técnicas aplicadas:

Figura 15. Resultados por técnica

	index	Model	Accuracy	Macro Precision	Macro Recall	Macro F1score	Weighted Precision	Weighted Recall	Weighted F1	Time taken
0	0	LogisticRegression1	0.853397	0.537702	0.407251	0.431129	0.842206	0.853397	0.840511	4487.917441
1	0	RandomForest1	0.572964	0.136345	0.079394	0.060789	0.380484	0.572964	0.418547	736.667724
2	0	LinearSVC1	0.850990	0.621663	0.471623	0.511963	0.840762	0.850990	0.842664	63.616501

Nota: Resultados por técnica donde se evidencia la precisión o calidad en las clasificaciones correctas, la cantidad de clasificaciones identificadas por cada modelo y la combinación entre la precisión y la sensibilidad

8.2.3 Selección del modelo

De las métricas anteriormente mencionadas, se puede decir que, la técnica de Máquinas de Soporte Vectorial y Regresión Logística arrojaron resultados muy favorables, pero la técnica de Soporte Vectorial fue la más eficiente al arrojar un 62.16% en la calidad de las clasificaciones correctas.

8.3 Prueba del modelo seleccionado con un caso simulado de PQRSF aplicado en la Universidad Católica Luis Amigo (Fase 5 y 6 Crisp- Dm)

8.3.1 Generar un plan de prueba que permita evaluar los resultados (Fase 5)

Dando respuesta a la fase cinco, se realizó un plan de prueba para los casos simulados, se revisaron los resultados obtenidos con el fin de evaluar si la técnica seleccionada daba respuesta a los objetivos descritos anteriormente. Los casos simulados se modelaron con base en las áreas de atención según las necesidades de los usuarios y las categorías que se tienen actualmente en el sistema, tomando en cuenta esto, se definieron los casos dentro del rango de parámetros que el sistema pueda evaluar, primero hay que tener en cuenta las necesidades de los usuarios reales y que la redacción abarque en lo posible sus necesidades y segundo que estas estén enmarcadas dentro de las categorías de atención.

- **Escenarios de prueba**
    - Estudiante de Pregrado

      Un estudiante de Pregrado que apenas inicia su segundo semestre debe ingresar al sistema académico para validar cual es el aula donde está programada la materia y revisar si tiene lecturas previas a la clase en la plataforma virtual, pero en vacaciones perdió los apuntes de sus contraseñas (Casos de prueba 1,3, 4, 5).

    - Estudiante de posgrados

      Un estudiante de posgrados llega por primera vez a la universidad necesita llegar en su vehículo y prestar un libro en la biblioteca para una consulta de su especialización (Casos de prueba 9, 10).

    - Empleado

      Un empleado debe asistir a una charla virtual de Gestión de Calidad donde se darán indicaciones sobre los procedimientos de su área y debe descargar los formatos de la intranet, pero se acuerda que el equipo no tiene sonido y que la contraseña la cambie hace poco, pero la ha olvidado (Casos de prueba 2,7).

- Usuario interesado

Un usuario necesita consultarle a su hermano cuanto es el valor de los grados pues ya ha terminado todas las materias. Y está interesado en la carrera de Ingeniería de sistemas y quiere conocer más sobre esta (Casos de prueba 6, 8).

- **Casos simulados para la prueba:**
  - **Caso simulado #1:** Buenos días, tengo problemas para ingresar al correo electrónico y lo requiero para revisar las tareas de clase, si lo pudieran solucionar lo antes posible les agradecería.
  - **Caso simulado #2:** Buenos días, estoy presentando problemas con mi equipo, tengo problemas con el sonido, mil gracias y quedo atenta
  - **Caso simulado #3:** Reciba un cordial saludo, Solicito amablemente su colaboración para que me ayuden con el ingreso al sistema académico. Quedo atenta a su respuesta.
  - **Caso simulado #4:** Buenos días; solicito su colaboración para que me ayuden con el ingreso al sistema académico, mil gracias y saludos.
  - **Caso simulado #5:** Tengo inconvenientes para ingresar al curso virtual en la plataforma Dicom y necesito ingresar al curso de inglés virtual.
  - **Caso simulado #6:** Buenos tardes, me podrían indicar el valor del semestre de la carrera de ingeniería de sistemas.
  - **Caso simulado #7:** Buenas tardes, necesito recuperar usuario y clave para acceder a la intranet.
  - **Caso simulado #8:** Necesito conocer el valor que tienen los derechos de grado.
  - **Caso simulado #9:** Buenos tardes, me pueden indicar cuál es el valor del pago del parqueadero.
  - **Caso simulado #10:** Buenas, me pueden indicar como es el préstamo de libros de la biblioteca.

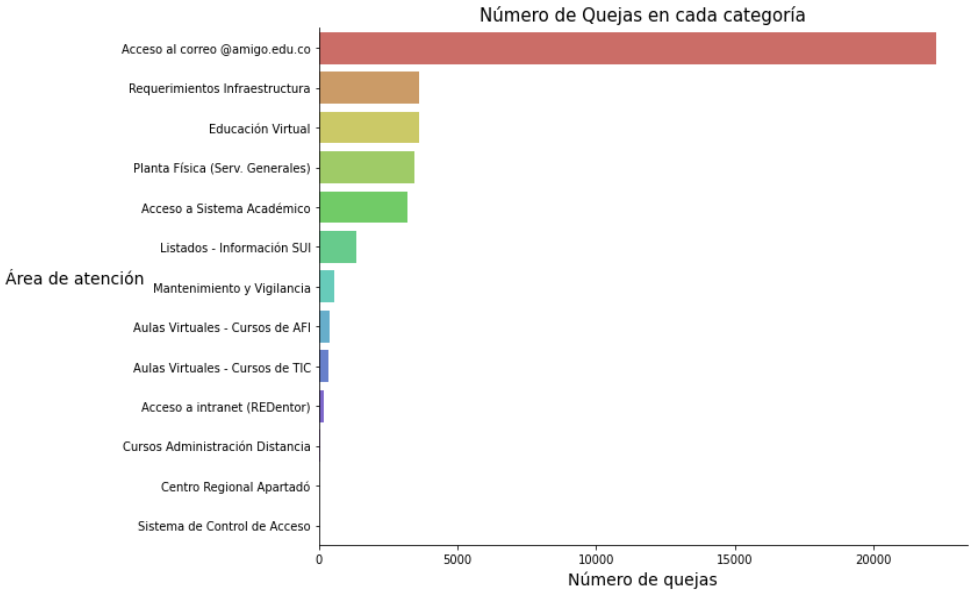
- **Resultados de los escenarios de prueba con casos simulados**

