

Pontificia Universidad Javeriana de Cali
Facultad de Ingeniería y Ciencias
Maestría en Ciencia de Datos

Análisis de sentimiento de las llamadas del Call Center de la Universidad Javeriana de Cali mediante el uso de Ciencia de Datos

Luis Miguel Semma Alvarez
Ivan Felipe Quiza Vargas
Nelson Andres Agreda Villota

Proyecto Aplicado para optar al título de Magister en Ciencia de Datos

Director: Dr. Gloria Inés Álvarez
Codirector: Dr. Diego Linares.

Junio 9 de 2025



Resumen

Campo	Contenido
Título	Análisis de sentimiento de las llamadas del Call Center de la Universidad Javeriana de Cali mediante Ciencia de Datos
Área de trabajo	Educación - Atención al Cliente
Tipo de proyecto	Aplicado
Autores	Luis M. Semma (8992231), Iván F. Quiza (8993339), Nelson A. Agreda (8985748).
Correo electrónico	lmsemma@javerianacali.edu.co, felipe4633@javerianacali.edu.co, 8985748@javerianacali.edu.co
Dirección y teléfono	Calle 147a 48 -53 Bogotá D.C, 3006598938 – 3152208872 - 311 3864474
Director	Dra. Gloria Inés Álvarez.
Vinculación del director	Docente.
Correo electrónico del director	galvarez@javerianacali.edu.co.
Codirector si aplica	Si
Codirector	Dr. Diego Luis Linares Ospina
Vinculación del director	Docente.
Correo electrónico del codirector	dlinares@javerianacali.edu.co.
Grupo o empresa que lo avala	Pontificia Universidad Javeriana de Cali.
Palabras clave	Call Center, Sentimiento, GCP, SVM, NMF, LIME.
Fecha de inicio	22 de enero 2024.
Fecha de finalización	20 de junio 2025.

Cuadro 1: Ficha resumen del proyecto

Resumen: El presente proyecto se enfocó en el análisis de sentimiento de las llamadas del Call Center de la Universidad Javeriana de Cali mediante técnicas de ciencia de datos. Su finalidad fue desarrollar una herramienta que permita analizar el sentimiento expresado en cada llamada y

clasificar las interacciones según su contenido emocional, contribuyendo a mejorar la calidad del servicio y apoyar la toma de decisiones informadas. Para ello, se emplearon técnicas como GCP Speech-to-Text para la transcripción de audios, modelos de clasificación supervisada como SVM y RNN y clasificación no supervisada con NMF. Los resultados muestran un buen desempeño, destacando la utilidad del modelo SVM-TFIDF con métricas de F1 y AUC positivas. Esta solución representa un avance significativo en la gestión de la información y atención al cliente en el entorno educativo

Índice general

1. Definición del problema	3
1.1. Planteamiento del problema.	3
1.2. Formulación del problema.	4
2. Objetivos del proyecto	6
2.1. Objetivo general	6
2.2. Objetivos específicos	6
3. Marco de referencia	7
3.1. Marco Teórico	7
3.1.1. Sentimiento	7
3.1.2. Análisis de sentimiento	8
3.1.3. Transcripción de audios	9
3.1.4. Limpieza de datos	10
3.1.5. Clasificación de datos	10
3.1.5.1. Clasificación supervisada	11
3.1.5.2. Clasificación no supervisada	11
3.1.6. Modelos de Machine Learning (ML)	12
3.1.6.1. Redes Neuronales (RNN)	12
3.1.6.2. Maquinas de soporte vectorial (SVM)	13
3.1.6.3. Random Forest (RDF)	13
3.1.6.4. Non-Negative Matrix Factorization (NMF)	13
3.1.6.5. Métricas de evaluación	14
3.1.7. Tecnología	15
3.1.7.1. GCP Google Cloud SPEECH-TO-TEXT	15
3.1.7.2. GCP Google Cloud Storage	16
3.1.7.3. Google Colab	16
3.1.7.4. Python	17
3.2. Antecedentes	17
4. Preparación de los datos	20
4.1. Prueba de concepto de los servicios de transcripción	20
4.1.1. Librería Whisper	20
4.1.2. AWS Transcribe	21
4.1.3. GCP Speech To Text	21

