



**Creación de una herramienta para clasificar y analizar las causas de las quejas por servicio
de TIGO usando NLP**

Jhan Carlos Sánchez Serna

Trabajo de prácticas para optar por el título de ingeniero industrial

Asesores

Asesor interno: Juan Camilo España Lopera

Asesor externo: Gisselle Doriee Cardenas Restrepo

Universidad de Antioquia

Facultad de ingeniería

Ingeniería industrial

Medellín, Colombia

2023

Cita	(Sánchez, 2023)
Referencia	Sánchez Serna, J. (2023). <i>Creación de una herramienta para clasificar y analizar las causas de las quejas por servicio de TIGO usando NLP</i> . [Práctica empresarial]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio César Saldarriaga.

Jefe departamento: Mario Alberto Gaviria Giraldo.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Tabla de Contenido

Resumen	7
Abstract	8
Introducción	9
1 objetivos	14
1.1 Objetivo general	14
1.2 Objetivos específicos	14
2 Marco teórico	15
3. Metodología	21
3.1 Etapas	21
3.2 Cronograma de actividades	29
3.3 Presupuesto	30
4 Resultados	31
5 Análisis	52
6 Conclusiones	54
Referencias	57

Lista de tablas

Tabla 1 Resultados de modelos	17
Tabla 2 Muestra del diccionario de la BU de FIJO	23
Tabla 3 Paquetes del desarrollo	24
Tabla 4 Palabras eliminadas del texto	25
Tabla 5 Comparación de la limpieza de texto	26
Tabla 6 Campos de la tabla consolidada	27
Tabla 7 Presupuesto del proyecto	30
Tabla 8 Filtros de la ‘vista general’	31
Tabla 9 Filtros seleccionados	31
Tabla 10 Volumen de tópicos y criterios	38
Tabla 11 Porcentaje de variación de tópicos y criterios	39
Tabla 12 Comportamiento del top 10 por región	47

Lista de figuras

Ilustración 1 Ingresos de PQR's	9
Ilustración 2 Cierre de PQR's	10
Ilustración 3 Reincidencia queja factura	11
Ilustración 4 Porcentaje de PQR's	12
Ilustración 5 Diagrama de la solución final	19
Ilustración 6 Diseño de la solución	22
Ilustración 7 Porcentaje de tipificación	28
Ilustración 8 Cronograma del proyecto	29
Ilustración 9 Recalamaciones de queja servicio por mes	32
Ilustración 10 Porcentaje de tipificación de la herramienta	33
Ilustración 11 Cantidad de quejas servicio por canal	34
Ilustración 12 Comportamiento del top 10 de tópicos agrupados	35
Ilustración 13 Comportamiento del top 10 de tópicos por mes	36
Ilustración 14 Nube de palabras	40
Ilustración 15 Participación por BU	40
Ilustración 16 Participación por UE	41
Ilustración 17 Participación de tópico por canal	43
Ilustración 18 Participación de tópico por ciclo de facturación	44
Ilustración 19 Volumen de quejas por servicio por regional	45
Ilustración 20 Volumen de quejas por servicio por departamento	46
Ilustración 21 Tópicos por participación	47
Ilustración 22 Vista general	48
Ilustración 23 Vista detalle	49
Ilustración 24 Vista regional	50

Siglas, acrónimos y abreviaturas

NLP	Natural Language Processing
BU	Unidad de Negocio
EU	Unidad Estratégica
SIC	Superintendencia de industria y comercio
CRC	Comisión de Regulación de Comunicaciones
PQR'S	Peticiones, quejas, reclamos y recursos

Resumen

Uno de los retos más comunes entre los empresarios es el servicio al cliente, debido a que éste es un aspecto importante para el éxito. La administración de las peticiones, quejas y reclamos (PQR's) es una de las actividades de la empresa con mayor incidencia en la fidelización del cliente, por tal motivo las empresas han entendido que la gestión de dicha información es una de las mejores formas de generar valor en las compañías. De ahí que el objetivo principal del proyecto sea crear un modelo de NLP para el proceso de quejas por servicio en la Dirección de Transformación de la Experiencia en TIGO Colombia, esta iniciativa pretende analizar el malestar del usuario, por medio de la clasificación de las causas de contacto, con ayuda de la información encontrada en los textos tanto para la radicación, es decir el problema, según la información suministrada por el cliente, como para el texto de la solución dadas a las PQR's. De esta forma, se busca tener insights de todo el conjunto de molestias, como por ejemplo conocer los tópicos que más participación tienen para cierta unidad de negocio y unidad estratégica en cierto periodo de tiempo para todas las quejas por servicio; actualmente existe un modelo NLP solo para quejas por facturas, por tal motivo se pretende tener a través de visualizaciones información que apoye a la toma de decisiones dentro de la empresa, así como también, mayor oportunidad de gestión de cada caso con relación al cliente, análisis más a fondo la tipología, entendimiento de variables importantes en la herramienta y adicionalmente, indagar más allá de las necesidades del cliente con relación a procesos de radicación y solución del problema.

Palabras clave: PQR'S, queja servicio, queja factura, sistemas de información, proceso, servicio al cliente, petición, queja, reclamo, recurso y NLP.

Abstract

One of the most common challenges among entrepreneurs is customer service since this is an important aspect for success. The management of requests, complaints and claims (PQR's) is one of the activities with greater impact on customer loyalty, for this reason companies have understood that the management of such information is one of the best ways to generate value. Hence, the main objective of the project is to create a NLP model for the process of service complaints in the Direction of Experience Transformation in TIGO Colombia, this initiative aims to analyze the user discomfort, through classification of the causes of contact, using information found in the texts for both: the filing, ie the problem, according to information provided by customer, as for text of solution given to the PQR's. In this way, it is sought to have insights of the whole set of annoyances, such as knowing topics that have more participation for a certain business unit and strategic unit in a period of time for all service complaints. Currently there is a NLP model only for complaints about invoices, for this reason it is intended to have through visualizations information that supports decision making within the company, as well as greater opportunity to manage each case in relation to customer, more in-depth analysis of the typology, understanding of important variables in the tool and additionally, to inquire beyond the customer's needs in relation to processes of filing and solving the problem.

Keywords: PQR'S, service complaint, invoice complaint, information systems, process, customer service, petition, complaint, claim, appeal and NLP.

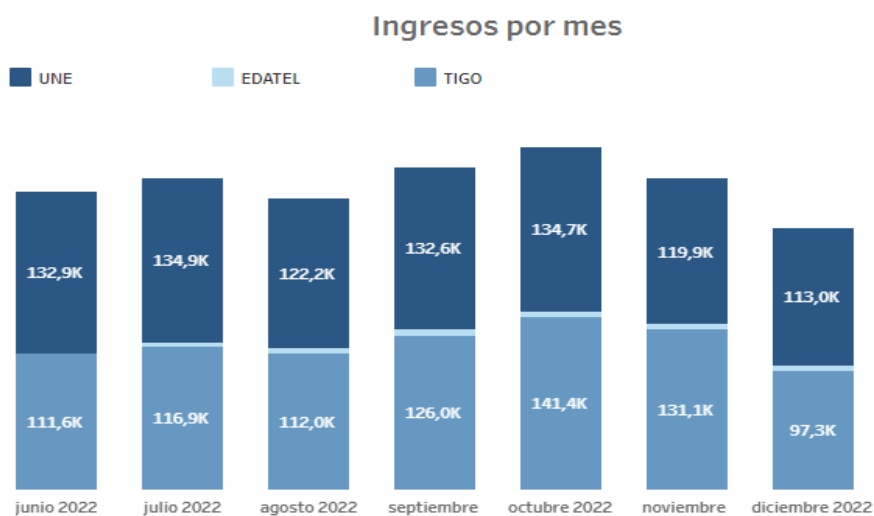
Introducción

Teniendo en cuenta que la ley 1755 de 2015 Colombia obliga a las empresas a tener un módulo de PQR's, conformado por: peticiones, quejas servicios, quejas facturas, reclamos y recursos, la empresa TIGO no es ajena a esta regulación y por medio de esta estructura, los clientes pueden exponer sus solicitudes en la forma y en el lenguaje que se desee, permitiéndole a la compañía poder conocer la voz del cliente respecto a sus inconformidades. Actualmente, existe una serie de canales en la empresa por donde puede ingresar el cliente y manifestar su malestar: carta, vía telefónica, visita física, página web, correo, redes sociales y WhatsApp. El tiempo máximo en el que se debe responder estas solicitudes no debe ser mayor a 15 días hábiles, tal como lo estipula la ley. Por otro lado, TIGO cuenta con un proveedor llamado INDRA el cual se encarga de hacerle cierre a cada una de las PQR's ingresadas.

El comportamiento del volumen de contactos de PQR's en los últimos siete meses no ha sufrido mayores variaciones, como se logra ver en la **Ilustración 1**, la cual muestra los ingresos de las tres compañías que cubre internamente TIGO, encontrando valores entre 210,3k y 276,1k.

Ilustración 1

Ingresos de PQR's

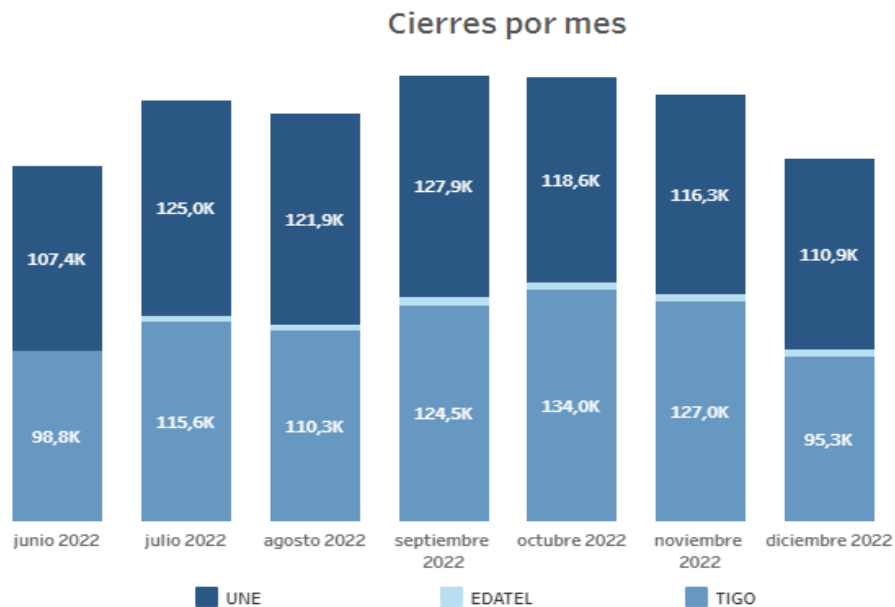


Fuente. (Tigo, 2022)

A su vez, el número de cierres, es decir la respuesta que se le da a cada uno de las PQR's es similar al volumen de ingresos, estas se pueden apreciar en la **Ilustración 2**, oscilando entre 206,2k y 252,4k con un horizonte de tiempo de los últimos siete meses.

Ilustración 2

Cierres de PQR's

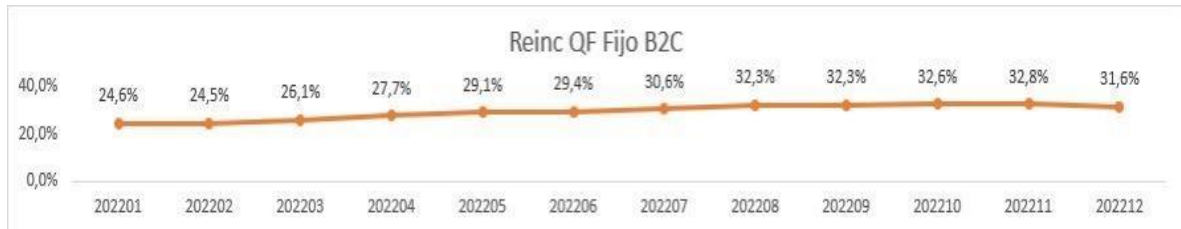


Fuente. (Tigo, 2022)

A pesar de no encontrar un patrón de crecimiento en las solicitudes de PQR's, con relación a la reincidencia de quejas factura para el área de FIJO B2C, es decir los servicios de telefonía, banda ancha, televisión para hogares visible en la **Ilustración 3**; para este tipo de solicitud puntualmente, si se ha hecho presente un incremento desde enero del 2022, hecho que genera una alerta para el área.

Ilustración 3

Reincidencia quejas factura



Fuente. (Tigo, 2022)

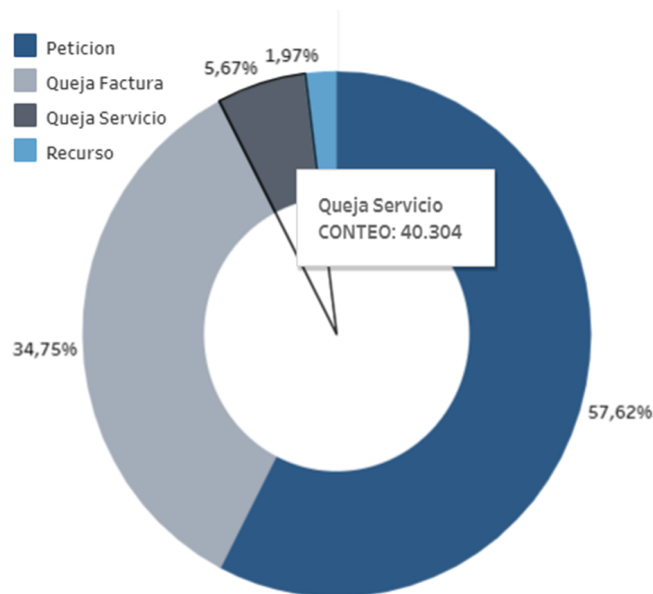
Con lo anterior, se pretende ilustrar el gran número de contactos que se tienen con relación a las PQR's y el alza de reincidencias para quejas factura generadas por los clientes. Al ser un alto volumen de información resulta complejo leer cada una de las solicitudes y asignarle una tipificación según el motivo del problema, teniendo en cuenta la diversa información del campo de texto según cada caso existente. De hecho, las personas encargadas de recibir las quejas deben asignar una categoría y subcategoría de cada caso, pero por diversos motivos como lo es la falta de trazabilidad de la información por los asesores, estas categorizaciones no son confiables y no es posible hacer uso de dicha información. De ahí, que sea necesario transformar los datos no estructurados de texto a data estructurada, para poder así extraer de una forma más explícita la información que el cliente transmite, con el objetivo de conocer con mayor detalle los problemas por las que una persona se siente inconforme con el servicio e ir más allá de las tipificaciones existentes.

Así mismo, en la actualidad existe un modelo NLP para quejas facturas, el cual representa el 34,4% de las PQR's como se observa en la **Ilustración 4**, el desarrollo de este modelo se encuentra en R, así como también un tablero en Tableau para visualizar lo encontrado en la herramienta, donde se logra mostrar la distribución de tópicos dependiendo la ubicación geográfica y demás unidades de negocio de la empresa. En este orden de ideas, es necesario conocer una porción más del espectro de PQR's debido a que el modelo actual solo muestra una parte de las quejas, se pretende construir un modelo de NLP por medio de Python para conocer la totalidad de las quejas, es decir, completar la información existente de quejas por facturas con la información de las quejas por servicio. Se elige este lenguaje de programación con la intención

de hacer una migración de R a Python, donde a su vez se muestre una priorización de tópicos en cada uno de los grupos de texto, puesto que, en ocasiones cada queja puede contener varias molestias y solo se pretende conocer el tópico con mayor fuerza, según la importancia que tenga para las diferentes áreas de la empresa. Además, que dicha priorización de tópicos no se encuentra en el modelo existente para quejas facturas. Cabe decir, que a pesar de que el volumen de peticiones es el más alto con un 57,62% del total de PQR's para el último trimestre del 2022, no se aplica NLP debido a que un área que está más controlada y se tiene mayor conocimiento respecto a lo que expresan los clientes, por lo que no es necesario crear la herramienta para este grupo de forma prioritaria.