Para la evaluación se tomaron como referencia 4 escenarios a los cuales se aplicó el modelo ML con el aprendizaje, obteniendo los siguientes resultados:

  - Para un escenario habitual en el que se puede desenvolver un estudiante de pregrado, la clasificación de los casos de prueba fue **satisfactoria** en un 100%, 4/4 casos obtuvieron clasificaciones exitosas.
  - En el segundo escenario el modelo logra realizar la clasificación en un 50%, ya que identifica la pregunta del parqueadero, por lo tanto, se debe verificar si las palabras de la otra solicitud no se encuentran en el porcentaje de entrenamiento y dado el caso se debería reentrenar.
  - Para el tercer escenario, donde un empleado se desenvuelve en su entorno laboral, el modelo logra la clasificación **satisfactoria en un 100% de** los 2 casos de prueba, siendo ambos muy distantes en cuanto al tipo de consulta.
  - Para el cuarto escenario donde la consulta hace referencias a valores, el modelo no logra realizar ninguna de las clasificaciones correctamente, se debe revisar como es el comportamiento de estas consultas, palabras y peso que tienen en la categoría.
  - Los resultados obtenidos de los casos simulados, clasificados por el modelo ML, permite inferir sobre la efectividad del modelo y la técnica seleccionada, ya que de 10 casos probados se logró tener la clasificación **satisfactoria** de 7 casos equivalente al 70%.
  - De acuerdo con las PQRSF evaluadas, el sistema, cuenta con un 31% de solicitudes con una clasificación errónea, el objetivo del indicador está pensado para realizar una reducción del 15%, donde el margen de error o de clasificaciones erradas sería 16%, el algoritmo deberá tener una efectividad de clasificación del 84% mínimamente para cumplir.
  - Lograr tener una clasificación **satisfactoria alta permitirá reducir los tiempos del servicio y se traduce para el usuario en una atención oportuna.**
  - Los tiempos de respuesta de las PQRSF es de 14,8 que son los a casos que recibieron alguna respuesta, pero no necesariamente en ese punto están solucionados o cerrado, el caso para esto el indicador es el tiempo de servicio que está en 17,1 horas.

- Los tiempos de servicio del promedio ponderado de las categorías de clasificación de las PQRSF es de 17,1 Horas, pero tenemos categorías que superan los tiempos de servicio llegando hasta 90,8 horas (Centro Regional Apartadó) y otros muy cercanos a los tiempos de servicio SLA en 42,5 (Cursos Administración Distancia)

Figura 16. Número de PQRSF por categoría

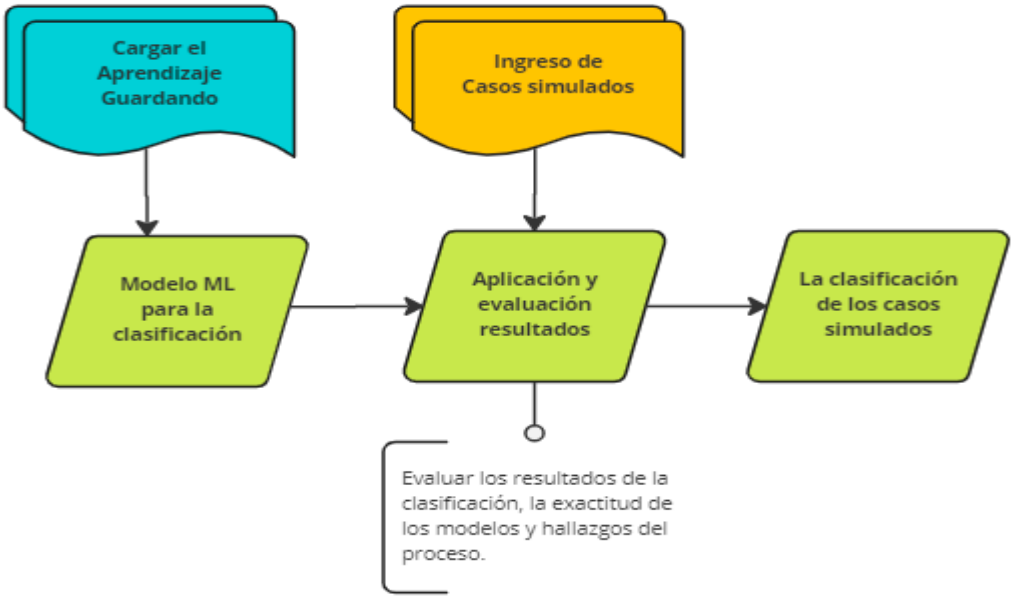


Fuente: Elaboración propia

La gráfica muestra que la mayoría de las solicitudes están relacionadas con el acceso al correo de dominio institucional, seguido de requerimiento de infraestructura, educación virtual, planta física, acceso al sistema académico.

8.3.2 Planear la implementación (Fase 6)

Figura 17Estructura plan de implementación



Fuente: Elaboración propia

En la figura 17 se representa el procedimiento que se realizará como evaluación experimental, donde se carga el aprendizaje del modelo previamente guardado, se aplica a los casos simulados al modelo ML, con el fin obtener una clasificación.