4.2.	Recolección y cargue de los datos	21
4.3.	Procesamiento de los datos	22
4.3.1.	Transcripción	22
4.3.2.	Etiquetado manual de los datos	22
4.3.3.	Preprocesamiento del texto	23
4.3.3.1.	Limpieza del texto	23
4.3.3.2.	Normalización	24
4.3.4.	Esquema del procesamiento	25
4.3.5.	Vectorización de los datos	25
4.3.5.1.	One Hot Encoding	26
4.3.5.2.	Term Frequency-Inverse Document Frequency (TF-IDF)	26
4.3.5.3.	Word2Vec	27
4.4.	Selección de modelo y etiquetado	27
4.4.1.	Modelo de clasificación supervisada	27
4.4.1.1.	Prueba de concepto modelo de clasificación	28
4.4.1.2.	Resultados prueba de concepto modelos de clasificación	30
4.4.2.	Etiquetado de los datos	32
5.	Modelado y Evaluación	33
5.1.	Entrenamiento del modelo SVM	33
5.1.1.	Resultados del entrenamiento del modelo SVM	33
5.2.	Modelo de clasificación no supervisada	36
5.2.1.	Non-Negative Matrix Factorization (NMF)	36
5.2.1.1.	Resultados modelo de NMF	37
6.	Análisis de resultados	39
6.1.	Análisis Clasificación Supervisada	40
6.2.	Análisis Clasificación No Supervisada	41
6.3.	Análisis Implementación	43
6.4.	Discusión de resultados orientado al negocio	43
6.4.1.	Canal Call Center	43
6.4.2.	Área de Mercadeo:	44
6.4.3.	Implementación de la aplicación:	44
7.	Implementación	45
7.1.	Esquema de la aplicación	45
8.	Conclusiones y Trabajos futuros	49
8.1.	Conclusiones	49
8.2.	Trabajos futuros	50
	Bibliografía	51

Índice de figuras

4.1. Esquema Procesamiento	25
4.2. Curva de ROC	31
4.3. Matriz de confusión SVM.	31
5.1. Curva de ROC SVM	34
5.2. Explicabilidad con LIME	35
6.1. Distribución de llamadas por sentimiento	40
6.2. Topics generados por NMF por cada sentimiento	41
7.1. Esquema Aplicación	45
7.2. Pantalla Cargue	46
7.3. Pantalla Procesamiento	47
7.4. Pantalla Resultado Modelo	47
7.5. Pantalla Evaluación Modelo LIME	48

Índice de cuadros

1.	Ficha resumen del proyecto	2
4.1.	Resumen de los hiperparámetros del modelo de red neuronal	29
4.2.	Resultados del F1-score por modelo y técnica de vectorización	30
5.1.	Evaluación general del modelo por clase	34
5.2.	Hiperparámetros del modelo NMF por cada sentimiento	37

Introducción

En el ámbito de las instituciones educativas, la atención a las necesidades y preocupaciones de los estudiantes y del personal fue crucial para mantener altos estándares de satisfacción y calidad.

La Universidad Javeriana de Cali, comprometida con la excelencia académica y el bienestar de su comunidad, enfrenta desafíos significativos en la gestión de las numerosas llamadas telefónicas que recibe su Call Center. Estas llamadas, que incluyen consultas, solicitudes, quejas y comentarios, relacionadas con los diferentes servicios de la universidad y programas académicos, estas llamadas son una fuente invaluable de información sobre la experiencia de los usuarios. Sin embargo, el volumen y la naturaleza de las grabaciones de audio dificultaron un análisis eficiente de cara identificar oportunidades de mejora en este servicio.

El proyecto se enfocó en el análisis de sentimiento de las llamadas del Call Center de la Universidad Javeriana de Cali mediante técnicas de ciencia de datos. El objetivo es desarrollar una herramienta que permita analizar el sentimiento expresado en cada llamada y clasificar las interacciones en función de su contenido emocional. Este análisis proporcionará puntos detallados sobre las necesidades y opiniones de los usuarios, mejorando la calidad del servicio y facilitando la toma de decisiones informadas.

A través del uso de técnicas avanzadas de aprendizaje automático y procesamiento de lenguaje natural, se espera desarrollar un modelo de clusterización que permita identificar una temática en específico según el sentimiento en las llamadas y proporcionar una herramienta para la visualización de los resultados. Este enfoque no solo optimizará la eficiencia del Call Center, sino que también ofrecerá una fuente de datos valiosa para la mejora continua de los servicios universitarios.

Este proyecto se llevó a cabo entre los años 2024 y 2025, en Santiago de Cali y aspira a establecer un precedente en el uso de ciencia de datos para mejorar la atención al cliente en el ámbito educativo. La implementación de esta solución tecnológica representa un avance significativo en la gestión de la información y en la capacidad de la universidad para responder proactivamente a las necesidades de su comunidad.

Definición del problema

1.1. Planteamiento del problema.

Las instituciones educativas están constantemente buscando formas innovadoras de mejorar la experiencia de sus estudiantes y usuarios. La Universidad Javeriana de Cali, como centro educativo comprometido con la excelencia académica y el bienestar de su comunidad, recibe una gran cantidad de llamadas telefónicas que abarcan una amplia variedad de consultas, solicitudes, quejas y comentarios por parte de estudiantes, profesores y personal administrativo. El proceso actual de gestión de estas llamadas se enfrenta a varios desafíos significativos. En primer lugar, el volumen de llamadas puede ser abrumador para el personal encargado de atenderlas, lo que puede llevar a retrasos en la respuesta y en la resolución de los problemas planteados. Además, la información recopilada a partir de estas llamadas se encuentra en formato de audio, lo que dificulta su análisis y aprovechamiento eficientes.