Ilustración 4

Porcentaje de PQR's.



Fuente. (Tigo, 2022)

Los clientes desempeñan un papel activo en la generación de información relevante, siendo fuentes de conocimiento valiosas para las empresas (Villarroel & Burton, 2014). Lo anterior, apalancado en el fuerte avance tecnológico, el cual ha permitido la ampliación de la oferta de canales disponibles para que las empresas recojan las opiniones de los clientes. De ahí que, las empresas que pueden gestionar los datos de lo que comunican los clientes con relación a

sus experiencias con el servicio de una manera regular son, de media, un 5% más productivas y un 6% más rentables que sus competidores (McAfee, 2012).

La gestión de las PQR's permite la trazabilidad y escalamientos internos con el objetivo de garantizar respuesta adecuada al cliente, así como la medición, seguimiento y cumplimiento con los acuerdos operativos y regulatorios (SIC, CRC) con el fin de fortalecer el servicio para lograr la excelencia operativa. Por consiguiente, gestionar las PQR's de forma eficaz constituye un requisito necesario para que las empresas puedan recuperar la confianza de un cliente insatisfecho, donde además la solución rápida y efectiva de cualquiera de ellas, puede propiciar incluso una mejor relación con el cliente que la que se tenía antes de que se produjera la insatisfacción (Rodríguez, 2018).

El proyecto pretende suministrar información más relevante de las quejas para la toma de decisiones y a su vez, generar mayor facilidad para la comprensión del campo de texto, tanto para la radicación, es decir el problema, según la información suministrada por el cliente, como para el texto de la solución. Un método bastante eficiente para este requerimiento es el proceso de NLP, el cual ofrece una solución potencial para tratar enormes volúmenes de datos no estructurados, haciendo posible hacer un seguimiento de lo que comunican los clientes y la misma empresa a través de texto (Villaverde, 2017).

También se busca encontrar patrones generales, categorización de los tópicos para ser capaces de traducir la información en acciones de mejora de los diferentes procesos. Por otro lado, nutrir los insights existentes, conociendo los malestares más comunes en los clientes en las quejas, representado por tópicos, análisis más a fondo la tipología, entendimiento de variables importantes en la herramienta, como también indagar más allá las necesidades del cliente con relación a procesos de radicación y solución de problemas, esto permitirá mayor oportunidad de gestión de cada caso con relación al cliente; por ejemplo acordar con el proveedor INDRA que haga más énfasis a ciertos motivos de contacto o a algunas PQR's en específico para evitar su reincidencia.

A continuación, se muestra el desarrollo de la creación de la herramienta a través de los objetivos, marco teórico, metodología, resultados, análisis, conclusiones y recomendaciones de posibles trabajos futuros.

1 objetivos

1.1 Objetivo general

Crear una herramienta para clasificar y analizar las causas de las quejas por servicio inmersas en la estructura de PQR's, por medio de tópicos y criterios, con el fin de generar insight oportunos, con información más detallada y de mejor calidad, de esta manera, mejorar la toma de decisiones y seguimiento de los motivos de contacto de los clientes de TIGO COLOMBIA.

1.2 Objetivos específicos

- Plantear una solución analítica que permita una mayor comprensión sobre las causas de las quejas por servicio.
- Diseñar y desarrollar tableros analíticos con los resultados del modelo NLP de quejas servicio para el público objetivo, para hacer visible las tendencias en el tiempo y facilitar la comprensión de la información en el proceso de PQR's.
- Realizar un preprocesamiento y organización de la información de las quejas de servicio para reportar de manera oportuna a las áreas implicadas y alta gerencia que le permitan tomar acciones y hacer seguimientos.

2 Marco teórico

Las empresas han entendido que deben ofrecer una experiencia al cliente de alto valor, esto está altamente influenciado por la alta competencia en el mercado. La visión tradicional de tener un buen producto, respaldado por una buena marca, ya no es suficiente. En este momento las empresas están tomando conciencia de la importancia de poner al cliente en el centro y de la necesidad de generar buenas experiencias en todas y cada una de las interacciones que tienen con él (Barreiro, 2016). Una gran herramienta para gestionar lo que dicen los usuarios, o clientes, es

NLP o preprocesamiento de lenguaje natural, siendo este el que utilizamos para comunicarnos entre nosotros, hablando o escribiendo, pero no es fácil para un ordenador entender este lenguaje. Esta es la razón por la que, desde hace muchos años, se ha intentado crear métodos para procesar este lenguaje natural y convertirlo en algo que el ordenador pueda entender. NLP Tiene como objetivo leer, descifrar y comprender información de forma que pueda utilizarse posteriormente. Las aplicaciones son: traductores, Microsoft Word, Grammarly, OK Google, Alexa, Siri, chatBots, clasificadores de texto, extracción contextual, análisis de sentimientos, entre otros (Zayat, 2019).

Una empresa que ha utilizado NLP es ETB. Para esta compañía bogotana de telecomunicaciones, es importante que las solicitudes de reclamación que se reciban sean tratadas y resueltas oportunamente, garantizando la satisfacción del cliente. Es por eso, que hasta el 2016 venía realizando una serie de iniciativas para mejorar la atención al cliente, a pesar de ello, estos esfuerzos no han resultado suficientes para mejorar su gestión y a su vez reducir la cantidad de reclamos que mensualmente se reciben (Espinel, 2019)

Así pues, en pro de mejorar la calidad y experiencia de los clientes, se diseña una metodología basada en analítica de datos, que le permita a la Empresa de Telecomunicaciones procesar de manera eficiente los reclamos de los clientes del segmento Hogares y Personas. Para la resolución del problema, se implementó modelos que usen NLP; utilizando el sentido dado al análisis de texto se quiere usar los contactos escritos y transcritos de las quejas recibidas en el área de experiencia al cliente para clasificarlas correctamente; al ingresar al área correcta, se dará un manejo óptimo e inmediato reduciendo considerablemente el método y tiempo de respuesta. De otra parte, con base en la aplicación de modelos de clasificación de texto (NLP), fue posible identificar con mayor certeza la naturaleza de quejas recibidas y se identificó las áreas responsables de darles solución (Espinel, 2019).

Seguidamente, se muestran los pasos que se realizaron para la creación de los modelos.

Extracción de datos: se descargan y consolidan los datos remitidos por la ETB, obteniendo la base inicial para el análisis.

Análisis de los datos recibidos: descriptivas y tablas de contingencia de la información recibida.

Identificación de variables relevantes: depuración y definición de base de datos con la que se trabaja los modelos.

Construcción de los modelos: se identifican los modelos apropiados para el desarrollo de la solución.

Análisis de resultados: se obtiene la clasificación y modelo óptimo para la entrega de la solución.

Impacto de la solución: detalle del impacto económico, estratégico y experiencia al cliente que teniendo en cuenta la aplicación del modelo se espera tener vs la gestión actual.

A través de modelos de clasificación empleando NLP, se busca encontrar solución al problema antes mencionado. A continuación, se muestran dichos modelos: Regresión Logística, Decision Tree Classifier, Random Forest, Gradient Boosting y Redes Neuronales para establecer la clasificación adecuada (Espinel, 2019).

El concepto de NLP según (Espinel, 2019) incluye lo siguiente:

- Para el tratamiento de Stopwords, fue usada la librería “nltk”, que contiene palabras muy usadas que no aportan a la diferenciación de las clases, adicional a estas palabras contenidas en este arreglo, fue creada una lista de las palabras que específicamente para este conjunto de datos se repetían en muchas de las observaciones.
- Se crea función para limpiar el texto que a partir de expresiones regulares elimina números dentro del texto, tildes entre otros caracteres que no permiten unificar las palabras con el mismo significado.
- Se aplica la función “stemming” en el texto dejando cada palabra en su forma raíz.

- Para el proceso de vectorización se definieron como hiperparámetros 800 features y 1 ngrama.
- Se usa el 30% de los datos para la validación de los modelos.

En la **Tabla 1** se puede apreciar los resultados finales:

Tabla 1

Resultado de modelos

ACCURACY						
Categoría	Procedimiento Actual (manual)	Modelo Regresión logística Multinomial	Modelo Decision Tree	Modelo Random Forest	Modelo Gradiente Boosting	Modelo Redes Neuronales
Comercial	75%	87%	81%	79%	79%	84%
Técnica	67%	96%	79%	79%	79%	88%
Facturación	65%	91%	91%	95%	95%	98%
Experiencia	69%	80%	94%	84%	84%	82%
Accuracy Modelo	69%	87,16%	83,42%	84,49%	84,49%	87,16%

RECALL					
Categoría	Modelo Regresión logística Multinomial	Modelo Decision Tree	Modelo Random Forest	Modelo Gradiente Boosting	Modelo Redes Neuronales
Comercial	87%	91%	89%	89%	81%
Técnica	71%	77%	77%	77%	83%
Facturación	91%	96%	91%	91%	93%
Experiencia	92%	83%	80%	80%	90%
Recall Modelo	86,25%	86,88%	84,40%	84,40%	86,76%

F1-SCORE					
Categoría	Modelo Regresión logística Multinomial	Modelo Decision Tree	Modelo Random Forest	Modelo Gradiente Boosting	Modelo Redes Neuronales
Comercial	87%	86%	84%	84%	83%
Técnica	82%	78%	78%	78%	85%
Facturación	91%	93%	93%	93%	95%
Experiencia	85%	88%	82%	82%	86%
F1-Score Modelo	87,03%	86,55%	84,37%	84,37%	87,16%

Fuente: Espinel, (2019)

El modelo de red neuronal presenta resultados similares a la regresión logística multinomial, en cuando a la precisión; a pesar de lo anterior, teniendo en cuenta la facilidad de implementación, velocidad de procesamiento para grandes cantidades de información, facilidad de interpretación de resultados e infraestructura de cómputo, se sugiere utilizar la regresión

logística multinomial, para la clasificación de reclamos de las UEN Hogares y Personas de ETB (Espinel, 2019).

Con base en la aplicación de modelos de clasificación de texto (NLP), fue posible identificar con mayor certeza la naturaleza de quejas recibidas y se identificó las áreas responsables de darles solución. Finalmente, con la correcta clasificación de las quejas, se tendrá un proceso controlado, con una identificación más acertada de los responsables de dar solución, lo que permitirá hacer un seguimiento más detallado y oportuno (Espinel, 2019).

Otro caso puntual donde se puede apreciar la aplicación de NLP, es un asistente virtual tipo chatBot, proyecto trabajado por el equipo de Mentes Colectivas de la Pontificia Universidad Javeriana. Este sistema permite la atención al usuario en horario no hábil, con el fin de acceder a recursos que atiendan de manera preliminar sus necesidades e informar al equipo de mentes colectivas de estos usuarios atendidos en esta franja horaria para realizar un seguimiento o agendamiento de una cita.

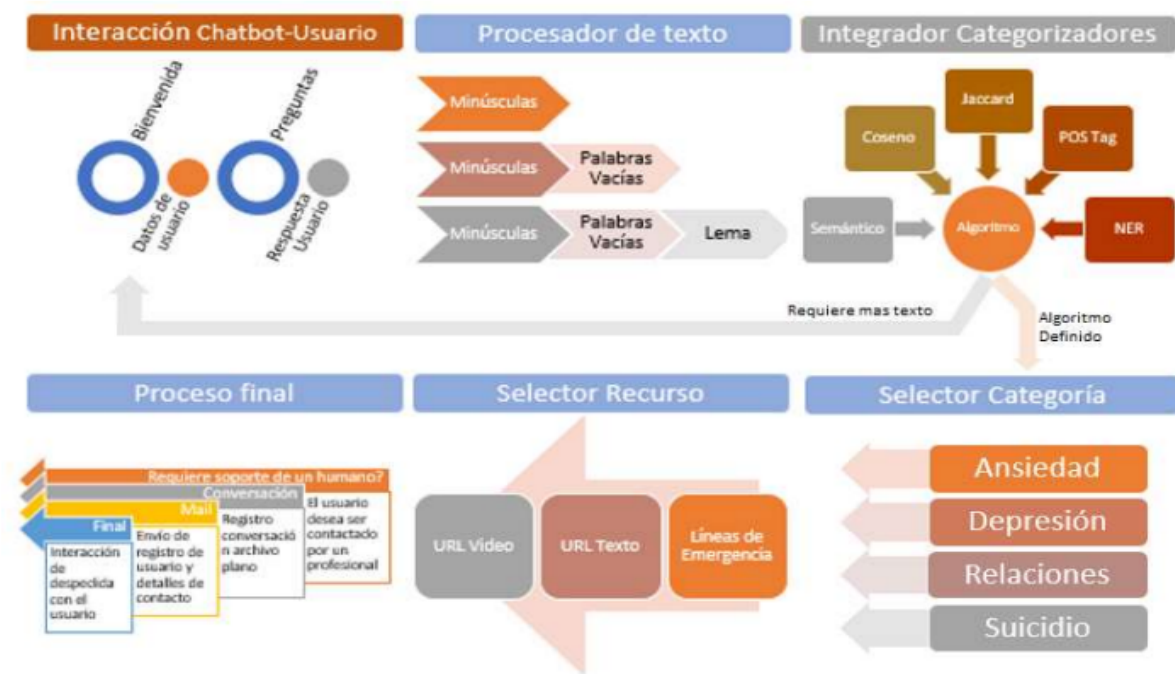
De ahí que el objetivo del proyecto sea generar un módulo de atención virtual tipo Chatbot que contenga las características principales para atender de manera preliminar a usuarios remotos que requieren acompañamiento emocional, llevando al usuario a resolver ciertas dudas iniciales, mientras un profesional está disponible para atenderlo directamente (Beltrán, 2022).

Con base en la respuesta del motivo de consulta dada por el usuario en NLP, esta respuesta es procesada basado en el análisis de procesamiento del lenguaje natural a través de categorizadores que, utilizando diferentes técnicas de similitud y teniendo en cuenta la categorización de insumos (textos) brindada por los expertos en cuatro áreas de interés, que son: ansiedad, depresión, relaciones y suicidio; CollectMind realiza la categorización de la frase del usuario en estas áreas de interés y posteriormente brinda un recurso relacionado. Adicionalmente, CollectMind le permitirá al usuario tener la opción de ser contactado por un profesional al siguiente día hábil de su consulta (Beltrán, 2022).

En la **Ilustración 5** se muestra el diagrama de solución final, donde se muestra la serie de procesos.

Ilustración 5

Diagrama de la solución final



Fuente: (Beltrán, 2022).

El diagrama de la solución anterior es realizado por medio de la metodología CRISP-DM, donde se pretende seguir con las fases que se muestran a continuación:

Comprensión del negocio: identificación de los procesos y canales mediante los cuales Mentec Colectivas ofrece sus servicios y realiza las atenciones a los individuos que buscan ser escuchados o recibir algún tipo de atención o recursos para atender su necesidad.

Entendimiento de los datos: recopilación de información, categorización y adecuación de la información suministrada por Mentec Colectivas. Estos insumos contemplan los

datos para Entendimiento de los datos entrenar el modelo, así como los diccionarios de datos que ayudarán en la definición del tipo de recurso para suministrarle al usuario.

Preparación de los datos: se realizará el proceso de alistamiento e inclusión de los datos, documentos y demás insumos, así como la exclusión de documentos que no sean relevantes Preparación de los datos para el proyecto o que no tengan la calidad necesaria para su uso.