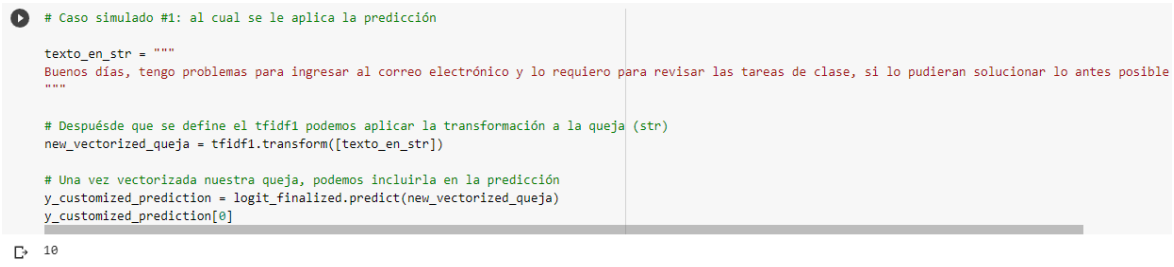
Una vez generadas las preguntas de los casos simulados (PQRSF) se ingresaron al modelo donde se realizó la clasificación para que este arroje la categoría a la cual serán asignados. En el entorno de Google Colab donde se implementó el código en Python se ingresaron los casos a través de la variable **texto_en_str** la cual se vectorizo, se cargan los datos del aprendizaje que previamente han sido guardados para realizar la predicción y evaluar los resultados.

Para ello se realizaron 10 casos simulados con el fin de evaluar los resultados de clasificación que arroja el modelo, a continuación, listamos los resultados

- caso simulado 1 = 10 (Acceso al correo @amigo.edu.co) ✅
- caso simulado 2 = 12 (Requerimientos Infraestructura) ✅

- caso simulado 3 = 1 (Acceso a Sistema Académico) ✅
- caso simulado 4 = 1 (Acceso a Sistema Académico) ✅
- caso simulado 5 = 13 (Educación Virtual) ✅
- caso simulado 6 = 18 (Planta Física (Ser. Generales)) ❌
- caso simulado 7 = 16 (Acceso a intranet (REDentor)) ✅
- caso simulado 8 = 13 (Educación Virtual) ❌
- caso simulado 9 = 24 (Mantenimiento y Vigilancia) ✅
- caso simulado 10 = 12 (Requerimientos Infraestructura) ❌

Figura 18Caso simulado 1

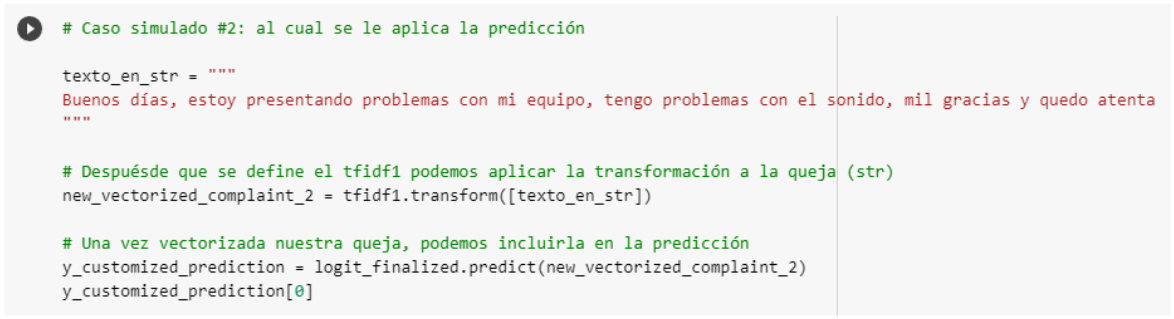


```
# Caso simulado #1: al cual se le aplica la predicción

texto_en_str = """
Buenos días, tengo problemas para ingresar al correo electrónico y lo requiero para revisar las tareas de clase, si lo pudieran solucionar lo antes posible
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
new_vectorized_queja = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(new_vectorized_queja)
y_customized_prediction[0]
```

```
10
```

Nota: *Caso simulado de un usuario con problemas para acceder al correo fue clasificado correctamente con el ID = 10 que corresponde a Acceso al correo @amigo.edu.co*

Figura 19Caso simulado 2

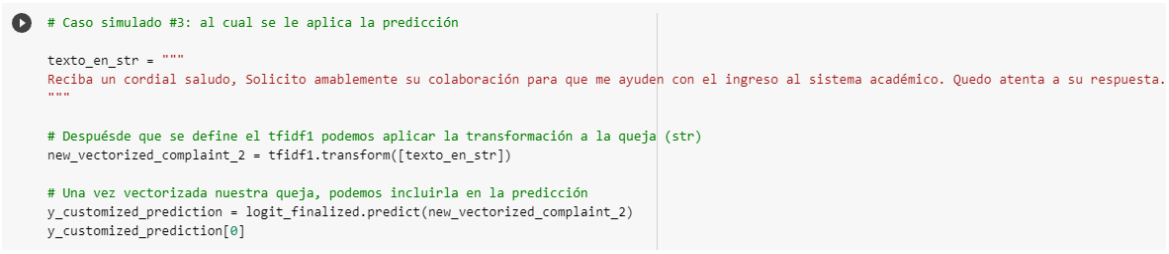


```
# Caso simulado #2: al cual se le aplica la predicción

texto_en_str = """
Buenos días, estoy presentando problemas con mi equipo, tengo problemas con el sonido, mil gracias y quedo atenta
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
new_vectorized_complaint_2 = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(new_vectorized_complaint_2)
y_customized_prediction[0]
```

```
12
```

Nota: Caso simulado de un usuario con problemas con su equipo de cómputo fue clasificado correctamente con el ID = 12 que corresponde al Requerimiento de infraestructura

Figura 20Caso simulado 3

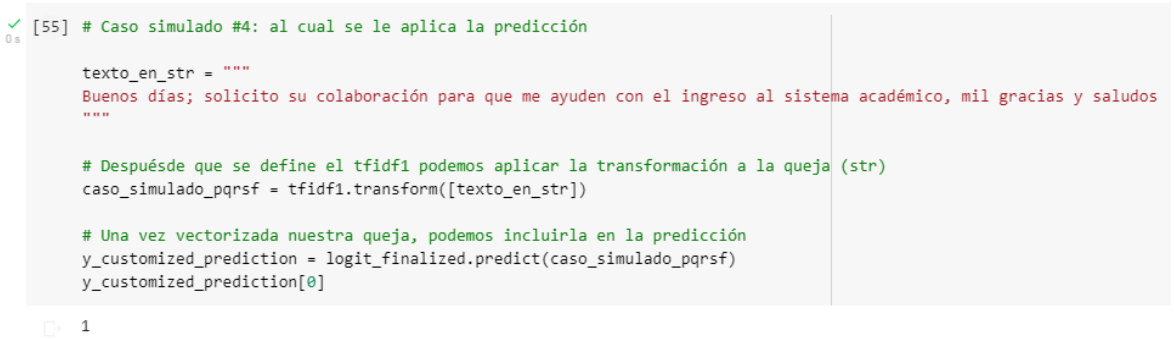


```
# Caso simulado #3: al cual se le aplica la predicción

texto_en_str = """
Reciba un cordial saludo, Solicito amablemente su colaboración para que me ayuden con el ingreso al sistema académico. Quedo atenta a su respuesta.
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
new_vectorized_complaint_2 = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(new_vectorized_complaint_2)
y_customized_prediction[0]
```

```
1
```