Uno de los mayores retos radica en la necesidad de comprender y evaluar el sentimiento y la intención detrás de cada llamada. Las llamadas pueden variar desde consultas simples y expresiones de satisfacción hasta quejas y reclamaciones más complejas. En este contexto, surge la necesidad de desarrollar un modelo analítico apoyado en la ciencia de datos que permita predecir el sentimiento hallado en la llamada. También, dependiendo de cada sentimiento, se buscó identificar el tema específico tratado en la llamada.

El principal reto desde la ciencia de datos fue desarrollar un sistema capaz de reconocer y clasificar el sentimiento expresado en cada llamada ya sea positivo, negativo o neutro, así como poder segmentar las diferentes llamadas que ingresan a esta línea. El desarrollo de esta solución no solo mejorará la eficiencia y la calidad del servicio al cliente de la universidad, sino que también proporcionará una invaluable fuente de información para la toma de decisiones basada en datos. Al comprender mejor las necesidades, preocupaciones y opiniones de su comunidad, la Universidad Javeriana de Cali estará en una posición privilegiada para implementar medidas proactivas que mejoren continuamente su oferta educativa y sus servicios institucionales.

La implementación de esta solución no solo mejorará la eficiencia y calidad del servicio al cliente de la universidad, sino que también proporcionará una fuente de información valiosa para la toma de decisiones estratégicas basadas en datos. Al comprender mejor las necesidades, preocupaciones y opiniones de su comunidad, la Universidad Javeriana de Cali estará en una posición privilegiada para anticiparse a los problemas, mejorar continuamente su oferta educativa y fortalecer los servicios institucionales. Asimismo, esta solución puede sentar las bases para una futura expansión del análisis de datos a otros puntos de contacto con la comunidad universitaria, integrando diferentes canales de comunicación en un enfoque más holístico de mejora continua.

1.2. Formulación del problema.

La Universidad Javeriana de Cali, reconocida por su compromiso con la excelencia académica y el bienestar de su comunidad, enfrenta un reto creciente en la gestión de un alto volumen de llamadas telefónicas. Estas interacciones, que incluyen consultas, solicitudes, quejas y comentarios de estudiantes, profesores y personal administrativo, son fundamentales para mantener una comunicación efectiva y garantizar la satisfacción de los usuarios. Sin embargo, el proceso de atención a estas llamadas presenta varios desafíos que ponen a prueba la capacidad de la universidad para responder de manera eficiente y oportuna a las necesidades de su comunidad.

Es importante para cualquier entidad o empresa conocer cómo están siendo atendidas las necesidades y se está haciendo la gestión telefónica, esto porque es uno de los canales de atención más importantes que se pueden tener, por tal razón identificar cualquier oportunidad de mejora basado en un análisis del proceso actual es ideal para cualquier organización o institución. Tener una lectura de estas llamadas y que sea de una manera automática y óptima con el apoyo de la inteligencia artificial sería una solución ideal para este proceso.

Uno de los desafíos importantes radica en el formato de la información, estas llamadas se registran en formato de audio, lo que genera tiempos elevados para el análisis sistemático de las llamadas.

Esto puede impedir que la universidad extraiga conclusiones útiles que podrían informar para tomar decisiones estratégicas y mejorar los servicios prestados en la universidad. En este contexto, el uso de herramientas tecnológicas avanzadas que permitan identificar y clasificar el sentimiento detrás de las interacciones telefónicas aumenta la capacidad de la institución para comprender a fondo las preocupaciones y necesidades de su comunidad. Teniendo en cuenta lo anterior sin un análisis adecuado de estas interacciones, las oportunidades para mejorar la experiencia del usuario y optimizar los procesos de atención se ven reducidas.

Las causas de estos problemas son diversas. En primer lugar, el elevado volumen de llamadas puede generar retrasos en la atención y afecta la calidad del servicio. Además, el reconocimiento

limitado del lenguaje natural, es decir, la capacidad de interpretar el contenido emocional y semántico de las llamadas limita la comprensión del verdadero sentimiento expresado por los usuarios y las personas encargadas del Call Center, lo que afecta la eficiencia de la detección del sentimiento buscando la resolución de problemas como en la calidad de la atención.

El desarrollo e implementación del análisis de sentimiento de las llamadas del Call Center de la Universidad Javeriana de Cali apoyado por la inteligencia artificial, trae varios retos importantes a tener en cuenta, el primero de ellos es lograr contar con un DataSet generado por un proceso de transcripción de audio, otro es que este dataset tenga la estructura y la calidad suficiente para contar con un insumo de datos para el análisis de sentimiento. Luego de esto el siguiente reto es el etiquetado, tener unos criterios claros y estandarizados para el proceso de etiquetado es vital para contar con un correcto entrenamiento del modelo y por último elegir las técnicas y modelos adecuados para la clusterización y entrenamiento es de vital importancia en el proceso de implementación.