Modelado: construcción de la herramienta junto con los componentes que ella demanda incluyendo los usados en el entrenamiento del Chatbot, que permita escoger un modelo adecuado con las necesidades del negocio. Una parte fundamental del proyecto es el uso de procesamiento de lenguaje natural mediante modelos pre entrenados.

Evaluación: se define una evaluación de frases que simulan la respuesta del cliente. con el fin de revisar la pertinencia en cada área de interés y poder medir las métricas de similitud en los categorizadores. Así mismo es importante la evaluación desarrollada por profesionales de Mentes Colectivas, quienes revisarán con frases previamente categorizadas en las cuatro áreas, validando si las respuestas corresponden a la categoría correcta.

Despliegue: documento donde se exponen los requisitos funcionales para poner en funcionamiento el chatbot.

El modelo recibe el texto ingresado por el usuario y los diccionarios antes realizados, lo que genera la categorización del grupo al que se puede asociar según las intenciones que se hagan visibles en el chat, Utilizando los insumos ofrecidos por los profesionales del portal de Mentes Colectivas, que fueron dispuestos en las pruebas evaluativas, CollectMind, permitió tener una precisión de 86.8% de la selección de la categoría, lo que ratifica que el enriquecimiento léxico de las c rporas ofrecen una mayor precisi n al modelo.

Con lo anterior, se evidencia la importancia de la herramienta de NLP por medio de diferentes campos de aplicación, para mejorar la experiencia al cliente, y poder así, obtener información más precisa para que se pongan sobre la mesa los diferentes planes de mejora con miras a aumentar la satisfacción del cliente.

3. Metodología

Para el desarrollo del proyecto se llevaron a cabo siete etapas, las cuales permiten parametrizar la secuencia de la construcción de la herramienta y su despliegue. Así mismo, se muestra el cronograma y el presupuesto utilizado a lo largo de todo el proyecto.

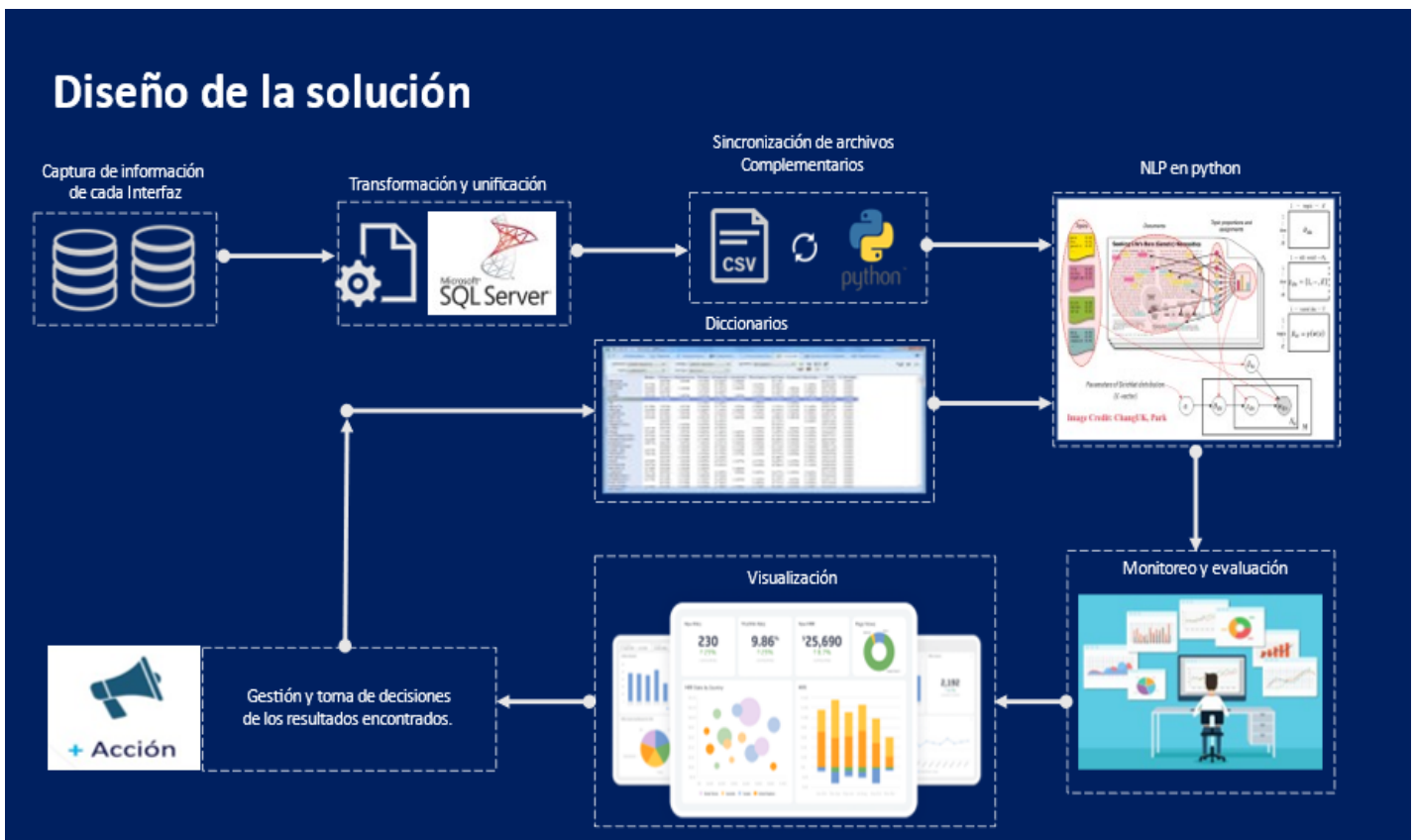
3.1 Etapas

1. **Conceptualización:** en esta etapa se inicia con el acercamiento del negocio, esto involucra cómo funcionan los procesos de PQR's y entendimiento de cómo se gestiona hasta el día de hoy, así como también, las necesidades y requerimientos existentes en la vicepresidencia, para ello se llevaron a cabo un par de reuniones donde se muestra la propuesta, inicialmente se da el aval para el proyecto y en las siguientes reuniones se puntualiza sobre las necesidades y requerimientos en la herramienta. Adicionalmente se comienza a examinar las bases de datos suministradas por la gerencia de Analytics, así como también, se inicia con el estudio de conceptos relacionados con NLP.
2. **Diseño de la solución:** se define de qué forma se aborda el problema según los requerimientos necesarios para las áreas involucradas, así mismo, como queda el proceso de gestión de las quejas servicio después de la creación de los tableros, cada cuanto se ejecuta, que persona queda encargada, que persona hace seguimiento a los planes de acción. Donde se tenga en cuenta, que los tópicos que mayor participación muestren deberán ser atacados y gestionados, logrando así

una disminución en estos casos. En la **Ilustración 6** se puede visualizar la estructura del diseño de la solución.

Ilustración 6

Diseño de la solución



3. Preparación de los datos: en la tercera etapa, se captura las fuentes de datos presentes en tres interfaces, EDATEL, TIGO y UNE. De ahí se realiza una unificación de la información y se realiza homologación, transformación y limpieza de las variables a trabajar a través de SQL SERVER. Estas son las bases involucradas en la unificación:

- [Reporting_Leader_Dev].[dbo].[TBL_PQRS_SIEBEL_COMPLETAS]
- [Reporting_Leader_Dev].[dbo].[TBL_PQRS_SIRA_COMPLETAS]
- [Reporting_Leader_Dev].[dbo].[TBL_PQRS_EDATEL_COMPLETAS]

Se hace la debida filtración para las quejas servicio, debido a que, en estas tablas se encuentran el total de PQR's, por otro lado, se tiene en cuenta los tres últimos tres meses, para ello se hace una selección móvil para que siempre muestre el periodo de tiempo indicado.

Finalmente se inicia con la construcción de los diccionarios de móvil y fijo por medio de archivos de Excel con formato xlsx, donde se plasman las palabras más representativas de los clientes que se comunican para generar una queja por servicio, asimismo a cada palabra se le asocia un tópico y criterio. Esta información es vinculada por medio de procesos de conexión con Python. A través de la **Tabla 2** se puede visualizar una pequeña muestra del diccionario para la unidad de negocio de FIJO.

Tabla 2

Muestra del diccionario de la BU de FIJO

Proceso	Palabra	Topico	Criterio
FIJO	wifi no llega hasta donde trabajo	falla en servicio	Wifi
FIJO	wifi no funciona dentro casa	falla en servicio	Wifi
FIJO	wifi mala	falla en servicio	Wifi
FIJO	wifi inestable	falla en servicio	Wifi
FIJO	WiFi funciona increíblemente mal	falla en servicio	Wifi
FIJO	wifi funciona increíblemente mal	falla en servicio	Wifi
FIJO	wifi dano	falla en servicio	Wifi
FIJO	wifi inestable	falla en servicio	Wifi
FIJO	vive cayendo	falla en servicio	Estabilidad General
FIJO	vive cayendo	falla en servicio	Estabilidad General

- 4. Construcción de herramienta de NLP:** la realización del desarrollo se hace en Python con módulos de procesos corridos en SQL Server. En la cual, para la realización del desarrollo se utilizan los siguientes paquetes vistos en la **Tabla 3**:

Tabla 3

Paquetes del desarrollo

	El módulo dateutil admite el análisis de fechas en cualquier formato de cadena, proporciona detalles
--	---

<p>python-dateutil</p>	<p>internos actualizados de la zona horaria mundial, ayuda a calcular los deltas relativos, así como también ayuda a calcular las fechas en función de reglas de recurrencia bastante flexibles. (jvatpoint, 2021).</p>
<p>openpyxl</p>	<p>Openpyxl es una biblioteca de Python que se utiliza para leer desde un archivo de Excel o escribir en un archivo de Excel. Los científicos de datos usan Openpyxl para el análisis de datos, la copia de datos, la extracción de datos, la elaboración de gráficos, las hojas de estilo, la adición de fórmulas, entre otros. (topcoder, 2021).</p>
<p>PyJWT</p>	<p>PyJWT es una biblioteca de Python que le permite codificar y decodificar JSON Web Tokens (JWT). JWT es un estándar abierto de la industria (RFC 7519) para representar reclamaciones de forma segura entre dos partes. (PyJWT, 2022).</p>
<p>Pyodbc</p>	<p>pyodbc es un módulo Python de código abierto que simplifica el acceso a las bases de datos ODBC. Implementa la especificación DB API 2.0. (pyodbc, 2022).</p>
<p>pandas</p>	<p>Pandas es un paquete de Python de código abierto que se usa más ampliamente para tareas de ciencia de datos/análisis de datos y aprendizaje automático. Está construido sobre otro paquete llamado Numpy, que brinda soporte para arreglos multidimensionales. Como uno de los paquetes de gestión de datos más populares, Pandas funciona bien con muchos otros módulos de ciencia de</p>

	datos dentro del ecosistema de Python. (ActiveState, 2022).
--	--

El desarrollo inicia con la homologación, limpieza y transformación de las entradas de texto, tanto para el campo problema, como para el campo solución. Donde se normaliza todas las palabras, quedando todas en minúscula, sin ninguna tilde o virgulilla. Por otro lado, se elimina cualquier carácter especial, como lo puede ser los signos de puntuación. Adicionalmente, en la **Tabla 4** se observa aquellas palabras eliminadas, como: conectores, conjunciones y palabras vacías que no aportan en la extracción de información de la herramienta.

Tabla 4

Palabras eliminadas del texto

yo	Tu	Tus
con	que	qu
qe	uno	un
una	el	ella
ello	ellos	ellas
esto	eso	de
por	porque	porq
en	me	mi
y	su	sus
se	ha	o
ah	son	la
al	es	los
de	las	a
lo	para	mis
mio	acerca	precisa
exacta	hay	tan

q	si	te
he	eh	e
muy	quisiera	han

En la **Tabla 5** se puede visualizar el antes y el después de la realización de la limpieza de la entrada de texto, con relación al campo problema del cliente, también que la misma limpieza se hace para la entrada de texto donde se encuentra la solución de la queja servicio. Cabe mencionar que, también se realiza el proceso de limpieza a los diccionarios ingresados al desarrollo, esto para evitar algún tipo de inconsistencia en los pasos siguientes.

Tabla 5

Comparación de la limpieza de texto

Problema antes de la limpieza	Problema después de la limpieza
ID Llamada: a24232a3-c5b5-4e01-828a-760178b9b11c Asesor que realiza la solicitud: kmontoyh Nombre de quién llama: daliz elena de la oz ortega Número de CC de quién llama: xxxxxx Autoriza que la notificación de su PQR se realice por correo electrónico?: SI Correo para envío de respuesta: xxxxxxx Autoriza que la respuesta le sea enviada por el servicio de mensajería Física?: NO Número de Contacto: xxxxxxx Móvil del reclamo o inconveniente: xxxxxx Por qué no se brindó SPC?: usuario solicita PQR Motivo de Petición: Otro tipo de Petición Requerimiento: usuaria se comunica indicando que una línea que tuvo con Tigo xxxxxx por falta de uso se la reciclaron, como tal ella perdió el número, y la persona que ahora tiene el número esta realizando estafas por medio de wassap porque abrió esta aplicación con el número xxxxxx le empezó a escribir a los contactos que ella tenía, solicitándole dinero haciéndose pasar por ella, usuaria solicita que se de la suspensión de esta línea porque la están perjudicando. y se está haciendo actos ilícitos	id llamada a24232a3c5b54e01828a760178b9b11c asesor realiza solicitud kmontoyh nombre quien llama daliz elena oz ortega numero cc quien llama xxxxxx autoriza notificacion pqr realice correo electronico si correo envio respuesta xxxxxxx autoriza respuesta le sea enviada servicio mensajeria fisica no numero contacto 3043343502 movil reclamo o inconveniente xxxxxx por no brindo spc usuario solicita pqr motivo peticion otro tipo peticion requerimiento usuaria comunica indicando linea tuvo tigo xxxxxx falta uso reciclaron como tal perdio numero persona ahora tiene numero esta realizando estafas medio wassap abrio esta aplicacion numero xxxxxx le empezò escribir contactos tenia solicitandoles dinero haciendose pasar usuaria solicita suspension esta linea estan perjudicando esta haciendo actos ilicitos

Luego de estar limpias las entradas de texto y diccionarios, se pasa al proceso de tipificación. Se extraen las palabras encontradas en las quejas de los clientes y se ubican en una lista, luego a través de ciclos, por cada comentario se busca todas las palabras que se puedan contener de acuerdo con lo que está en los diccionarios. Al final, se obtiene una lista con todas las palabras encontradas, dado el caso que no encuentre palabras en un input de texto, la lista queda vacía.

Por último, se consolida esta información, la cual se ubica en una tabla de SQL SERVER, junto con una serie de variables provenientes de la unificación de fuentes de información y la adición de otras variables necesarias para la visualización, en la **Tabla 6** se puede observar los campos existentes, junto con su descripción.

Tabla 6

Campos de la tabla consolidada

Campos	Descripción
RADICADO	Código que se le asigna a cada queja.
IDENTIFICACION	Cédula de la persona que presenta la queja.
FECHA_INGRESO	Fecha en la que ingresa la queja.
FECHA_CIERRE	Fecha en la que se le da respuesta y cierre a la queja.
PERIODO	Año y mes del ingreso de la queja.
EDAD_PQR	Número de días en el que se le da respuesta y cierra de la queja.
TIPO	Tipo de PQR, en este caso todas corresponden a queja servicio.
FAVORABILIDAD	Indica si la respuesta dada al cerrarse la queja es: favorable, desfavorable o parcialmente favorable.
MEDIO_INGRESO	Canal por el que ingresa la queja: redes sociales, medios escritos, teléfono, tiendas, visita y página web.
ESTADO	Indica si la queja aún está abierta o si esta se encuentra cerrada.
BU	La unidad de negocio indica si la queja corresponde a clientes con servicios fijos o móviles.
UE	La unidad estratégica indica si la queja corresponde a cliente natural o cliente tipo empresa.
FILIAL	Indica la filial que cobija al cliente: tigo, une o edatel.
CICLO	Indica el ciclo de facturación en el que se encuentra el cliente.
DEPARTAMENTO	Departamento en el que se encuentra el cliente.
MUNICIPIO	Municipio en el que se encuentra el cliente.
REGIONAL	Indica la región en la que se encuentra el cliente: andina, Bogotá, costa y sur.
FUENTE_DATOS	Indica la fuente de datos a la que corresponde el cliente: SIRA= TIGO, EDATEL = EDATEL o SIEBEL = UNE.
Id_Survey	Representa el mismo valor que el radicado.