Nota: Caso simulado de un usuario con problemas para acceder al sistema académico fue clasificado correctamente con el ID = 1 que corresponde al Sistema Académico

Figura 21. Caso simulado 4

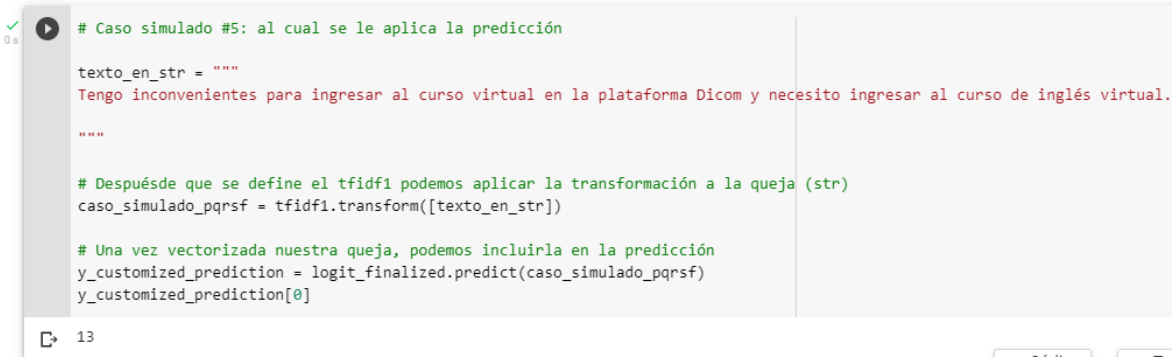


```
[55] # Caso simulado #4: al cual se le aplica la predicción

texto_en_str = """
Buenos días; solicito su colaboración para que me ayuden con el ingreso al sistema académico, mil gracias y saludos
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
1
```

Nota: Caso simulado de un usuario con problemas para acceder al sistema académico fue clasificado correctamente con el ID = 1 que corresponde al Sistema Académico

Figura 22.Caso simulado 5

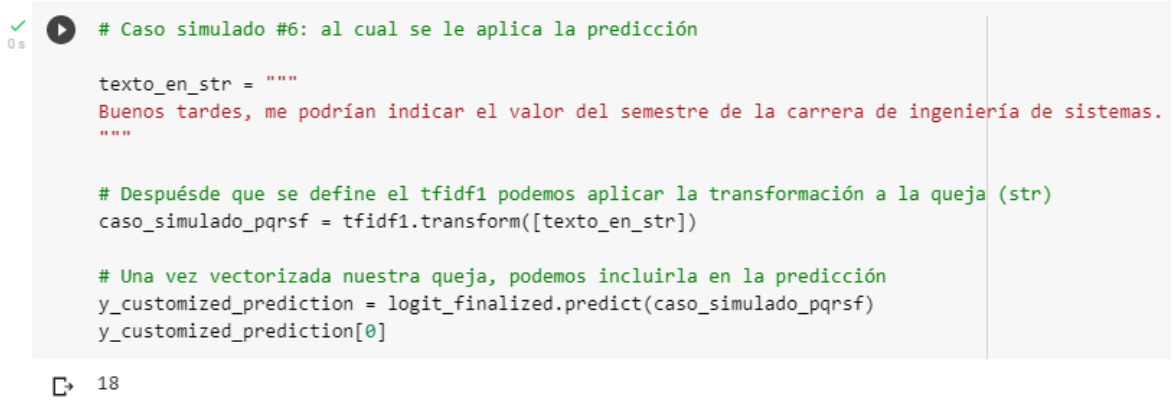


```
# Caso simulado #5: al cual se le aplica la predicción

texto_en_str = """
Tengo inconvenientes para ingresar al curso virtual en la plataforma Dicom y necesito ingresar al curso de inglés virtual.

"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
13
```

Nota: Caso simulado de un usuario con problemas para acceder al sistema académico DICOM fue clasificado correctamente con el ID = 13 que corresponde a Educación Virtual

Figura 23. Caso simulado 6

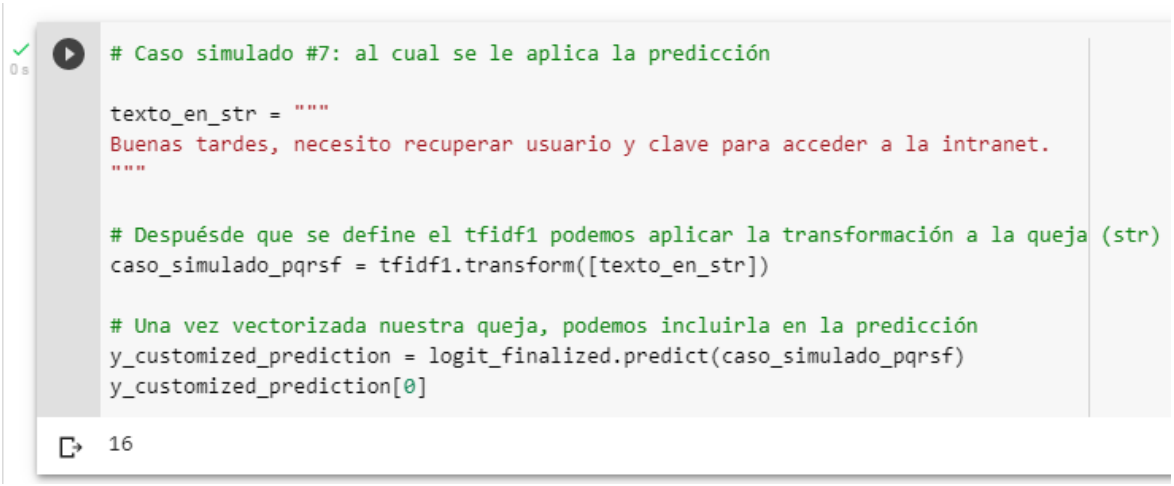


```
# Caso simulado #6: al cual se le aplica la predicción

texto_en_str = """
Buenos tardes, me podrían indicar el valor del semestre de la carrera de ingeniería de sistemas.
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
18
```