En resumen, la Universidad Javeriana de Cali se enfrenta a un desafío que tiene en cuenta la operatividad del Call Center como la experiencia general de sus usuarios. La implementación de un sistema de análisis de sentimiento, basado en técnicas avanzadas de ciencia de datos, puede

ser clave para mejorar la capacidad de respuesta y la eficiencia en la atención, brindando además una fuente valiosa de información para la toma de decisiones estratégicas a largo plazo y puede significar una oportunidad para entrar a evaluar los distintos procesos que se manejan al interior de la universidad.

Objetivos del proyecto

2.1. Objetivo general

Desarrollar una herramienta para el análisis de sentimiento de las grabaciones de llamadas del Call Center de la Universidad Pontificia Javeriana de Cali utilizando técnicas de aprendizaje automático, para mejorar la clasificación de las consultas y la evaluación de la satisfacción del usuario.

2.2. Objetivos específicos

- Realizar el preprocesamiento de los datos de las grabaciones de las llamadas del Call Center para el uso en los modelos.
- Desarrollo del modelo de clasificación de las llamadas para identificar el nivel de satisfacción, por medio del análisis de sentimiento.
- Elaborar un modelo de clustering para segmentar las llamadas partiendo del análisis de sentimiento obtenido en el resultado del modelo de clasificación.
- Elaborar un informe que permita visualizar de manera tabular y resumida los resultados obtenidos por los modelos.
- Desarrollo de una interfaz que permita ejecutar el modelo en tiempo real para ver sus resultados.

Marco de referencia

3.1. Marco Teórico

3.1.1. Sentimiento

En psicología, un sentimiento se concibe como una contestación emocional que sucede cuando se tiene un estímulo, y se presenta por una persona, un objeto, un animal o alguna situación particular. Este es la respuesta de un proceso donde las emociones se unen con la interpretación sensata de lo que vivimos [1]. Por otro lado, las emociones son reacciones que se dan de forma inmediata, casi instintiva, que sacuden ante algo que afecta en ese instante. Los sentimientos, por el contrario, son más elaborados: se dan cuando se da el espacio de reflexionar sobre esas emociones y les damos un significado [2].

Si ponemos un ejemplo, podemos pensar en un instante en el que se está con un grupo de amigos y alguno cuenta una historia que los hace reír a todos. Esa alegría que sienten, ese lazo con el momento es un sentimiento. Las emociones son penetrantes pero efímeras, como un chispazo; los sentimientos, por el contrario, tienen más profundidad, como cuando tienes una idea que se queda contigo y tus pensamientos por un buen rato [3]. Una buena forma de verlo es imaginar las emociones como notas individuales y los sentimientos como la melodía completa que forman [4].

La diferencia entre los dos es clara: las emociones son respuestas de forma automáticas que da el cuerpo, a diferencia de los sentimientos, los cuales requieren de un procesamiento mental el cual debe tener un contexto y un sentido [4]. Esto fue primordial para hacer los analizar de los textos, debido a que las palabras pueden irradiar tanto las reacciones emocionales inmediatas como las formas más profundas de una persona. Adicional, los sentimientos tienden a influir en las decisiones, comportamientos e incluso en nuestra salud, lo que sugiere que estudiarlos en textos puede ayudar a entender cómo actúan las personas en distintos contextos [5].

La comparación con la música resulta particularmente útil, porque los sentimientos son complejos y no siempre se pueden reducir a categorías simples. Por eso, las herramientas para analizarlos deberían captar esa riqueza en lugar de limitarse a clasificaciones rígidas [3].

En la psicología, se tienen muchas teorías las cuales permiten explicar el inicio de las emociones. Dichas teorías genera un marco para entender los sentimientos. La teoría de la evolución indica dice que las emociones se dan para ayudar a sobrevivir y relacionarnos. La teoría de James-Lange indica que primero se debe experimentar una reacción física (como un aumento del ritmo cardíaco) y luego interpretamos la emoción [6]. La teoría de Cannon-Bard indica que la emoción y la reacción física pueden suceder de forma simultánea [7]. Schachter-Singer plantea que las reacciones físicas anteceden a la emoción [8], pero esta se define según el contexto. Por último, la teoría de la emoción

construida sugiere que las emociones no son innatas, sino que las construimos al interpretar nuestras experiencias [9] [10]. Cada teoría ofrece una perspectiva distinta y dependiendo del enfoque del análisis, podemos usar una para identificar patrones en textos, como el contexto o las descripciones de reacciones físicas.

3.1.2. Análisis de sentimiento

El análisis de sentimiento radica en identificar e interpretar las emociones y opiniones indicadas en un contenido. Al imaginar que revisamos los comentarios y reseñas de un restaurante, de esta forma podemos observar lo siguiente: ¿los clientes quedaron contentos con la comida o por el contrario les generó descontentos o molestias por el servicio? Eso es lo que se busca en este tipo de proceso: identificar el tono emocional detrás de las palabras [11]. Esto no se trata solo de identificar palabras buenas y negativas o positivas y negativas, Por el contrario, se debe percibir las emociones, actitudes y los puntos de vista que se tienen en el contenido del texto [12]. Esto consiste en examinar una gran cantidad de información escrita para identificar patrones relevantes.