Customer id	Representa la cedula del cliente.
Type_survey	Indica si la tipificación corresponde a una entrada de texto de problema o solución, así como también su BU.
Comment	Indica el ingreso de texto sin limpieza.
Approved Comment	Indica el ingreso de texto con limpieza.
COMENTARIO_VALIDO	Indica si el comentario es válido o no; si es válido marca un '1', sino marca '0'.
Topico	Tópico asociado a la queja.
Criterio	Criterio asociado a la queja.
Polarity	Indica la polaridad de la entrada de texto, si es: positiva, negativa y neutra.
Word	Indica la palabra extraída del campo de texto.
CORTE	Indica si la fecha de ingreso pertenece a fecha corte o a fecha complemento.

5. Monitoreo y evaluación de tipificación: se realiza una evaluación a la tipificación que está realizando la herramienta por medio de la ecuación que se encuentra en la **Ilustración 7**, donde la variable que está en el numerador ‘quejas servicio con tópico’, hace referencia a aquellos reclamos que fueron válidos y donde el tópico es diferente de nulo y la variable del numerador ‘quejas servicio válidas’, hace referencia al conteo del campo de comentarios válidos. Esto se calcula para conocer qué tan acertados son los resultados que arroja, así mismo, periódicamente se realiza un monitoreo de las ejecuciones para validar que la herramienta siga arrojando los resultados esperados a través del tiempo. Adicionalmente, en los primeros meses los diccionarios se estarán alimentando hasta tener un nivel significativo de palabras con tópicos y criterios, luego se deberán actualizar y nutrir cada seis meses, sustentando en que, entre más completos están los diccionarios, mejores resultados se obtendrán de la herramienta.

Ilustración 7

Porcentaje de tipificación

$$PORCENTAJE DE TIPIFICACIÓN = \frac{\text{quejas servicio con tópico}}{\text{quejas servicio válidas}} * 100$$

6. Visualización y análisis: en la sexta etapa se visualiza la información por medio de un tablero por el aplicativo Tableau, el cual se actualiza automáticamente todos

los días a las 7:00 AM. Así pues, este permite plasmar las diferentes tipificaciones de las quejas, generando focos para la toma de decisiones, para que luego, se le pueda dar la debida gestión en dicha área. En las siguientes secciones se presenta los resultados de las visualizaciones y su respectivo análisis.

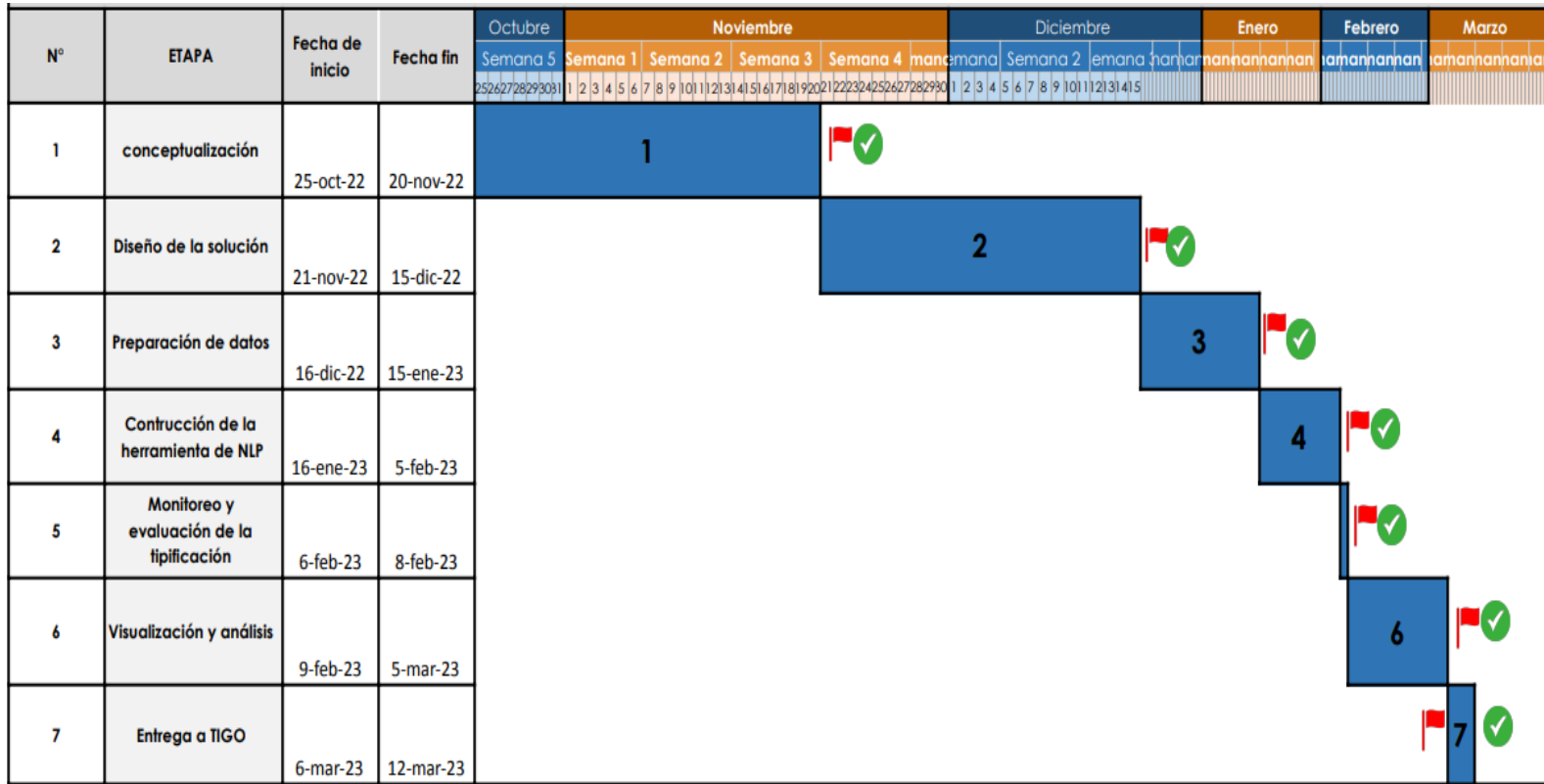
7. **Entrega a TIGO:** en la última etapa se presenta los resultados y se realiza las entregas a la gerencia con los principales insights encontrados en la herramienta de NLP. Adicionalmente, se reciben todas las retroalimentaciones por las diferentes áreas que hacen uso de la herramienta, para poder que se tenga el mayor provecho por el mayor número de personas posibles.

3.2 Cronograma de actividades

En la **Ilustración 8** se puede visualizar cada una de las etapas mencionadas anteriormente, con su respectiva fecha de inicio y fecha de finalización, así como también los hitos presentes en el proyecto y si fue cumplida la etapa en el proceso estipulado. Para una mejor visualización del cronograma ingrese al siguiente enlace: [cronograma](#).

Ilustración 8

Cronograma del proyecto



3.3 Presupuesto

Los recursos necesarios para el desarrollo del proyecto se encuentran descritos en la **Tabla 7**, donde los montos con mayor valor se encuentran en la mano de obra y el computador donde se realiza todo el desarrollo de la herramienta.

Tabla 7

Presupuesto del proyecto

Presupuesto del proyecto	
Descripción	Valor total
Computador	\$ 4.630.000
Mano de obra (6 meses)	\$ 8.880.000
Licencia de microsoft (6 mes)	\$ 399.396
Licencia de Tableau (1 mes)	\$ 345.636
TOTAL	\$ 14.255.032

4 Resultados

El despliegue final del proyecto se hace a través de un tablero de visualización en Tableau donde a través de tres vistas se plasma la información encontrada en la herramienta de NLP. La primera vista llamada ‘VISTA GENERAL’ muestra información general de la tipificación asociado a volúmenes de contactos, porcentaje de tipificación, volumen y porcentaje de quejas servicios por canal, cantidades de tópicos y criterios y nube de palabras. Esta vista, viene acompañada de cinco filtros vistos en la **Tabla 8** junto a su apertura.

Tabla 8

Filtros de la ‘VISTA GENERAL’

CAMPO FILTRO	APERTURA
PERIODO	202301, 202302, 202303
BU	FIJO, MOVIL
UE	B2B, B2C
PROBLEMA/SOLUCIÓN	PROBLEMA FIJO, PROBLEMA MOVIL, SOLUCIÓN FIJO, SOLUCIÓN MÓVIL
CORTE	CORTE, COMPLEMENTO

Por efectos prácticos se presentan en la **Tabla 9** los resultados con los siguientes filtros, debido a que son los más usados en el tablero.

Tabla 9

Filtros seleccionados

CAMPO FILTRO	APERTURA
PERIODO	202301, 202302, 202303
BU	FIJO
UE	B2C
PROBLEMA/SOLUCIÓN	PROBLEMA FIJO
CORTE	CORTE

A continuación, se muestran con detalle cada uno de los esquemas realizados. En la **Ilustración 9** se muestra un gráfico de barras que permite conocer el volumen de reclamaciones por queja servicio para los últimos tres meses, así mismo, se muestra el número de entradas de texto que fueron válidas con tópico, válidas sin tópico y no válidas; las no validas hace referencia

a esas quejas servicio que son nulas o que después de hacer la limpieza no se pudo extraer ninguna palabra para tipificar. Para el primer trimestre del año se tiene niveles de reclamación normales con relación a valores de meses anteriores, aunque para el mes de febrero se presenta un alza significativa de hasta 24.000 reclamaciones. Adicionalmente, es visible que la cantidad de entradas de texto válidas con tópico representa la gran mayoría del espectro de datos para cada vez.

Ilustración 9

Reclamaciones de queja servicio por mes

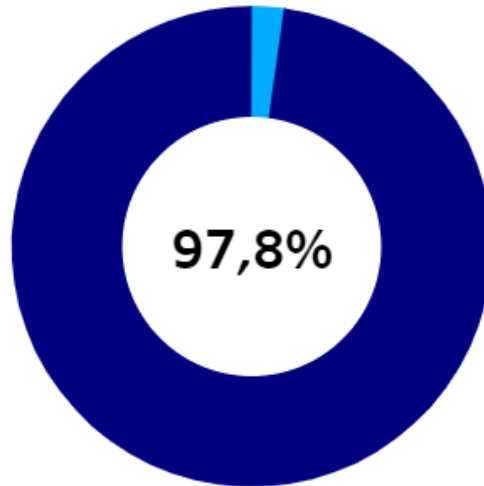


En la **Ilustración 10** se puede observar un gráfico circular de anillo, el cual representa el porcentaje de tipificación proveniente de la fórmula de la **Ilustración 7**, donde el color azul oscuro representa la cantidad con tipificación y el azul celeste la cantidad sin tipificación. Se considera que dentro de TIGO es aceptable la tipificación si supera el 80%, por lo que en este caso la actuación de la herramienta con relación a los diccionarios tiene un muy buen comportamiento.

Ilustración 10

Porcentaje de tipificación de la herramienta

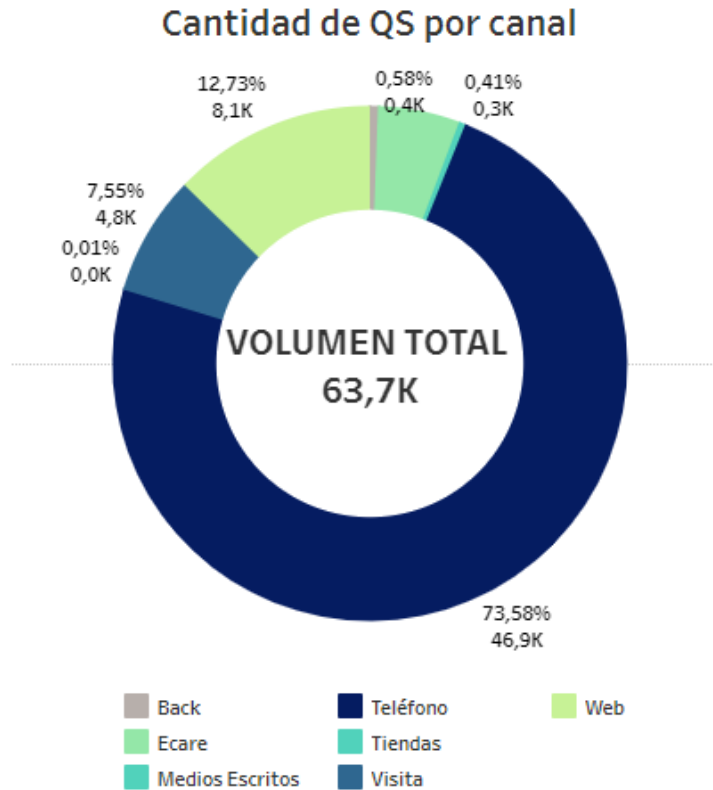
Porcentaje de tipificación



En la **Ilustración 11** se representa el volumen total de quejas servicios que ingresan a los diferentes canales existentes en la empresa, donde dicho valor se ubica en el centro del círculo de anillo, adicionalmente se muestra la cantidad en miles y el porcentaje de participación de cada uno de los canales de 'Back', 'Ecare', 'Medios Escritos', 'Teléfono', 'Tiendas', 'Visita' y 'web'. Para los tres meses se tiene un aproximado de 64.000 reclamaciones en todos los canales de atención para el primer trimestre del año. Por otro lado, el canal de vía telefónica representa el del 73,58% del total de contactos por PQR's, seguido de la página web, la cual representa el 12,73%, siendo estos, los que generan el Pareto de contactos por canal.