Nota: Caso simulado de un usuario preguntando sobre el valor semestral de la carrera de ingeniería de sistema fue mal clasificado con el ID = 18 que corresponde a Planta Física (Ser. Generales). No fue satisfactoria ya que clasificó una etiqueta que no tiene relación con el texto descrito.

Figura 24.Caso simulado 7



```
# Caso simulado #7: al cual se le aplica la predicción

texto_en_str = """
Buenas tardes, necesito recuperar usuario y clave para acceder a la intranet.
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
16
```

Nota: Caso simulado de un usuario con problemas para acceder a la internet, fue clasificado correctamente con el ID = 16 que corresponde a Acceso a intranet (REDntor)

Figura 25.Caso simulado 8

```
# Caso simulado #8: al cual se le aplica la predicción

texto_en_str = """
Necesito conocer el valor que tienen los derechos de grado."""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
13
```

Nota: Caso simulado de un usuario preguntando sobre el valor del derecho de grado fue mal clasificado con el ID = 13 que corresponde a Educación virtual. No fue satisfactoria ya que clasificó una etiqueta que no tiene relación con el texto descrito.

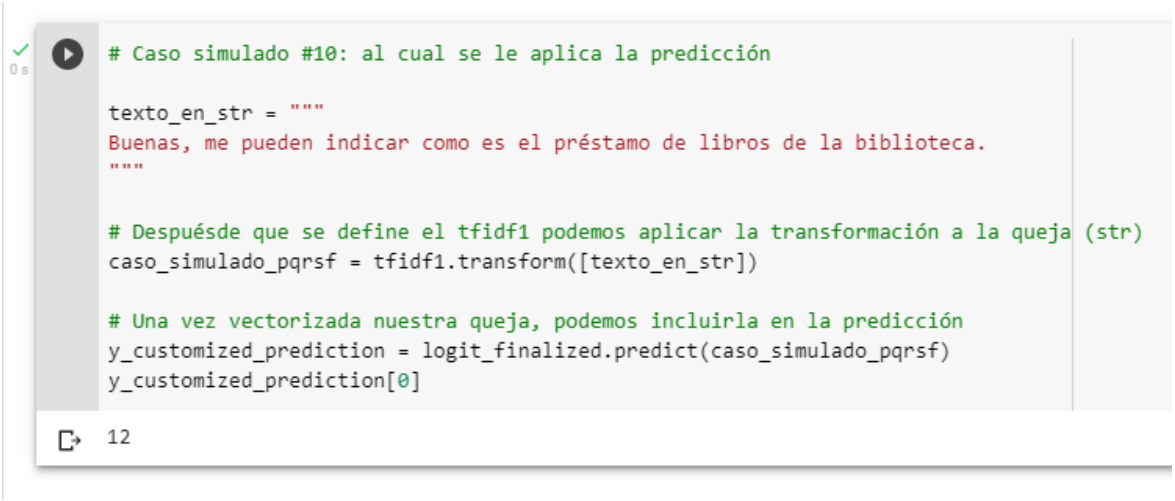
Figura 26.Caso simulado 9

```
# Caso simulado #9: al cual se le aplica la predicción

texto_en_str = """
Buenos tardes, me pueden indicar cuál es el valor del pago del parqueadero.

"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
24
```

Nota: Caso simulado de un usuario con problemas para el pago del parqu eadero, fue clasificado correctamente con el ID = 24 que corresponde a Mantenimiento y vigilancia

Figura 27.Caso simulado 10



```
# Caso simulado #10: al cual se le aplica la predicción

texto_en_str = """
Buenas, me pueden indicar como es el préstamo de libros de la biblioteca.
"""

# Despuésde que se define el tfidf1 podemos aplicar la transformación a la queja (str)
caso_simulado_pqrsf = tfidf1.transform([texto_en_str])

# Una vez vectorizada nuestra queja, podemos incluirla en la predicción
y_customized_prediction = logit_finalized.predict(caso_simulado_pqrsf)
y_customized_prediction[0]
```

```
12
```

Nota: Caso simulado de un usuario preguntando sobre los préstamos de libros en la biblioteca fue mal clasificado con el ID = 12 que corresponde a Requerimientos Infraestructura. No fue satisfactoria, ya que clasificó una etiqueta que no tiene relación con el texto descrito.

9. DISCUSIÓN

El uso de la metodología Crisp-DM, en este trabajo de grado permitió crear un modelo de minería de texto que posibilitara la clasificación de las PQRSF. Por lo tanto, de los tres objetivos específicos propuestos, se pudieron alcanzar los tres en su totalidad. Además, al margen de estos objetivos, se logró identificar que se puede realizar una mejora en la clasificación si se agrega el campo asunto dentro de los datos elegidos para el proceso, debido a que su contenido ayudaría a mejorar el modelo.

Lo anterior es coherente con lo hallado por Galán (2015), quien expuso que al aplicar la metodología CRISP-DM, le permitió encontrar un plan de extracción, normalización, y codificación de datos para la realización de procesos de minería de datos cuatrimestrales. Lo que quiere decir que esta metodología permite optimizar procesos en el contexto universitario.

En las fases 1, 2 y 3 se realizó un mayor esfuerzo debido a que se debía conocer el negocio, comprender los datos y realizar la preparación de los mismos sobre el conjunto de datos en el que se va a trabajar.