Con recursos como el análisis del lenguaje natural y la inteligencia artificial, somos capaces de estudiar textos de plataformas sociales, críticas de productos o reportajes para captar las emociones de las personas. No obstante, el idioma humano es complicado. Una expresión como “está bien” puede reflejar entusiasmo o simplemente aceptación, dependiendo de la situación. El análisis de sentimiento se puede abordar en varios niveles:

- **Identificación de polaridad:** Establece si un texto tiene una carga positiva, negativa o es neutral. En ocasiones, se emplean escalas más específicas, como "muy positivo." "ligeramente negativo"[12].
- **Reconocimiento de emociones:** Detecta emociones concretas, como felicidad, tristeza o ira [12].
- **Análisis centrado en aspectos:** Se interesa por factores concretos, como lo que se opina sobre el sabor de un platillo comparado con su costo en un restaurante [12].
- **Análisis de intención:** Intenta comprender cuáles son las intenciones de la persona, como comprar, quejarse o hacer una sugerencia [12].

El nivel que se elija dependerá del propósito del análisis. Si solo se necesita conocer el ánimo general, la polaridad es suficiente. Sin embargo, si se busca profundizar en las causas de esas emociones, es necesario un enfoque más exhaustivo.

Hay varios métodos para llevar a cabo el análisis:

- **Métodos léxicos:** Utilizan diccionarios que asignan valores emocionales (positivos o negativos) para evaluar el texto [13]. Herramientas como VADER, que se han desarrollado para redes sociales, reconocen incluso emoticonos y lenguaje coloquial [14]. Son veloces, pero pueden no captar matices como el sarcasmo.

- **Métodos de aprendizaje automático:** Se entrenan algoritmos con ejemplos para que reconozcan sentimientos de manera independiente [12]. Modelos como Naive Bayes, Máquinas de Vectores de Soporte (SVM) o transformadores como BERT son comunes. Ofrecen alta precisión, pero requieren grandes volúmenes de datos y potencia de cálculo.
- **Métodos híbridos:** Combinan enfoques léxicos y de aprendizaje automático para lograr un equilibrio entre sencillez y precisión [12].

Cada método presenta pros y contras. Los léxicos son sencillos de implementar, pero menos efectivos con textos que tengan ambigüedades. El aprendizaje automático proporciona mayor exactitud, pero consume más recursos. Los híbridos tratan de encontrar un balance.

3.1.3. Transcripción de audios

La conversión de audio es el proceso de transformar el contenido de un archivo de audio en texto escrito. Es similar a tomar notas detalladas durante una entrevista o un podcast. Este procedimiento es fundamental en investigaciones donde se examina el idioma hablado, ya sea para identificar temas o captar emociones [15].

Existen distintas maneras de generar la transcripción:

- **Literal:** Registra de forma exacta cada palabra, incluye silencios, muletillas y todo tipo de errores. Esta forma es útil cuando el proyecto desea analizar el estilo del habla de las personas.
- **Lectura corregida:** Esta forma elimina cosas no necesarias para ayudar a la lectura.
- **Inteligente:** En esta alternativa lo que se es concentrarse en el contenido importante y principal, dejando a un lado los aspectos irrelevantes.
- **Editada:** Aquí se realiza un resumen del texto sin cambiar su esencia.
- **Narrativa:** Registra de forma muy detallada el texto, esta opción es empleada mayormente en temas legales.

La manera de generar la transcripción va a depender del objetivo del análisis que se requiera. Si se requieren todos los detalles exactos, la transcripción literal sería la más adecuada; por lo contrario, si solo se busca tener el contenido esencial, una transcripción limpia la opción adecuada es la inteligente [16].

Las transcripciones son demasiado útiles, debido a que ofrecen un registro fidedigno el cual puede ser validado en distintas ocasiones para obtener significados más concretos. Los métodos de transcripción pueden ser:

- **Manual:** En este caso una persona realiza el proceso de escucha y posteriormente escribe. Es un método preciso y confiable, sin embargo, este puede tardar de 4 a 5 horas por cada hora de audio. Teniendo en cuenta lo anterior, se puede indicar que no es óptimo en términos de agilidad.

- Automático: En esta opción, un programa realiza todo el proceso de forma rápida. Sin embargo, es importante nombrar que en esta opción puede generar algunos errores cuando los audios son de baja calidad.
- Híbrido: Esta opción combina el uso de software con revisión humana para asegurar una mayor precisión.

Para lograr una mejor calidad en los resultados, es importante usar un micrófono de buena calidad y estar en un ambiente adecuado. Aplicaciones como Amberscript u Otter. ai hacen que esta actividad o proceso sea más fácil de realizar [17].