Ilustración 11

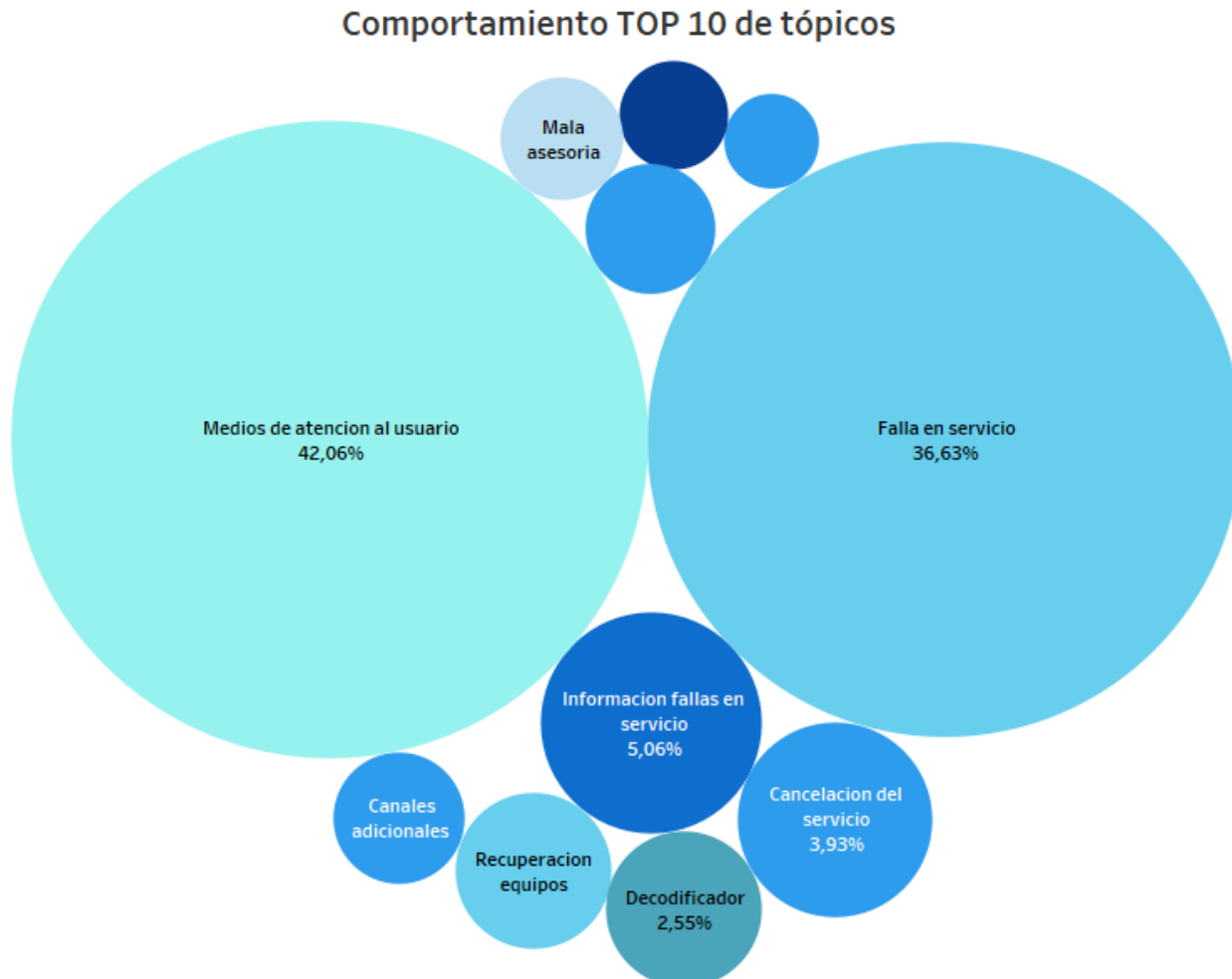
Cantidad de quejas servicios por canal



En la **Ilustración 12** se puede observar el top 10 de tópicos mostrando el porcentaje de participación de cada uno. De ahí que, medios de atención al usuario sea el tópico con más reclamaciones para el primer trimestre del año con un 42,06% del total, seguido por el tópico de falla en servicio con un 36,63%, donde juntos representan casi el 80% de la participación de tópicos. Cabe mencionar que, si bien por efectos de tamaño de los círculos no se muestra la información completa, si se pone el cursor del ratón se puede visualizar el nombre del tópico y su porcentaje.

Ilustración 12

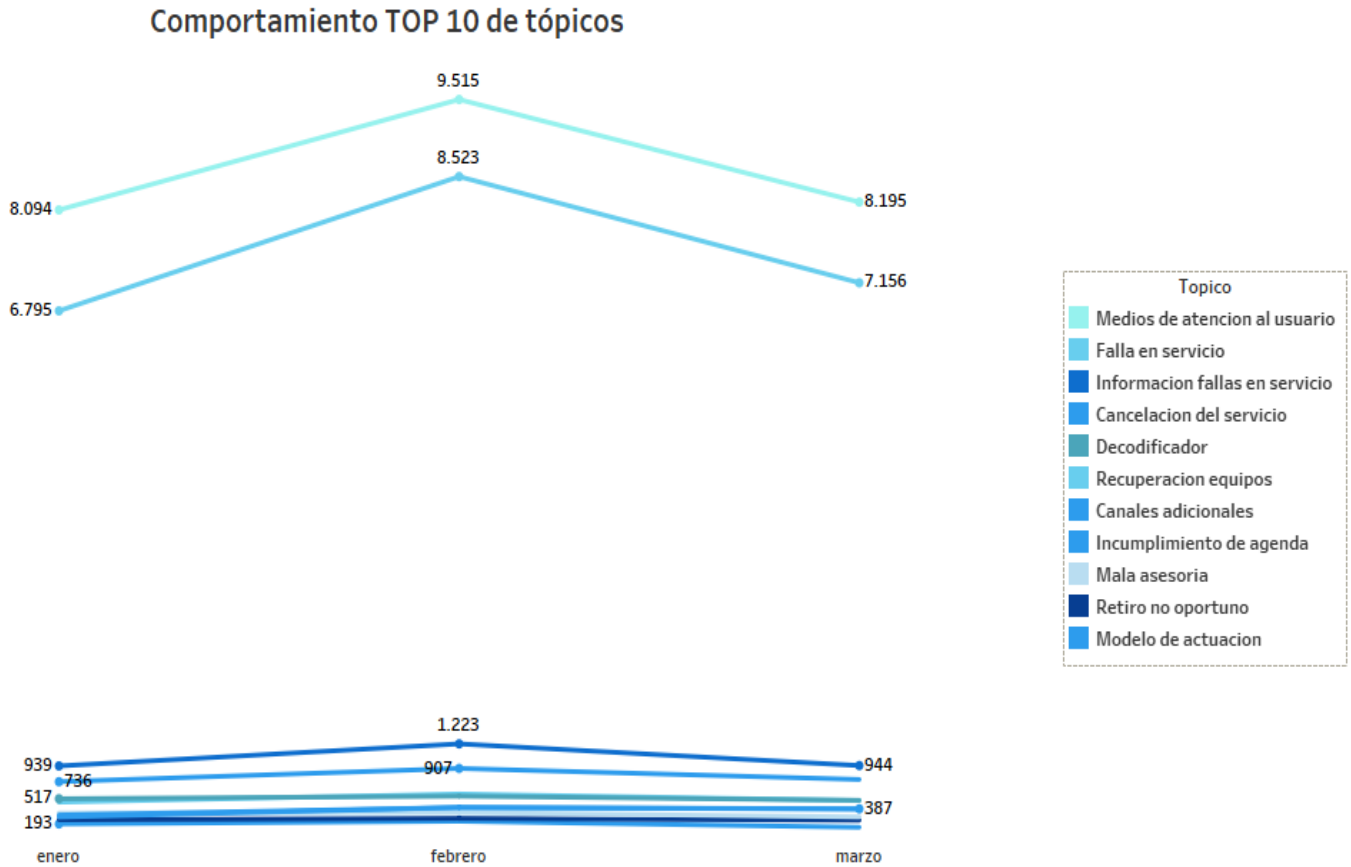
Comportamiento del TOP 10 de tópicos agrupados



En la **Ilustración 13** se puede visualizar el comportamiento del top 10 de tópicos a través del tiempo, mostrando el volumen de cada uno. Los principales hallazgos se encuentran en la reducción de los tópicos entre febrero y marzo, de resto se presenta una estabilidad para los demás tópicos.

Ilustración 13

Comportamiento del top 10 de tópicos por mes



En la **Tabla 10** se visualiza el gráfico correspondiente a los volúmenes de todos los tópicos y criterios durante los tres meses más recientes, donde a su vez, según la intensidad del color representa que tan grande el valor asociado a la categorización. Dónde ‘medios de atención al usuario’ y ‘falla en servicio’ presentan los volúmenes más altos durante el trimestre mostrado. Adicionalmente, en el gráfico se puede desplegar los criterios asociados a cada tópico mostrando también la cantidad de volúmenes.

Tabla 10

Volumen de tópicos y criterios

Volumen de tópicos y criterios

Topico	202301	202302	202303
Medios de atencion al usuario	8.094	9.515	8.195
Falla en servicio	6.795	8.523	7.156
Informacion fallas en servicio	939	1.223	944
Cancelacion del servicio	736	907	765
Decodificador	517	550	497
Recuperacion equipos	473	578	494
Canales adicionales	307	404	384
Incumplimiento de agenda	281	405	387
Mala asesoria	334	339	283
Retiro no oportuno	247	264	241
Hurto de cable	174	235	167
Modelo de actuacion	193	227	150
Suplantacion	12	34	43
Total general	19.102	23.204	19.706

En la **Tabla 11** se visualiza el porcentaje de variación de cada uno de los tópicos y criterios existentes, el grafico busca mostrar esos puntos de inflexión donde hay aumentos pronunciados en las diferentes categorizaciones, para ello en los casos donde hay incremento se representa con un triángulo rojo y si disminuye se muestra con un triángulo color verde. Los hallazgos más relevantes se encuentran en el mes de febrero donde hubo incremento en todos los tópicos, donde, además, el incremento más alto fue para ‘Suplantacion’ con 183,3%, para el mes de marzo este tópico fue el único que nuevamente tuvo incremento, siendo de 26,5%.

Tabla 11

Porcentaje de variación de los tópicos y criterios.

porcentaje de variación

Topico		202301	202302	202303
Medios de atencion al ..	▷		▲ 17,6%	▼ -13,9%
Falla en servicio	▷		▲ 25,4%	▼ -16,0%
Informacion fallas en s..	▷		▲ 30,2%	▼ -22,8%
Cancelacion del servicio	▷		▲ 23,2%	▼ -15,7%
Decodificador	▷		▲ 6,4%	▼ -9,6%
Recuperacion equipos	▷		▲ 22,2%	▼ -14,5%
Canales adicionales	▷		▲ 31,6%	▼ -5,0%
Incumplimiento de ag..	▷		▲ 44,1%	▼ -4,4%
Mala asesoria	▷		▲ 1,5%	▼ -16,5%
Retiro no oportuno	▷		▲ 6,9%	▼ -8,7%
Hurto de cable	▷		▲ 35,1%	▼ -28,9%
Modelo de actuacion	▷		▲ 17,6%	▼ -33,9%
Suplantacion	▷		▲ 183,3%	▲ 26,5%

En la **Ilustración 14** se muestra una nube de palabras por frecuencia, donde según los filtros y tópicos seleccionados de alguna de las gráficas antes expuestas, se muestran las palabras más representativas de cada tipificación.

Ilustración 14

Nube de palabras

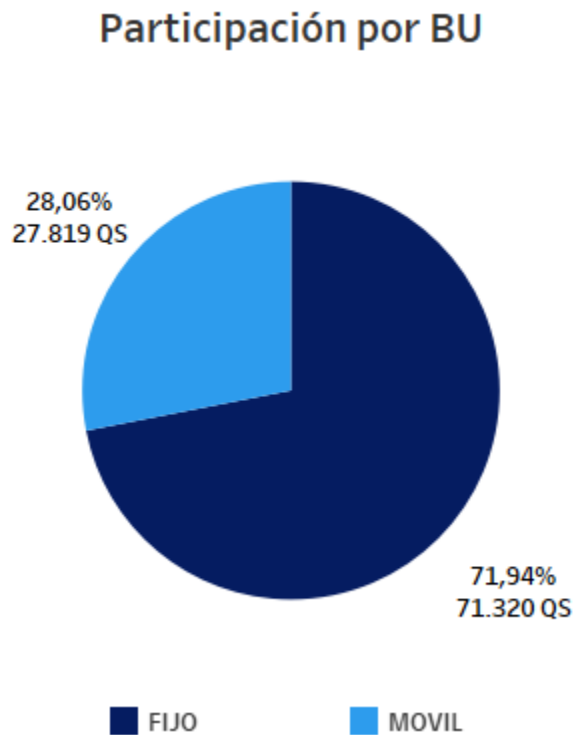
Nube de palabras



La segunda vista del tablero muestra cuatro gráficos representando con mayor detalle los valores de la información con relación a los campos de: BU, UE, canal y ciclo de facturación. Los filtros utilizados en esta vista son los mismos utilizados en la ‘vista general’, más el filtro de tópicos los cuales varían de acuerdo con su unidad de negocio. Con relación con el primer gráfico ubicado en la **Ilustración 15** se puede observar la participación por unidad de negocio, donde el azul oscuro representa fijo y el azul celeste representa móvil, la gráfica muestra el porcentaje y el volumen de las cantidades. Donde la mayoría de entradas de BU corresponde a fijo con 71,94% y el otro 28,06% corresponde a móvil.

Ilustración 15

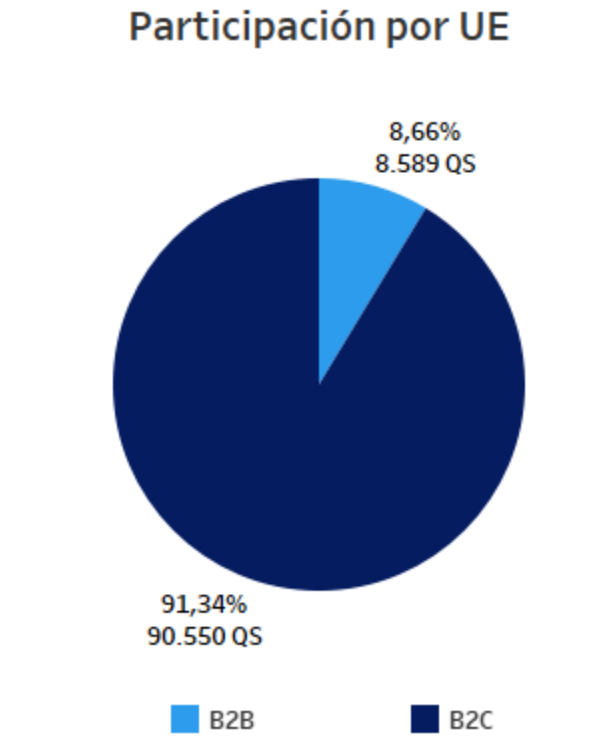
Participación por BU



En la **Ilustración 16** se puede visualizar el comportamiento de la unidad estratégica, mostrando el porcentaje y volumen de participación de las cantidades de clientes B2C y B2B. Donde gran parte de los clientes son B2C, es decir clientes naturales, siendo el 91,34% del total y el restante corresponde a clientes con razón social, es decir pertenecientes a empresa, siendo el 8,66%.