Se realizaron consultas a la base de datos en SQL para obtener información suficiente, tener entendimiento de los datos y cómo estos se acoplaban con el proceso. Cuando ya se tenía claridad de donde y como se encontraba la información se realizó la consulta para armar el dataset.

Se realizó una selección de técnicas de modelado para la construcción del modelo, en el lenguaje de programación python ya que este nos permite implementar una gran variedad de librerías para ML, gráficas e implementación de mediciones para su evaluación, la herramienta para realizar el despliegue se escogió Google Colab ya que gracias a su versatilidad nos facilitó la aplicación de las técnicas, evaluación, así mismo nos permitió determinar cuáles eran las más adecuadas para nuestro objetivo principal. Por último, una vez obtenidos los modelos, se analizaron para determinar la adecuación o no de los mismos. En este caso determinamos que las técnicas 1 y 3 (Regresión logística y Máquinas de soporte vectorial) podrían ser válidas para nuestros objetivos y se descartó la técnica 2 (Random Forest) por no ser lo suficientemente eficiente.

Para una implementación a nivel empresarial se debía profundizar en cada una de las etapas de la minería de texto, realizando una inspección más a fondo sobre los datos, validando que toda la información esté bien categorizada con su debida etiqueta, en la preparación de los datos a realizar las etapas de lematización y PNL o procesamiento de lenguaje natural, para tener una mejor calidad de los datos antes de la implementación de las técnicas de ML, en el modelamiento de los datos es importante implementar técnicas compuestas que logren una mejor precisión y sensibilidad, además de implementar métricas como Curva ROC, coeficiente de determinación, o R2, el método de la silueta, para realizar una evaluación de los modelos.

En la implementación de este proyecto de trabajo de grado fue necesario utilizar PyScript, ya que este permite crear aplicaciones ricas de Python en el navegador utilizando la interfaz de HTML, en este se incluirá el código Python donde se aplicará el aprendizaje del modelo ML entrenado guardado previamente, que será aplicado por medio de un formulario conectado con la API de osTicket. En adelante, los pasos a seguir siguen siendo los mismos que se han descrito en el desarrollo de trabajo, que van desde la comprensión del negocio hasta la implementación.

Con respecto a la clasificación de datos etiquetados, en el caso de investigación un sistema de PQRSF es una tarea de complejidad que tiene diferentes factores que influyen, en el comportamiento del modelo ML y de los resultados. Por lo tanto, se recomienda que en cada periodo se realice una retroalimentación de los datos para mejorar la clasificación mediante el reentrenamiento.

Según Paramesh & Shreedhara (2019) el rendimiento de los sistemas de clasificación tradicionales se puede mejorar aún más mediante el uso de varios conjuntos de técnicas de clasificación, como combinar las predicciones de diferentes modelos. De los diferentes modelos de clasificación para este ejercicio se considera que la técnica de máquinas de soporte vectorial (SVM) tuvo un mejor desempeño en comparación con otros modelos.

Por otro lado, la mayoría de los artículos investigados que fueron en inglés permitieron conocer que con técnicas de machine learning se ha logrado generar modelos más precisos para una clasificación, en cambio, con el idioma español es un poco más complejo de generar, en este trabajo, se desarrollaron varias técnicas de clasificación para lograr identificar a cuál área pertenece la PQRSF de la Universidad Católica Luis Amigó,

dando como resultado que el mejor para estos casos serían las máquinas de soporte vectorial.

10. CONCLUSIÓN

Tomando en cuenta el sistema de información, el cual permite medir la capacidad de respuesta inmediata y oportuna a los usuarios tanto internos y externos, en este trabajo se evidenciaron hallazgos que permitieron detectar que la información no cuenta con filtro que agilice la respuesta de manera pertinente. En este sentido, es importante determinar, cuál es la forma más propicia de generar un sistema de clasificación de información suministrada por los usuarios mediante un algoritmo que permita clasificar las peticiones, quejas, reclamos, sugerencias y felicitaciones que lleguen de manera directa a cada una de las áreas responsables de gestión de la universidad y su atención se dé manera eficiente, eficaz y efectiva (pertinente para los usuarios).

En cuanto a la caracterización del proceso de PQRSF de la Universidad Católica Luis Amigó, se logra concluir que, una vez identificado como funciona el sistema actual de PQRSF, cuáles han sido las fallas detectadas en el proceso y su manejo internamente se busca implementar un modelo de clasificación automático para lograr obtener un mejoramiento en los tiempos de respuesta y de esta manera mejorar la satisfacción de los usuarios.

El proceso más complejo durante el desarrollo del algoritmo de clasificación está relacionado con la limpieza de los datos, este está incluido en la fase tres de la metodología CRISP-DM. ya que en algunos casos la fuente o descripción de la PQRSF no era clara, tenía redacciones incoherentes, caracteres especiales, correos electrónicos entre otros que no permitían realizar una clasificación efectiva, por lo tanto, se debieron aplicar técnicas de limpieza como son la tokenización y los stopword donde el primero separaba las palabras y el segundo elimina aquellas palabras que no representan mayor valor para la clasificación.

Después de haber aplicado el Machine Learning (ML) y verificar que es eficiente para la clasificación de las PQRSF, se determina que la técnica más eficiente para la clasificación es la de máquinas de soporte vectorial, ya que logró clasificar una mayor cantidad de solicitudes maximizando el margen de separación entre las clases, así mismo, se logró identificar que cada categoría tiene una cantidad de datos diferentes, es decir, las que tienen más registros son las que permiten obtener una mayor cantidad de aciertos; por

otro lado, si las palabras empleadas para la clasificación de las PQRSF tienen palabras que generen peso en diferentes categorías pueda darse a una mala clasificación.

Después de haber aplicado el Machine Learning (ML) y verificar que es eficiente para la clasificación de las PQRSF, se determinó que la técnica más eficiente para la clasificación es la de máquinas de soporte vectorial, ya que logró clasificar una mayor cantidad de solicitudes maximizando el margen de separación entre las clases, así mismo, se logró identificar que cada categoría tiene una cantidad de datos diferentes, es decir, las que tienen más registros son las que permiten obtener una mayor cantidad de aciertos; por otro lado, si las palabras empleadas para la clasificación de las PQRSF tienen palabras que generen peso en diferentes categorías pueda darse a una mala clasificación.