3.1.4. Limpieza de datos

La limpieza de datos hace referencia a la preparación de la información para hacer un análisis previo, eliminar errores, inconsistencias o datos que no son relevantes o importantes para el proceso [18]. Este es un paso esencial, debido a que la información que se gestiona de forma incorrecta puede llevar a conclusiones equivocadas. Esto es similar a ordenar un escritorio o un espacio antes de iniciar a trabajar en alguna tarea: si dicho espacio está desorganizado, el rendimiento de lo que se vaya a realizar se va a ver afectado. Realizar una limpieza efectiva de forma previa, mejora la calidad de los resultados que se obtengan, ahorran tiempo y reduce los gastos. Cuando se habla de los textos, se tienen algunas técnicas como se relacionan a continuación:

- Manejo de valores ausentes: Se rellenan con datos consistentes o se eliminan.
- Eliminación de duplicados: Se remueven registros que se repiten.
- Estandarización: Se unifican formatos, como fechas o tipografías.
- Preprocesamiento: Se borran signos de puntuación, se convirtió todo a letras minúsculas, se eliminan palabras vacías (como “y” o “de”), se segmenta el texto y se aplican técnicas como lematización o stemming.

En todos los ámbitos relacionados a la estrategia de los datos, la calidad de los mismos es fundamental y como sabemos la ciencia de los datos dentro de cualquier proceso hace parte de la estrategia de los datos, por ende la calidad de los datos es muy importante en estos procesos, pero no es solo por esta razón, la exactitud y eficiencia del proceso depende mucho de la calidad y completitud de los datos, por tal motivo es parte fundamental de cualquier proceso relacionado a esta práctica.

3.1.5. Clasificación de datos

La clasificación es una tarea de aprendizaje que consiste en predecir una etiqueta o categoría para una entrada dada, basándose en datos con algunas características. El objetivo es que el modelo aprenda a distinguir entre diferentes clases a partir de un conjunto de ejemplos, para luego poder clasificar correctamente nuevos datos [19].

En el caso de la clasificación de texto, según Sebastiani (2002), este proceso implica el uso de técnicas de aprendizaje automático para entrenar clasificadores capaces de categorizar textos en función de su contenido. El enfoque tradicional se basa en representar cada documento mediante un vector de características, por ejemplo, palabras o n-gramas— y utilizar algoritmos como Naive Bayes, SVM o redes neuronales para aprender patrones, esta técnica ha demostrado ser efectiva en análisis de sentimiento y organización automática de contenido digital [20].

3.1.5.1. Clasificación supervisada

La clasificación supervisada es una técnica de aprendizaje automático que utiliza conjuntos de datos previamente etiquetados para entrenar algoritmos. Su objetivo es que, una vez entrenados, estos algoritmos sean capaces de clasificar correctamente nuevos datos no vistos[21]. En los análisis de sentimiento que se realizan, la clasificación supervisada involucra realizar entrenamientos de los modelos con textos o audios, los cuales ya han sido etiquetados previamente con sus sentimientos correspondientes (por ejemplo, positivo, negativo, neutral). El objetivo con el cual se realiza esto es para que el modelo aprenda la relación entre las características de los datos (como palabras o características de audio) y las etiquetas de sentimiento, con lo cual podrá predecir de forma asertiva los sentimientos de nuevos datos.

Algunos de los algoritmos que se utilizan en la categoría de supervisados para el análisis de sentimiento, incluyendo Naive Bayes, Máquinas de Vectores de Soporte (SVM), Regresión Logística, Árboles de Decisión, Random Forests y Redes Neuronales (incluyendo CNNs, RNNs y Transformers). La eficiencia de los algoritmos deriva de la calidad y cantidad de todos los datos que se hayan entrenado y etiquetado, así mismo, como las técnicas de extracción de características utilizadas para representar los datos. La clasificación supervisada generalmente logra una mayor precisión cuando se dispone de suficientes datos etiquetados [22]

3.1.5.2. Clasificación no supervisada

La clasificación no supervisada, al contrario de la supervisada, aprende de datos sin necesidad de etiquetar previamente para así evidenciar patrones inherentes o agrupaciones dentro de los datos [17]. En el análisis de sentimiento, las metodologías no supervisadas no necesitan de datos de entrenamiento previamente preetiquetados. Por el contrario, utilizan distintos enfoques para deducir el sentimiento a partir de los datos en sí.

Los enfoques más comunes para la clasificación no supervisada en los análisis de sentimiento tienen métodos basados en léxicos, técnicas de clustering, modelado de temas y el uso de incrustaciones de palabras. Las metodologías basadas en vocabularios utilizados en diccionarios de palabras previamente compilados con polaridades de sentimiento asociadas para fijar el sentimiento de un texto. Las técnicas de clustering aglomeran textos o audios relacionados con el sentimiento en función de la similitud de sus características [18]. El modelado de temas identifica los temas subyacentes en una colección de textos, donde algunos temas pueden estar asociados con un sentimiento particular [23]. El análisis de sentimiento no supervisado es útil en escenarios con datos limitados o sin etiquetar y para la exploración inicial de datos [24].