Ilustración 16

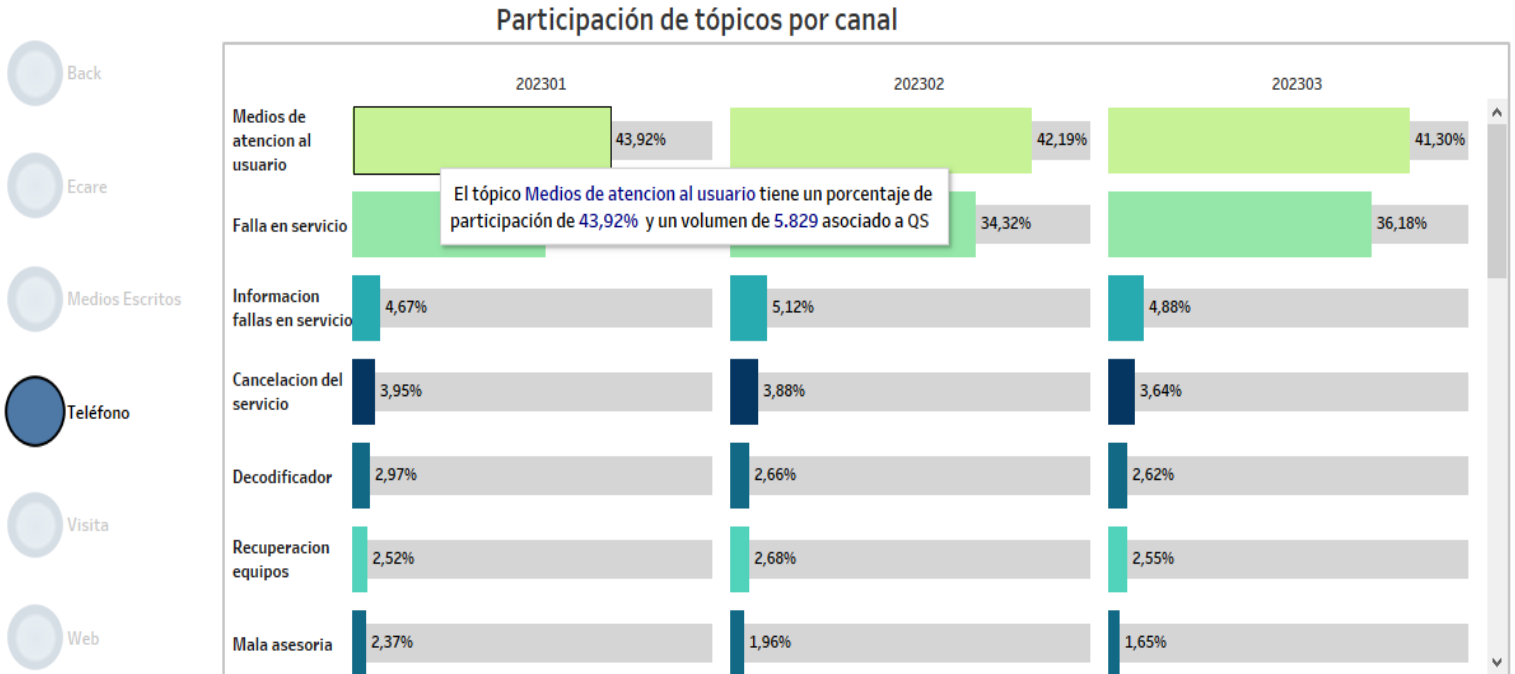
Participación por UE



En la **Ilustración 17** se puede visualizar los tópicos discriminados por periodo y en donde se puede ver por diferentes canales de ingreso de la queja servicio, filtrados por los botones ubicados en la parte izquierda del gráfico. Para el caso específico del canal vía teléfono, ‘Medios de atención al usuario’ y ‘Falla en servicio’ sigue siendo en los tres meses los tópicos más representativos en volúmenes.

Ilustración 17

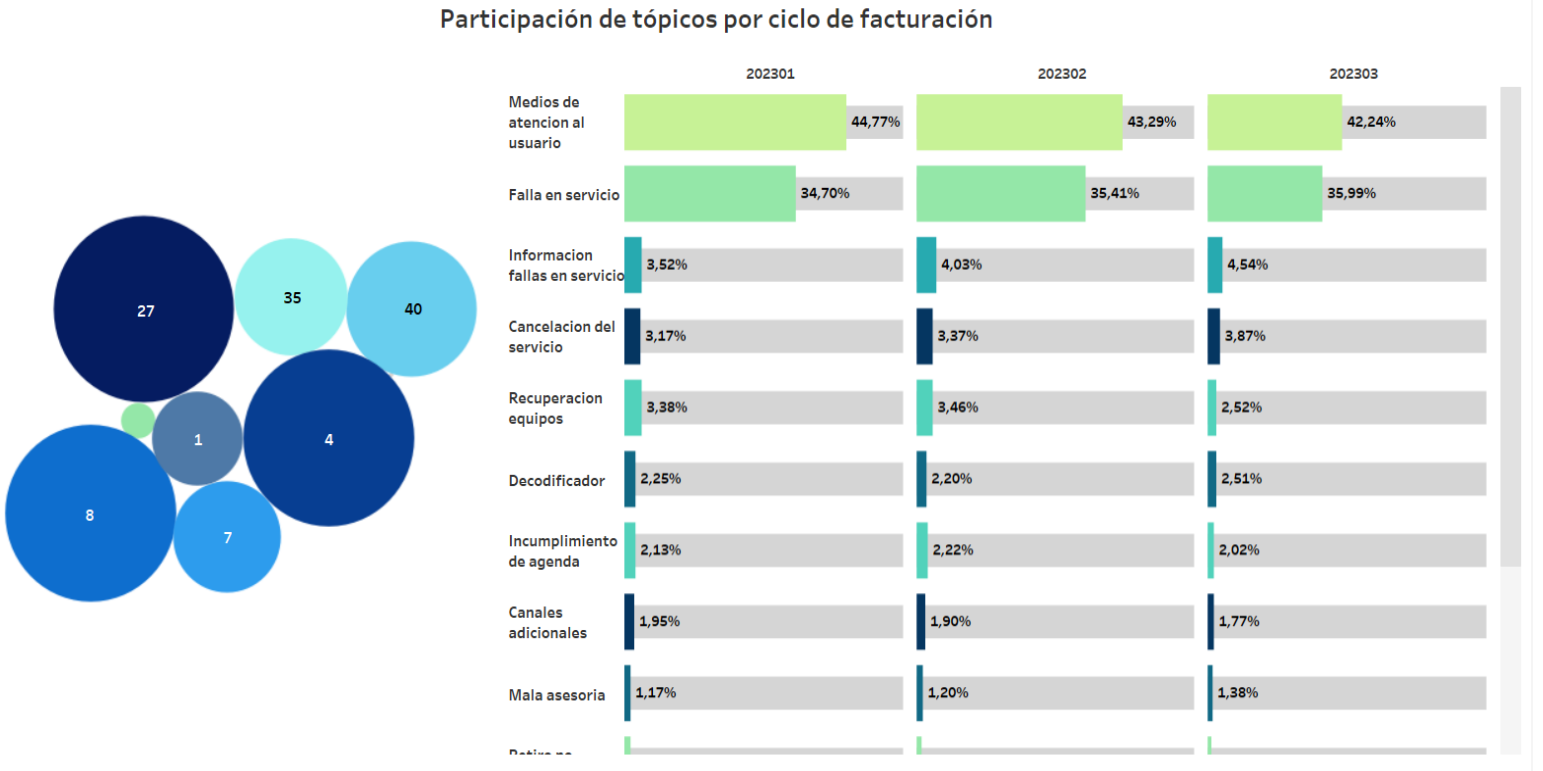
Participación de tópicos por canal



En la **Ilustración 18** se puede visualizar los tópicos discriminados por ciclo de facturación, siendo este el tramo de tiempo desde que se inicia la factura hasta que acaba, para el área de PQR's es de vital importancia indagar los tópicos más representativos en la agrupación de ciclos de facturación. Por lo que, es posible filtrar por medio de los círculos ubicados en la parte izquierda del gráfico representando los ciclos de facturación, cabe mencionar que dichos ciclos, entre más diámetro tienen mayor es el número de usuarios pertenecientes a este. No se logra percibir ningún comportamiento inusual respecto a la información que se ha encontrado en el resto de gráficos mostrados con anterioridad.

Ilustración 18

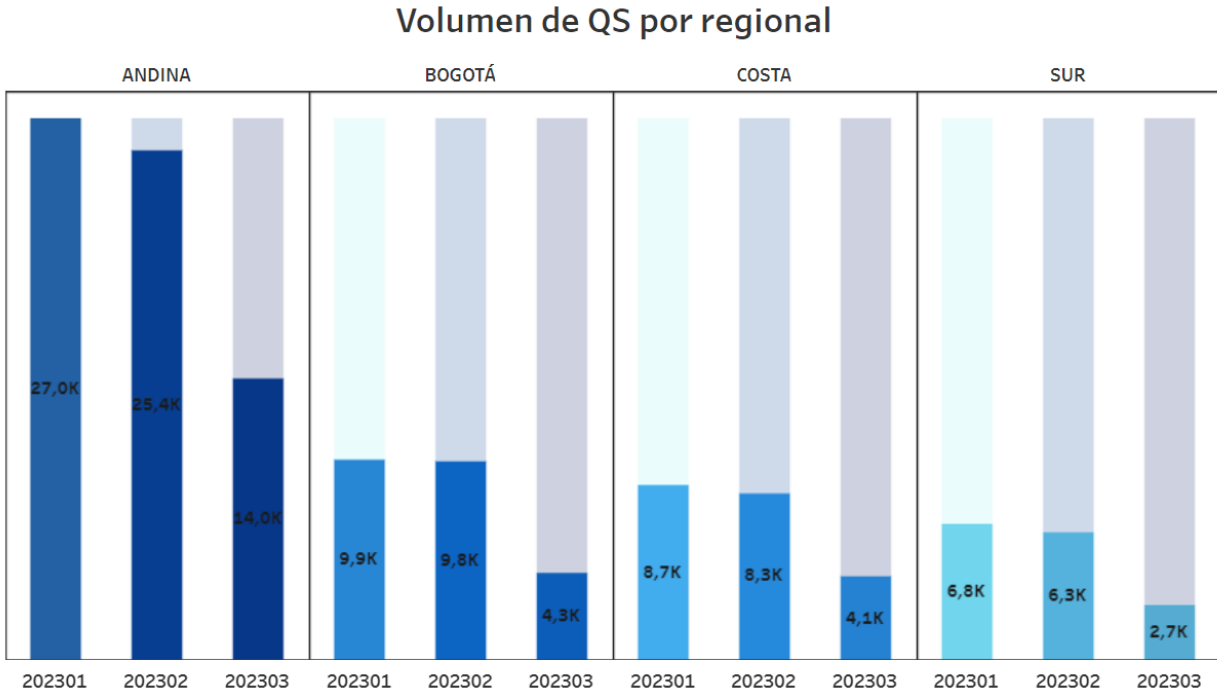
Participación de tópicos por ciclo de facturación



En la tercera vista llamada ‘vista regional’, se muestran aquellas gráficas relacionadas a la información de tópicos y criterios distribuida por zonas geográficas del país, donde se puede visualizar causas comunes en ciertos lugares donde tiene presencia TIGO. Por otro lado, se muestran los mismos filtros de la anterior vista; periodo, BU, UE, tópico, problema/solución y corte, adicionalmente se le suman los filtros de ‘Regional’ (andina, Bogotá, costa y sur) y un filtro llamado ‘Departamento’ donde se encuentra el distrito capital y los 32 departamentos. En la **Ilustración 19** se visualiza el volumen de quejas por servicio por las agrupaciones de regionales existentes, en donde para los tres meses la región andina presenta los números más altos de contactos seguido de Bogotá y costa que tienen un comportamiento muy similar y por último la región sur del país. Por otro lado, es importante mencionar que para el mes de marzo se generó una reducción de quejas por servicio en todas las regiones.

Ilustración 19

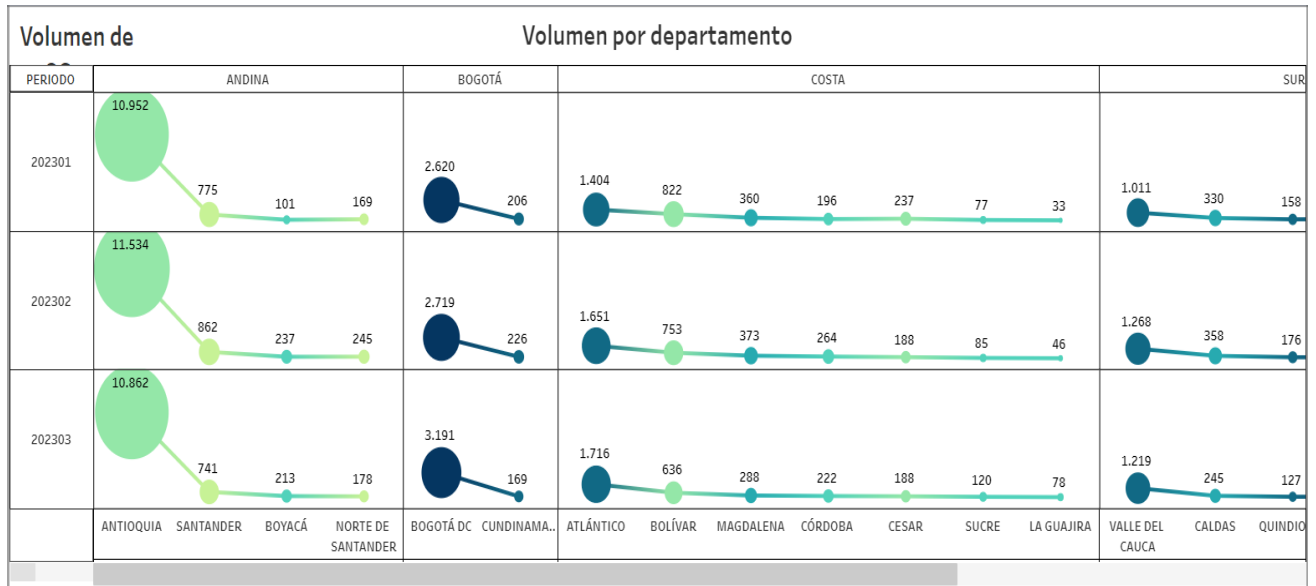
Volumen de quejas por servicio por regional



En la **Ilustración 20** se puede apreciar el volumen de quejas por servicio por departamento para el primer trimestre del año, donde se encuentra que gran parte de los reclamos son generados desde el departamento de Antioquia, seguido muy por atrás por el distrito capital Bogotá, después atlántico y de la región sur, el Valle del Cauca. Los departamentos anteriormente mencionados son los lugares donde se centra el Pareto de reclamaciones.

Ilustración 20

Volumen de quejas por servicio por departamento



En la **Tabla 12** se encuentra el comportamiento del top 10 por región, mostrando el volumen total y el porcentaje de participación de todos los tópicos por mes, además, según la intensidad del color este indica la cantidad de contactos de reclamaciones. Así pues, se evidencia que el ‘medios de atención al usuario’ y ‘Falla en servicio’ son los dos tópicos que generan los mayores números de quejas por servicio, y que, asimismo, la región andina posee el mayor volumen de estos tópicos.