Una vez aplicadas las tres técnicas de clasificación se podría decir que las técnicas de regresión logística y máquinas de soporte vectorial son las más eficientes para el modelo, pero al revisar la variable de precisión que determina la calidad de las clasificaciones correctas y el Recall que indica la cantidad de clasificaciones que el modelo fue capaz de identificar se encontró que la más eficiente es la clasificación realizada con la técnica de máquinas de soporte vectorial.

Teniendo en cuenta el objetivo de probar el modelo, se obtuvo como resultado, la efectividad de la técnica seleccionada, ya que de 10 casos probados se logró tener la clasificación efectiva de 7 equivalentes al 70%, el 30% restante al no estar dentro del dominio para evaluar, no logra la clasificación efectiva y es por esto que se debe mejorar el proceso mediante el reentrenamiento de los datos en unos tiempos específicos.

Finalmente, la aplicación de los casos simulados arrojó que algunas categorías de clasificación como: Acceso al correo @amigo.edu.co, Requerimientos Infraestructura, Educación Virtual, Planta Física (Serv. Generales), Acceso a Sistema Académico, Acceso a intranet (REDentor), obtuvieron mejores resultados que otros y una efectividad del 70% en el tema de clasificación correcta, pero al compararlo con el total de los dato se observa que el modelo de clasificación tiene más efectividad en las categorías, lo que permite entender que donde más datos se tienen el modelo logra una predicción de 0.84% de la precisión ponderada.

11. REFERENCIAS (APA)

Amine, M., El Filali, S., Aarika, K., Benlahmar, E., Rachida, A. y Debauche, O. (2021). Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis, Procedia Computer Science, Volume 191, Pages 487-492, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.07.062. https://www.sciencedirect.com/science/article/pii/S1877050921014629

Arvinder, K. & Chopra, D. (2016). Comparación de herramientas de minería de texto. En Conferencia internacional sobre confiabilidad, tecnologías de infocom y optimización (ICRITO) (Tendencias y direcciones futuras), 7-9 de septiembre de 2016.

Asamblea Departamental de Antioquia (s.f.). PQRSD. [en línea] (consulta realizada el 8 de septiembre de 2022). https://www.asambleadeantioquia.gov.co/pqrsd/

Asamblea Nacional Constituyente (1991). Constitución Política de Colombia. https://pdba.georgetown.edu/Constitutions/Colombia/colombia91.pdf

Becerril, I. y Villa, G. (2018). Reestructuración del sistema de atención a quejas y reclamos. Revista Ciencia Administrativa,2:65-80. http://eds.a.ebscohost.com/eds/detail/detail?vid=2&sid=f34b7812-2201-4c02-a6c0-230d7773f940%40sdc-v-sessmgr03&bdata=Jmxhbmc9ZXMmc2l0ZT1lZHMtbGl2ZQ%3d%3d#AN=138598749&db=bth

Berry, M. W. y J. Kogan (2010). Text Mining: Applications and Theory. Chicester, GB: Wiley, 2010.

Calancha Zúñiga, N. (2011). Breve aproximación a la técnica del árbol de decisiones. https://www.academia.edu/29655526/Aproximaci%C3%B3n_a_la_T%C3%A9cnica_de_%C3%81rbol_de_Decisiones?bulkDownload=thisPaper-topRelated-sameAuthor-citingThis-citedByThis-secondOrderCitations&from=cover_page

Congreso de la República de Colombia (junio 30 de 2015). Ley 1755. Por medio de la cual se regula el Derecho Fundamental de Petición y se sustituye un título del Código de Procedimiento Administrativo y de lo Contencioso Administrativo. http://www.secretariasenado.gov.co/senado/basedoc/ley_1755_2015.html

Congreso de la República de Colombia (diciembre 23 de 2014). Ley 1740. Por la cual se desarrolla parcialmente el artículo 67 y los numerales 21, 22 y 26 del artículo 189 de la constitución política, se regula la inspección y vigilancia de la educación superior, se modifica parcialmente la ley 30 de 1992 y se dictan otras disposiciones. https://www.mineducacion.gov.co/portal/normativa/Leyes/350383:Ley-1740-de-Diciembre-23-de-2014

Congreso de la República de Colombia (marzo 6 de 2014). Ley 1712. Por medio de la cual se crea la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional y se dictan otras disposiciones. https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=56882

Galán Cortina, V. (2015). Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario. Universidad Carlos III de Madrid. https://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf

Hassani, H., Beneki, C., Unger, S., Taj Mazinami, M. y Reza M. (2020). Text Mining in Big Data Analytics. Big Data Cogn. Comput. 2020, 4, 1; doi:10.3390/bdcc4010001. https://www.mdpi.com/2504-2289/4/1/1

Jiménez – Beltrán J. y Cadena-Carter, M. (2015). Software libre para la gestión de peticiones, quejas, reclamos y felicitaciones con énfasis en la cocreación. Facultad de Ingeniería de Sistemas Universidad Autónoma de Bucaramanga Bogotá, Colombia. https://repository.unab.edu.co/bitstream/handle/20.500.12749/3388/2015_Articulo_Jimenez_Beltran_Javier_Hernan.pdf?sequence=2&isAllowed=y

Matplotlib.org (s.f). Matplotlib: visualización con Python[en línea] (Consulta realizada el 20 de octubre de 2022). https://matplotlib.org/

Numpy. Org (s.f). NumPy documentation. [en línea] (Consulta realizada el 20 de octubre de 2022). https://numpy.org/doc/stable/

pandas.pydata.org (s.f). 10 minutos para pandas [en línea] (Consulta realizada el 20 de octubre de 2022). https://pandas.pydata.org/docs/user_guide/10min.html

Paramesh, S.P., Shreedhara, K.S. (2019). Automated IT Service Desk Systems Using Machine Learning Techniques. In: Nagabhushan, P., Guru, D., Shekar, B., Kumar, Y.