3.1.6. Modelos de Machine Learning (ML)

Los modelos de aprendizaje automático son herramientas avanzadas que examinan datos para identificar patrones. En el análisis de sentimiento, son sumamente efectivos para procesar valoraciones, comentarios o publicaciones [25]. Algunos modelos comunes incluyen:

- Naive Bayes: Sencillo y eficaz, óptimo para grandes volúmenes de datos.
- Máquinas de Vectores de Soporte (SVM): Útiles para clasificaciones complejas.
- Regresión Logística: Proporciona resultados que se pueden entender como probabilidades.
- Bosques Aleatorios: Integra múltiples árboles de decisión para aumentar la exactitud.
- Redes Neuronales: Modelos como BERT o GPT, capaces de reconocer matices sutiles.

La elección del modelo depende de la dificultad del análisis y de la cantidad de información disponible.

3.1.6.1. Redes Neuronales (RNN)

Las Redes Neuronales Recurrentes (RNN) son una arquitectura frecuente en el aprendizaje profundo, concebidas para manejar información secuencial como texto, sonido o secuencias temporales. En contraposición a las redes neuronales convencionales, las RNN poseen conexiones que generan ciclos, lo que les facilita conservar una memoria de las entradas anteriores en la secuencia [26]. Esta característica resulta especialmente beneficiosa para labores como el estudio de emociones, donde el contexto de las palabras previas puede tener un impacto en la comprensión del sentimiento presente. Las RNN tienen la capacidad de modelar dependencias secuenciales en los datos, lo que las convierte en idóneas para labores de procesamiento del lenguaje natural, incluyendo el estudio de las emociones [27]. Hay diversas arquitecturas de RNN, entre las que se incluyen las RNN estándar, las Redes Neuronales Bidireccionales Recurrentes (BRNN), las Redes de Memoria a Corto Plazo (LSTM) y las Unidades Recurrentes Cerradas (GRU). Las RNN convencionales pueden experimentar problemas al aprender dependencias a largo plazo debido al problema del gradiente de fuga. Las BRNN analizan la secuencia de entrada en ambas direcciones con el objetivo de incrementar la exactitud. Las LSTM y las GRU están concebidas para tratar el problema de la dependencia a largo plazo a través de la implementación de sistemas de memoria. En el estudio de emociones, las RNN tienen la capacidad de analizar una serie de palabras en una reseña o un tweet y anticipar la percepción general (positiva, negativa o neutral) [27]. El modelo adquiere la habilidad de relacionar determinadas palabras o secuencias de palabras con emociones concretas, basándose en datos de entrenamiento clasificados. La habilidad de las RNN para tomar en cuenta el contexto secuencial las convierte en más eficaces para captar las sutilezas del lenguaje y optimizar la exactitud del análisis de emociones en contraposición a los modelos que abordan cada palabra de manera autónoma [28].

3.1.6.2. Maquinas de soporte vectorial (SVM)

Las Máquinas de Soporte para Vectores (SVM) son algoritmos de aprendizaje supervisado empleados en la clasificación y la regresión. Dentro del análisis de emociones, las SVM tienen la capacidad de categorizar textos en diversas clases de emociones, tales como positivo, negativo o neutral, hallando el hiperplano ideal que divide los datos en diferentes categorías. Las SVM resultan especialmente útiles en espacios de alta dimensión, como los que se hallan frecuentemente en los datos de texto, donde cada palabra o n-grama puede ser visto como una característica. Generalmente, el procedimiento para utilizar SVM para la categorización de texto conlleva diversos pasos: preprocesamiento del texto, extracción de características (como TF-IDF, bag-of-words), entrenamiento del modelo SVM con datos etiquetados y evaluación del desempeño del modelo. Las SVM pueden emplear diversas funciones kernel (lineal, polinómica, RBF, entre otras) para gestionar datos que pueden ser separados linealmente y no linealmente. Las SVM han probado ser un enfoque sólido para el estudio de las emociones, frecuentemente alcanzando una alta exactitud en la categorización del sentimiento en distintas actividades [29].

3.1.6.3. Random Forest (RDF)

Random Forest (RDF) es un algoritmo de aprendizaje automático de conjunto empleado en tareas de clasificación y regresión. Opera creando varios árboles de decisión durante la formación y produciendo la clase que representa el modo de las clases (para la clasificación) o la predicción media (para la regresión) de los árboles individuales. En el estudio de las emociones, Random Forest puede emplearse para categorizar el sentimiento en textos, estudiando las palabras y sus frecuencias [29].

Los beneficios de emplear Random Forest abarcan su habilidad para gestionar grupos de datos grandes, su resistencia frente al sobreajuste y su habilidad para ofrecer estimaciones sobre la relevancia de las características. Para el estudio de emociones, se realiza un preprocesamiento del texto y se obtienen las características (como el bag-of-words, TF-IDF) con el fin de entrenar el modelo de Random Forest. El algoritmo genera diversos árboles de decisión, cada uno capacitado en un subgrupo aleatorio de los datos y un subgrupo aleatorio de las características. La predicción definitiva de la emoción se fundamenta en el voto predominante de las predicciones de todos los árboles [29]. Random Forest ha probado su eficacia en el estudio de las emociones, frecuentemente consiguiendo un buen desempeño en la categorización de las emociones en diversos grupos de datos [30].