Tabla 12

Comportamiento del top 10 por región

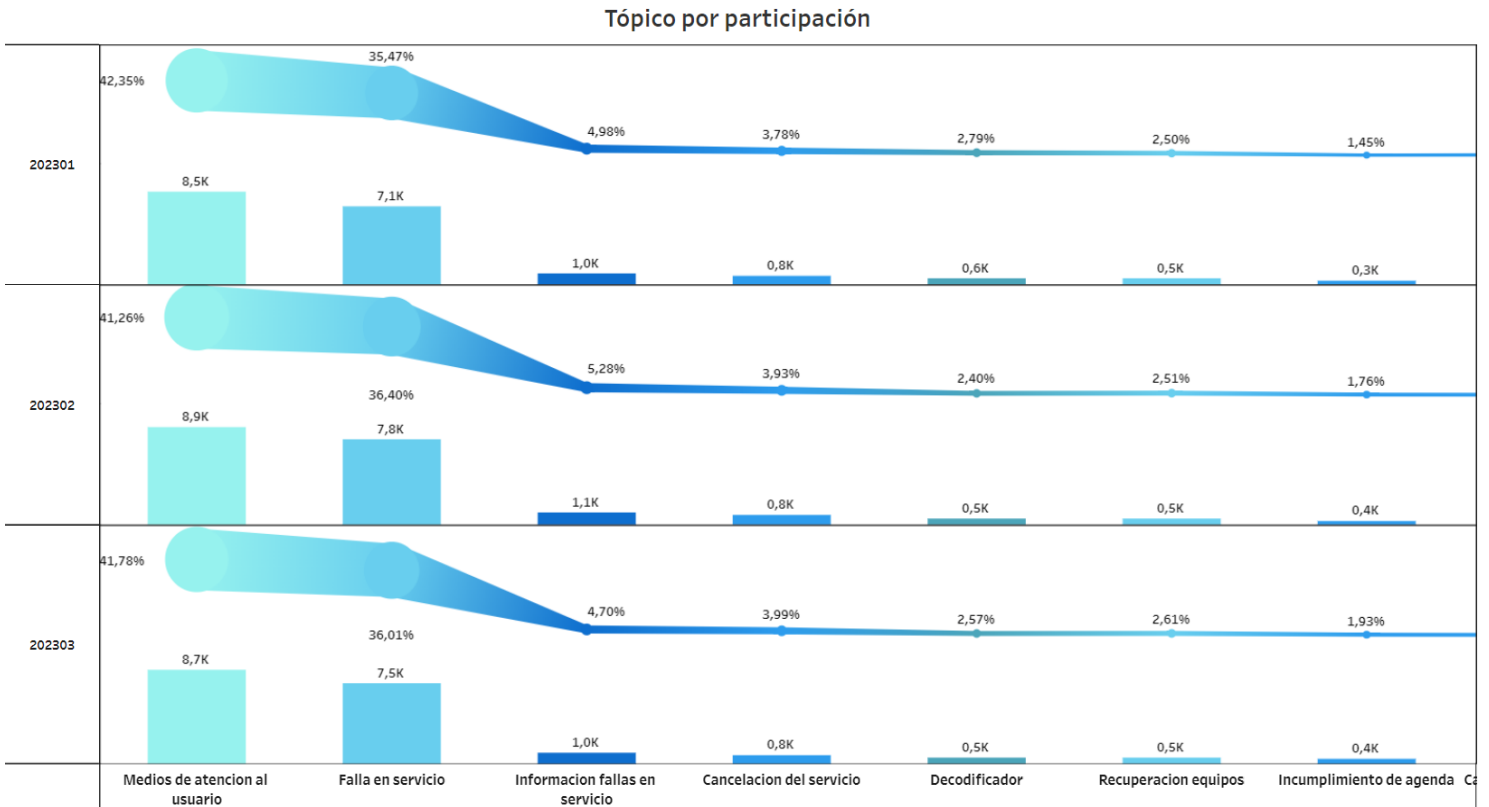
Comportamiento del Top 10 por región

Topico	ANDINA			BOGOTÁ			COSTA			SUR		
	enero	febrero	marzo	enero	febrero	marzo	enero	febrero	marzo	enero	febrero	marzo
Medios de atención al usuario	5,1K 43,8%	5,4K 42,9%	5,0K 42,6%	1,2K 43,8%	1,2K 40,5%	1,4K 43,2%	1,3K 42,7%	1,4K 42,3%	1,4K 42,8%	0,8K 41,0%	0,9K 40,4%	0,9K 43,2%
Falla en servicio	4,0K 34,5%	4,5K 35,9%	4,1K 35,1%	1,0K 34,7%	1,1K 38,1%	1,3K 38,3%	1,3K 41,2%	1,3K 40,9%	1,3K 39,9%	0,8K 41,5%	0,9K 38,7%	0,8K 40,2%
Información fallas en servicio	0,6K 5,1%	0,7K 5,2%	0,6K 5,1%	0,2K 5,9%	0,2K 6,0%	0,2K 5,0%	0,2K 4,9%	0,2K 5,4%	0,1K 4,6%	0,1K 4,4%	0,1K 6,0%	0,1K 3,4%
Cancelación del servicio	0,5K 3,9%	0,5K 4,2%	0,5K 4,5%	0,1K 4,1%	0,1K 4,2%	0,1K 3,2%	0,1K 3,7%	0,1K 3,2%	0,1K 4,0%	0,1K 3,5%	0,1K 3,9%	0,1K 3,1%
Decodificador	0,4K 3,5%	0,3K 2,6%	0,3K 2,9%	0,1K 2,4%	0,1K 2,7%	0,1K 2,3%	0,0K 1,6%	0,1K 1,6%	0,1K 2,0%	0,0K 1,9%	0,1K 2,2%	0,1K 2,6%
Recuperación equipos	0,4K 3,0%	0,4K 2,8%	0,3K 2,9%	0,1K 2,4%	0,1K 2,3%	0,1K 2,4%	0,0K 1,3%	0,1K 1,9%	0,1K 2,0%	0,0K 2,1%	0,1K 2,5%	0,1K 2,7%
Incumplimiento de agenda	0,2K 1,7%	0,2K 2,0%	0,3K 2,4%	0,0K 1,5%	0,1K 1,7%	0,0K 1,4%	0,0K 1,0%	0,0K 1,3%	0,0K 1,5%	0,0K 1,0%	0,0K 1,6%	0,0K 1,4%
Canales adicionales	0,2K 1,6%	0,2K 1,7%	0,2K 1,9%	0,1K 1,9%	0,0K 1,6%	0,1K 1,9%	0,0K 1,5%	0,0K 1,3%	0,0K 1,1%	0,0K 2,1%	0,1K 2,3%	0,0K 1,9%
Mala asesoría	0,2K 2,0%	0,2K 1,7%	0,2K 1,8%	0,1K 2,3%	0,0K 1,7%	0,0K 1,4%	0,0K 1,3%	0,0K 1,2%	0,0K 1,1%	0,0K 1,2%	0,0K 1,4%	0,0K 0,8%
Modelo de actuación	0,1K 0,9%	0,1K 1,0%	0,1K 0,8%	0,0K 1,2%	0,0K 1,2%	0,0K 0,9%	0,0K 0,6%	0,0K 0,9%	0,0K 0,9%	0,0K 1,3%	0,0K 1,0%	0,0K 0,8%

En la **Ilustración 21** se presentan todos los tópicos con el volumen y el porcentaje de participación para el primer trimestre del año, esta gráfica permite tener una visión general de tópicos para la vista regional. Con relación a los resultados del gráfico, no presentan variación con la información antes expuesta.

Ilustración 21

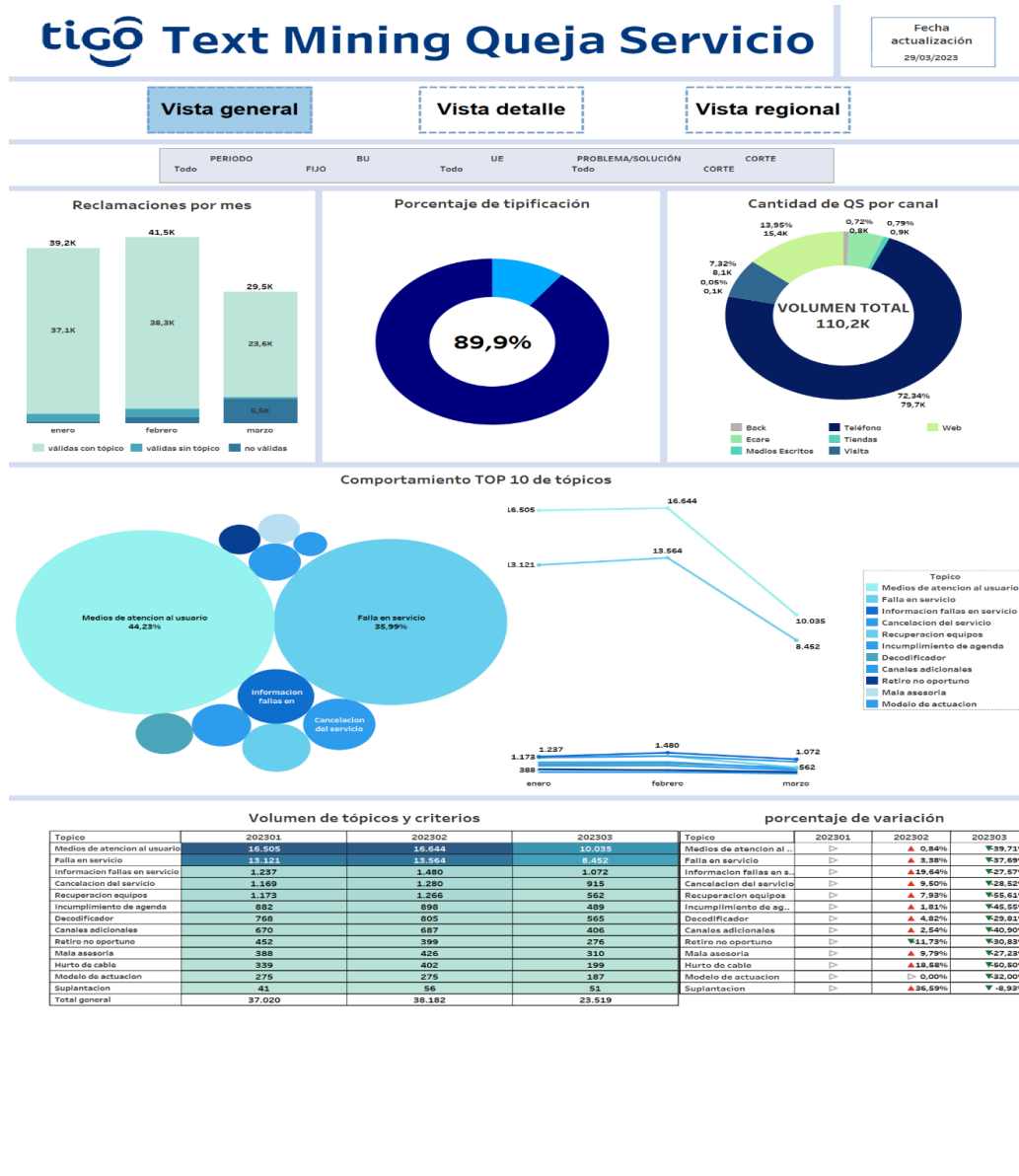
Tópicos por participación



Los gráficos anteriormente mostrados corresponden al tablero de visualización de la herramienta de NLP para las quejas servicio del espectro de PQR's, en la **Ilustración 22**, **Ilustración 23** e **Ilustración 24** se muestran las tres vistas del tablero en forma conjunta.

Ilustración 22

Vista general



Nube de palabras



Ilustración 23

Vista detalle

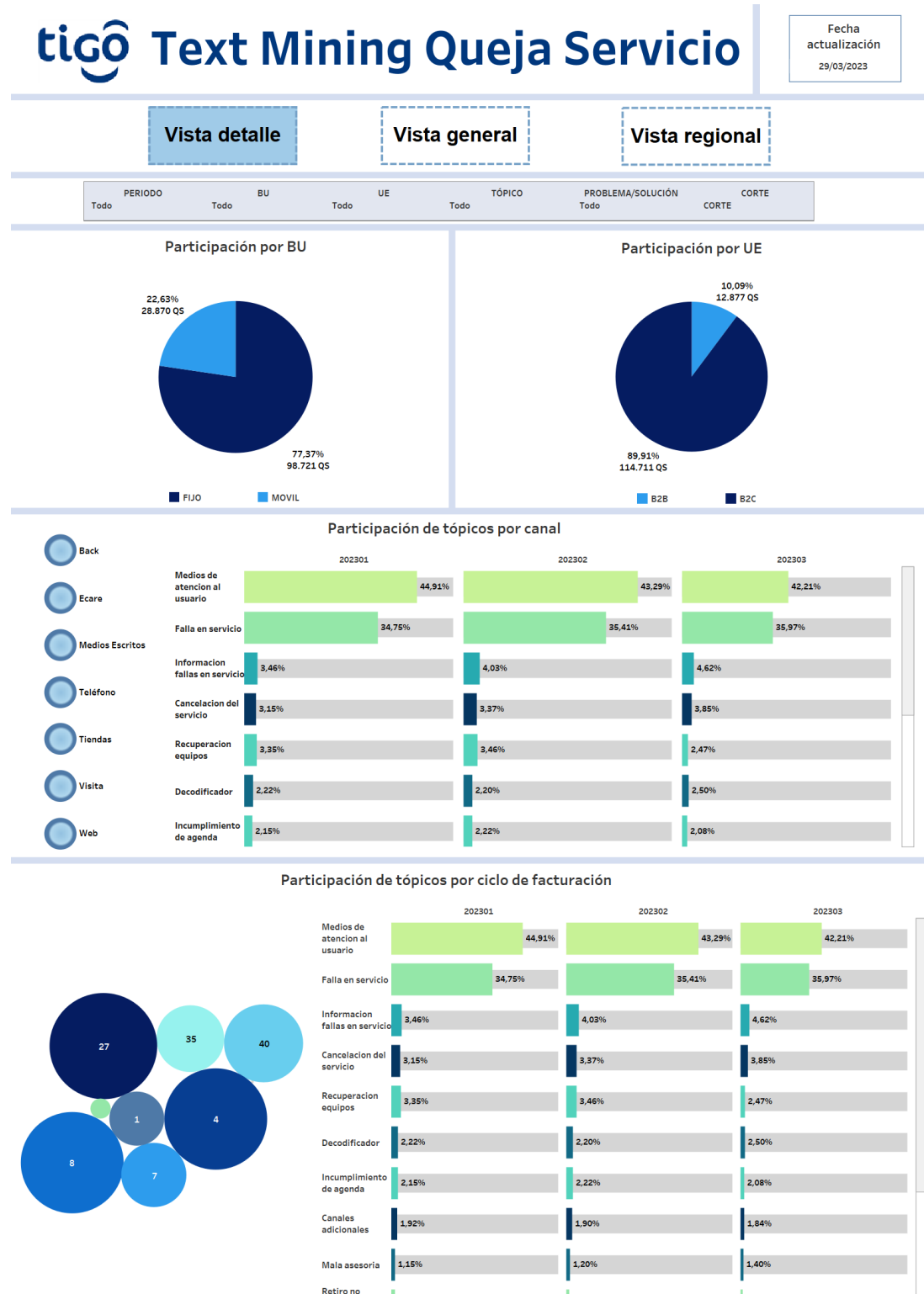


Ilustración 24

Vista regional

tigo Text Mining Queja Servicio

Fecha actualización
29/03/2023

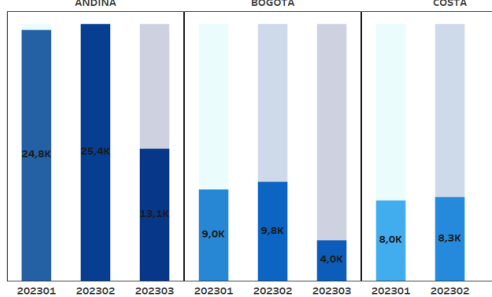
Vista regional

Vista general

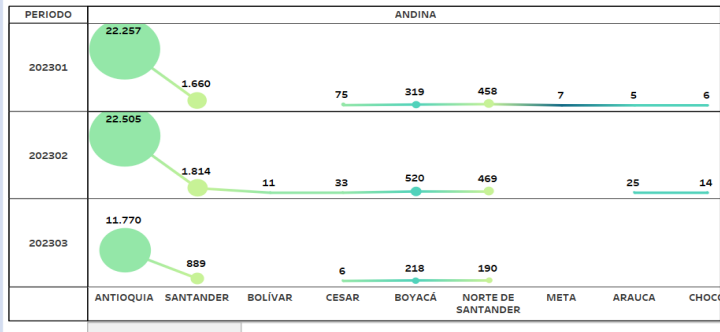
Vista detalle

PERIODO	BU	UE	TÓPICO	REGIONAL	DEPARTAMENTO	PROBLEMA/SOLUCIÓN	CORTE
Todo	Todo	Todo	Valores múltiples	Valores múltiples	Todo	Todo	CORTE