(eds) Data Analytics and Learning. Lecture Notes in Networks and Systems, vol 43. Springer, Singapore. https://doi.org/10.1007/978-981-1t3-2514-4_28

Revina, A., Buza, K., y Meister, V. (2020). IT Ticket Classification: The Simpler, the Better. *IEEE Access*, vol. 8, pp. 193380-193395. https://ieeexplore.ieee.org/document/9234428

Riefer, M. & Ternis, S. & Thaler, T. (2016). Mining Process Models from Natural Language Text: A State-of-the-Art Analysis. https://www.researchgate.net/publication/298020400_Mining_Process_Models_from_Natural_Language_Text_A_State-of-the-Art_Analysis

Subasi, A. (2020). Capítulo 3 – Técnicas de Aprendizaje Automático en Aprendizaje Automático Practico para el Análisis de Datos con Python. ISBN 9780128213797. Técnicas de aprendizaje automático - ScienceDirect

Tolciu, D., Sacarea, C. and Matei, C. (2021). "Analysis of Patterns and Similarities in Service Tickets using Natural Language Processing," in *Journal of Communications Software and Systems*, vol. 17, no. 1, pp. 29-35, February 2021, doi: 10.24138/jcomss. v17i1.1024. https://jcoms.fesb.unist.hr/10.24138/jcomss.v17i1.1024/#

Vallalta Rueda, J. (s.f). aprendizaje supervisado y no supervisado. En Health Data Miner [en línea] (Consulta realizada el 9 de septiembre de 2022). https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/

Yan, Z., Yun, Q. y Cuixia, L. (2017). Improved KNN Text Classification Algorithm with MapReduce Implementation. http://ieeexplore.ieee.org/document/8248509/

Honglei, Z., Zhenbo, Z., Hongjun, Z., Irfan, M. y Asim A. (2022). Big data-assisted social media analytics for business model for business decision making system competitive analysis, Information Processing & Management,Volume 59, Issue 1,102762,ISSN 0306-4573. https://www.sciencedirect.com/science/article/pii/S0306457321002430

Weber, R. (2000). Data Mining en la Empresa y en las Finanzas Utilizando Tecnologías Inteligentes. Revista Ingeniería De Sistemas Volumen XIV, N.º 1, junio. 01-revista def.-OK (uchile.cl)

Zhong, N., Li, Y. and Wu, S. (2012)"Effective Pattern Discovery for Text Mining," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 1, pp. 30-44, Jan. 2012, doi: 10.1109/TKDE.2010.211. https://ieeexplore.ieee.org/document/5611523

Zicari, P., Folino, G., Guarascio, M., & Pontieri, L. (2021). Discovering accurate deep learning based predictive models for automatic customer support ticket classification. Proceedings of the ACM Symposium on Applied Computing, 1098–1101. **https://doi.org/10.1145/3412841.3442109**

Fuentes electrónicas

El libro de python (s.f.). cadenas Python. [en línea] (Consulta realizada el 22 de octubre de 2022). Cadenas en Python | El Libro De Python

HolyPyThon.com (s.f). Logística Regression Optimization. [en línea] (Consulta realizada el 22 de octubre de 2022). https://holypython.com/log-reg/logistic-regression-optimization-parameters/

IBM.com (agosto 17 de 2021). Conceptos básicos de ayuda de CRISP – DM [en línea] (Consulta realizada el 22 de octubre de 2022). https://www.ibm.com/docs/es/spss-modeler/saas?topic=guide-deployment

Seaborn.pydata.org (s.f).: Visualización estadística de datos. [en línea] (Consulta realizada el 22 de octubre de 2022). Seaborn: Visualización estadística de datos — Documentación de Seaborn 0.12.1 (pydata.org)

Anexo 1: Consulta SQL del sistema de PQRS

Consulta ticket interpuestos por los usuarios (total de registros históricos)

```
SELECT ticket_thread.ticket_id, help_topic.topic, ticket__cdata.subject,
ticket_thread.body, ticket_thread.created, ticket.closed, ticket_thread.thread_type
FROM ticket_thread
INNER JOIN ticket__cdata ON ticket_thread.ticket_id = ticket__cdata.ticket_id
INNER JOIN ticket ON ticket_thread.ticket_id = ticket.ticket_id
INNER JOIN help_topic  ON ticket.topic_id = help_topic.topic_id
AND thread_type = "M"
ORDER BY `ticket_thread`.`ticket_id` ASC
```

Mostrando registros 0 - 29 (38,981 total, La consulta tardó 0.6786 seg) [ticket_id: **5 - 42**]

Consulta de ticket interpuestos por los usuarios para el primer periodo de 2022

```
SELECT ticket_thread.ticket_id, help_topic.topic, ticket__cdata.subject,
ticket_thread.body, ticket_thread.created, ticket.closed, ticket_thread.thread_type
FROM ticket_thread
INNER JOIN ticket__cdata ON ticket_thread.ticket_id = ticket__cdata.ticket_id
INNER JOIN ticket ON ticket_thread.ticket_id = ticket.ticket_id
INNER JOIN help_topic  ON ticket.topic_id = help_topic.topic_id
AND thread_type = "M"
AND `ticket_thread`.`created` > '2021-12-31'
```

Mostrando registros 0 - 29 (4,214 total, La consulta tardó 0.2798 seg)

Consulta ticket reasignados (mal etiquetados) primer periodo del año 2022

```
SELECT ticket_thread.ticket_id, help_topic.topic, ticket__cdata.subject,
ticket_thread.body, ticket_thread.created, ticket.closed, ticket_thread.thread_type
FROM ticket_thread
INNER JOIN ticket__cdata ON ticket_thread.ticket_id = ticket__cdata.ticket_id
INNER JOIN ticket ON ticket_thread.ticket_id = ticket.ticket_id
INNER JOIN help_topic  ON ticket.topic_id = help_topic.topic_id
AND thread_type = "N"
AND `ticket_thread`.`created` > '2021-12-31'
WHERE `title` LIKE '%Ticket transferred%'
```

Mostrando registros 0 - 29 (1,302 total, La consulta tardó 0.0133 seg)

El campo **thread_type** Guarda los valores M, N , R

M= Tickets , N=Tickets Cambio de Estado(Ticket transferido de Departamento, Cerrados, Asignados), R = Tickets Respondidos