Volumen de QS por regional



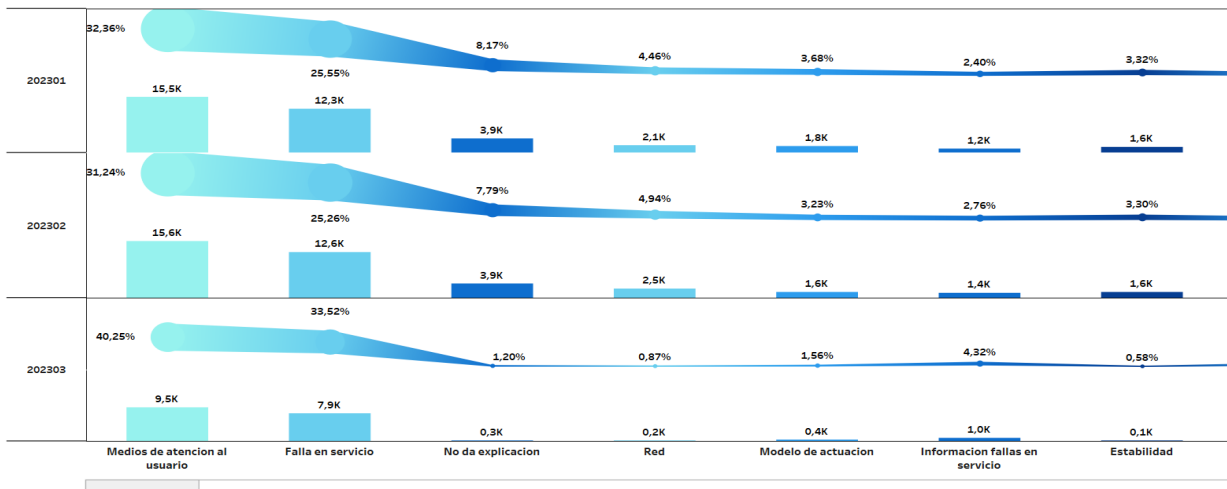
Volumen por departamento



Comportamiento del Top 10 por región

Topico	ANDINA			BOGOTÁ			COSTA			SUR		
	enero	febrero	marzo	enero	febrero	marzo	enero	febrero	marzo	enero	febrero	marzo
Medios de atención al usuario	9,6K 40,9%	9,5K 40,0%	5,4K 42,9%	2,1K 26,4%	2,0K 24,0%	1,6K 41,3%	2,3K 32,6%	2,4K 32,5%	1,5K 40,9%	1,5K 27,5%	1,7K 28,9%	1,0K 40,4%
Falla en servicio	7,1K 30,4%	7,3K 30,5%	4,3K 33,8%	1,5K 19,5%	1,7K 20,0%	1,3K 34,9%	2,1K 29,0%	2,1K 28,3%	1,4K 38,1%	1,6K 27,5%	1,6K 27,5%	0,9K 36,5%
No da explicación	1,1K 4,5%	1,1K 4,6%	0,1K 0,6%	1,3K 17,0%	1,4K 16,5%	0,1K 2,2%	0,8K 11,3%	0,8K 10,8%	0,1K 1,6%	0,7K 13,0%	0,6K 10,9%	0,1K 2,7%
Red	0,6K 2,5%	0,6K 2,6%	0,0K 0,4%	0,7K 8,3%	0,9K 10,8%	0,1K 1,5%	0,4K 5,5%	0,5K 6,2%	0,1K 1,5%	0,5K 9,0%	0,5K 8,6%	0,0K 1,9%
Modelo de actuación	0,6K 2,4%	0,5K 2,3%	0,2K 1,3%	0,7K 8,7%	0,6K 7,4%	0,1K 2,4%	0,3K 3,5%	0,3K 3,5%	0,1K 1,7%	0,3K 4,7%	0,2K 3,5%	0,0K 1,8%
Información fallas en servicio	0,7K 2,9%	0,8K 3,3%	0,6K 4,8%	0,2K 2,2%	0,2K 2,5%	0,2K 4,4%	0,2K 2,4%	0,2K 2,9%	0,2K 4,3%	0,2K 2,9%	0,2K 2,9%	0,1K 3,2%
Estabilidad	0,4K 1,8%	0,5K 2,0%	0,0K 0,2%	0,5K 6,6%	0,6K 6,7%	0,0K 1,1%	0,3K 4,5%	0,3K 4,6%	0,0K 0,7%	0,3K 5,8%	0,3K 5,0%	0,0K 1,7%
Cancelación del servicio	0,7K 2,9%	0,7K 3,1%	0,6K 4,4%	0,2K 2,0%	0,2K 2,0%	0,1K 2,9%	0,2K 2,2%	0,2K 2,2%	0,1K 3,4%	0,1K 1,7%	0,1K 2,2%	0,1K 2,8%
Recuperación equipos	0,7K 3,2%	0,8K 3,2%	0,3K 2,7%	0,2K 2,0%	0,1K 1,8%	0,1K 1,9%	0,1K 1,8%	0,2K 2,1%	0,1K 1,7%	0,1K 1,7%	0,1K 2,2%	0,1K 2,3%
Incumplimiento de agenda	0,5K 2,3%	0,5K 2,2%	0,3K 2,4%	0,1K 1,1%	0,1K 1,2%	0,1K 1,7%	0,1K 1,9%	0,1K 1,9%	0,1K 1,6%	0,1K 1,1%	0,1K 1,3%	0,0K 1,6%
Decodificador	0,5K 2,3%	0,5K 2,2%	0,3K 2,6%	0,1K 1,0%	0,1K 1,1%	0,1K 2,1%	0,1K 0,9%	0,1K 1,0%	0,1K 1,8%	0,1K 1,1%	0,1K 1,4%	0,1K 2,3%
Canales adicionales	0,4K 1,7%	0,4K 1,7%	0,2K 1,8%	0,1K 0,8%	0,1K 0,9%	0,1K 1,6%	0,1K 1,3%	0,1K 0,9%	0,0K 1,0%	0,1K 1,1%	0,1K 1,5%	0,0K 1,6%
Mala asesoría	0,3K 1,4%	0,3K 1,4%	0,2K 1,8%	0,1K 1,8%	0,2K 2,0%	0,1K 1,3%	0,1K 1,3%	0,1K 1,1%	0,0K 0,9%	0,1K 1,2%	0,1K 1,2%	0,0K 0,8%
Fallas prestación servicio	0,2K 0,8%	0,2K 0,9%	0,0K 0,2%	0,2K 2,6%	0,2K 3,0%	0,0K 0,6%	0,1K 1,7%	0,2K 2,1%	0,0K 0,7%	0,1K 2,4%	0,2K 2,8%	0,0K 0,3%

Tópico por participación



5 Análisis

Cuando las empresas aprovechan los datos permiten el alcance de relaciones, estrategias y empoderamiento necesario para dar respuesta de una manera oportuna a sus clientes así lo menciona Durán en el 2021, es por ello, que TIGO en la búsqueda del aprovechamiento de la analítica de datos se ha centrado en generar iniciativas que permitan mejor la toma de decisiones rápidas y mejor informadas. Con relación a lo anterior, el área de PQR's de la empresa ha venido trabajando en miras de simplificar los procesos de análisis de experiencias de los clientes por medio de tableros de visualización que ayudan a ver los datos de una forma más estructurada y consolidada.

El equipo de PQR's viene centrando sus energías en lograr minimizar tiempos en los que la manualidad de las operaciones consume gran parte del día a día. De ahí que, que la herramienta creada para clasificar y analizar las causas de las quejas por servicio usando NLP apunte al direccionamiento existente en la empresa enfocado a las soluciones analíticas.

A través de los resultados presentados en el anterior apartado se muestran los diferentes elementos que se pueden extraer en el despliegue final de la herramienta. Donde, con los filtros seleccionados, se pudo encontrar una mayor participación de inconformidades con relación a los medios de atención del usuario y sobre las fallas en el servicio, los cuales son los puntos de inflexión más críticos con relación a la experiencia al cliente de la empresa. Por otro lado, el tópico de suplantación si bien no presenta grandes cantidades en el número de reclamaciones, si presenta un aumento representativo en los dos últimos períodos analizados, por lo que sería una arista la cual debe ser revisada con lupa.

Es importante mencionar que los resultados analizados corresponden a un primer acercamiento de inputs a la herramienta, es por ello, que los diccionarios que recibe el desarrollo fueron ingresados tanto para clientes móviles como para clientes fijos con una construcción básica de palabras, tópicos y criterios, donde fue necesario la creación de diccionarios sin mucha

minucia, con el fin de poder comprobar el funcionamiento de la herramienta. Y es que, a pesar de que el porcentaje de tipificación tuvo muy buenos resultados, debido a que este indicador mira, de las palabras que son válidas, cuantas logran tener alguna tipificación, pero esto no asegura que se haga una correcta asignación de tópico según la palabra. Lo anterior, sustentado en que, fue necesario reciclar información de los diccionarios construidos para la herramienta de NLP de queja factura y cada letra de las PQR's tiene un naturaleza diferente respecto a lo que el cliente manifiesta., por lo que, la intención de cada la palabra respecto al tópico, criterio y priorización es diferente. En la actualidad el área involucrada viene trabajando en la construcción de unos diccionarios más cercanos al funcionamiento de las quejas por servicio y donde se tiene se tiene claridad sobre la importancia de que se encuentren nutridos y actualizados dichos elementos de entrada a la herramienta, en vista de que el éxito de la herramienta depende en su mayoría de que tan bien estén fundamentados.

El éxito de la herramienta analítica a través del tiempo está muy ligado al monitoreo constante de posibles fallas que puedan ocurrir para evitar que se convierta en un recurso con poca o nula usabilidad por parte del personal interesado. Así mismo, es importante que se sepa sobre la existencia de la herramienta para que se maximice el potencial de uso y, además se explore la herramienta de manera acertada, permitiendo la extracción y aprovechamiento de información relevante para buscar soluciones en los enfoques indicados.

A través de esta herramienta de NLP se puede identificar en dónde está fallando la compañía y donde es posible encontrar aquellos focos que tienen mayor acción de mejora. Asimismo, la iniciativa funciona como una especie de semáforo de alerta, donde mes a mes se analiza aquellos valores que son poco usuales en el comportamiento habitual de volúmenes en tópicos y criterios.

La mitigación de los problemas de los clientes, no solo se solucionan con la detección de sus causas, sino que también depende del accionar de la compañía, por lo que, saber que áreas intervenir, junto a los temas y responsables dentro del ecosistema gerencial es de suma

relevancia. De igual forma, se debe hacer una mesa de trabajo con el personal involucrado, para empezar a buscar una causa raíz y lograr mitigar el impacto.

6 Conclusiones

Dada a la falta de entendimiento encontrada en la estructura de PQR's, más propiamente en las quejas por servicio se opta por la implantación de una herramienta por medio de la técnica de NLP para clasificar y analizar las causas por las que los clientes comunican sus inconformidades categorizadas como quejas por servicio en la empresa TIGO COLOMBIA. Para su creación se tuvo en cuenta las necesidades y requerimientos existentes en la vicepresidencia de experiencia al cliente. Asimismo, a través de herramientas como SQL SERVER, Python, y Tableau se tejió el sistema de tipificación de texto.

Dentro, del despliegue final de información se presenta un tablero con tres vistas, que permite ver de forma detallada según la mirada del lente, discriminado por unidades de negocio, por unidades estratégicas, por lapsos de tiempo y por ubicaciones geográficas. Por lo que, es posible hacer estudio puntual a cada combinación de entendimiento para la empresa.

La herramienta de NLP muestra esos puntos críticos, en donde se tiene mayor acción de mejora, funcionando como una especie de semáforo de alerta, donde mes a mes, en el mejor de los casos, se pueden percibir un volumen constante de ciertas tipificaciones, pero si comienza a generarse alteración en el volumen de quejas, como incrementos para ciertas categorizaciones, se puede extraer estos comportamientos anómalos o atípicos para ser atacados.

Uno de los puntos positivos de esta implementación es la optimización de tiempo a la hora de buscar las principales causas por las que un cliente se comunica, debido a que con anterioridad se revisaban las quejas por servicio de forma manual, generando sesgo en los hallazgos según los grupos de textos analizados. Por otra parte, entre más información y conocimiento se tenga sobre los malestares del cliente con base fiable y veraz, más amplio será el abanico de posibilidades de tomar las decisiones más oportunas, en cuestión a los puntos a atacar para generar planes de acción adecuados.

7 Recomendaciones

Un enfoque que podría complementarse con la herramienta de clasificación de NLP es el clustering de texto, el cual puede ofrecer el potencial de descubrir estructuras implícitas y temas significativos que pueden emerger del contenido de los grupos de texto. Por lo que, el clustering podría servir para generar insight más nuevos y poco visibles que aporten al entendimiento de lo que comunica el cliente.

Por otro lado, a futuro se podría unificar todas las letras en un solo tablero de visualización, donde utilizando la herramienta creada, se pueda ingresar los diccionarios para cada una de las PQR's y poder de forma consolidada toda la información de reclamaciones de los clientes.

Referencias

- ActiveState. (2022, 9 agosto). *What Is Pandas in Python? Everything You Need to Know*.
<https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/>
- Barreiro, A. (2016). La importancia del diseño en la Experiencia de Cliente. Escuela de organización industrial. Recuperado 15 de enero de 2013, de
https://asociaciondec.org/wp-content/uploads/2016/11/Dise%C3%B1o_en_CX_EOI_Ficha3.pdf
- Béltran, D. (2022). TGMISC221002 CHATBOT MENTES. Repositorio Pontificia Universidad Javeriana.
https://repository.javeriana.edu.co/bitstream/handle/10554/61844/475-attachment-1655445398-Memoria_TGMISC221002.pdf?sequence=1
- Duran, J. (2021). *La importancia y los beneficios del uso de datos en nuestra empresa*. PYMESWORLD. <https://pymesworld.com/importancia-datos-empresa/>
- Espinel, J. (2019, 14 diciembre). DISEÑO DE UNA METODOLOGÍA ANALÍTICA PARA EL TRATAMIENTO DE QUEJAS Y RECLAMOS PARA EL SECTOR HOGARES Y MOVILIDAD ETB. Repositorio Universidad de los Andes.
<https://repositorio.uniandes.edu.co/bitstream/handle/1992/43935/u830389.pdf?sequence=1>
- Javatpoint. (2021). *Dateutil module in Python - Javatpoint*. www.javatpoint.com.
<https://www.javatpoint.com/dateutil-module-in-python>
- McAfee, A. (2012, 8 octubre). Big Data: The Management Revolution. Harvard Business Review. <https://hbr.org/2012/10/big-data-the-management-revolution>
- PyJWT. (2022). *Welcome to PyJWT — PyJWT 2.6.0 documentation*.
<https://pyjwt.readthedocs.io/en/stable/>
- Pyodbc. (2022, noviembre 16). PyPI. <https://pypi.org/project/pyodbc/>
- Rodríguez, N. (2018). *Análisis y desarrollo de las PQRS en Industrias Metálicas El Jordán para lograr una mejora en el control de sus procesos en la ciudad de Bogotá*. Unilibre. <https://repositorio.unilibre.edu.co/bitstream/handle/10901/15891/ARTICULO%20Especializacion.%20.pdf?sequence=2&isAllowed=y>
- Topcoder. (2021, 5 noviembre). *Excel Automation with Openpyxl in Python*.
<https://www.topcoder.com/thrive/articles/excel-automation-with-openpyxl-in-python>

Villarroel, F. & Burton, J. (2014). Analyzing Customer Experience Feedback Using Text Mining: A Linguistics- Based Approach. *Journal of Service Research*, 14–17, 17-21.

Villaverde, N. (2017). Nuevas técnicas estadísticas: Text Mining en Web. *ruc.udc*.
https://ruc.udc.es/dspace/bitstream/handle/2183/19947/VillaverdeMedina_Nicole_TFG_2017.pdf?sequence=2

Zayat, A. (2019, septiembre). Análisis Automático de Experiencia de Cliente en los Canales de Atención Digitales. UAM. Recuperado 15 de enero de 2023, de
https://repositorio.uam.es/bitstream/handle/10486/688797/zayat_mata_ana_tfm.pdf?sequence=1