3.1.6.4. Non-Negative Matrix Factorization (NMF)

La Factorización de Matrices No Negativas (NMF) es un método de álgebra lineal y un conjunto de algoritmos utilizados en el análisis multivariante. En este procedimiento, una matriz V se divide usualmente en dos matrices W y H , con la característica de que las tres matrices carecen de elementos negativos. Esta ausencia de negatividad facilita la revisión de las matrices generadas y, en usos como el tratamiento de espectrogramas de audio o la actividad muscular, la no negatividad es intrínseca

a los datos que se toman en cuenta. En el estudio de texto, la NMF puede emplearse para modelar temas, permitiendo identificar temas ocultos en un conjunto de documentos [31].

Dentro del estudio de las emociones, la NMF podría emplearse para detectar patrones de palabras vinculadas a emociones positivas o negativas. La NMF, al dividir la matriz de palabras por documento en dos matrices no negativas, puede descubrir los temas o elementos subyacentes que fomentan la emoción manifestada en el texto. Aunque la NMF no se considera un clasificador directo de sentimientos como el SVM o Naive Bayes, puede funcionar como un método eficaz para disminuir la dimensionalidad y identificar características para entender la estructura interna de los datos de sentimientos [31].

3.1.6.5. Métricas de evaluación

Las medidas de evaluación son esenciales para medir el desempeño de los modelos de análisis de emociones y categorización de texto [32]. Estas métricas ofrecen indicadores cuantitativos de cuán eficazmente un modelo está anticipando la emoción o categorizando el texto en las categorías adecuadas [33]. Las métricas habituales de evaluación comprenden:

- Accuracy: El vínculo entre la cantidad de predicciones acertadas y la cantidad total de estas [32].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

donde TP son verdaderos positivos, TN verdaderos negativos, FP falsos positivos y FN falsos negativos [33].

- Precisión: La correlación entre la cantidad de positivos verdaderos y la cantidad total de positivos estimados. Evalúa la precisión de las predicciones favorables [29].

$$\text{Precisión} = \frac{TP}{TP + FP}$$

- Recall (Sensibilidad): El vínculo entre la cantidad de positivos verdaderos y la cantidad total de positivos reales. Evalúa la habilidad del modelo para reconocer todas las ediciones positivas [29].

$$\text{Recall} = \frac{TP}{TP + FN}$$

- Índice F1 (F1-índice): El promedio armónico de la exactitud y el recall. Ofrece una evaluación balanceada del desempeño de un modelo, particularmente cuando las clases se encuentran desequilibradas [29].

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$$

- AUC (Zona bajo la curva ROC): Evalúa la habilidad del clasificador de diferenciar entre clases. Frecuentemente se emplea para dificultades de clasificación binaria [29].

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN}$$

donde la curva ROC representa TPR contra FPR para distintos umbrales.

La selección de la métrica de evaluación está determinada por la tarea concreta y los propósitos de la pesquisa. Por ejemplo, en un contexto en el que es crucial identificar todas las instancias positivas, el recall podría ser el indicador principal. En algunas situaciones, donde la precisión de las predicciones positivas es más relevante, la precisión podría ser más significativa. La calificación F1 resulta beneficiosa cuando se persigue un balance entre precisión y recall.

3.1.7. Tecnología

Se reconoció que realizar estos procesos sin tecnología sería casi imposible especialmente dadas las capacidades computacionales disponibles en la actualidad. Por esta razón se decidió por utilizar diversos servicios de GCP (Google Cloud Platform) para el desarrollo e implementación de este proyecto. Google como nube, ofrece una robusta gama de herramientas, particularmente en el ámbito del procesamiento de datos.

Además de los servicios de GCP seleccionados para esta implementación, el lenguaje Python desempeñó un papel fundamental en el desarrollo de la solución. Esta solución, basada en datos, fue diseñada para analizar, clasificar y agrupar las llamadas del Call Center de la universidad, tomando en cuenta el sentimiento detectado en cada una.

Estas herramientas proporcionan múltiples características como precisión potenciada por Inteligencia Artificial, funciones de análisis, compatibilidad lingüística e integración con otras herramientas de investigación [34]. Al seleccionar una herramienta, es necesario tener en cuenta aspectos como la exactitud, el costo, la rapidez y las demandas particulares de la investigación [35]. La presencia de servicios de transcripción, ya sean automatizados o humanos, ofrece a los investigadores alternativas fundamentadas en su presupuesto, restricciones temporales y necesidades de exactitud. La presencia de servicios de transcripción, ya sean automatizados o humanos, ofrece a los investigadores alternativas fundamentadas en su presupuesto, restricciones temporales y necesidades de exactitud. Las herramientas impulsadas por Inteligencia Artificial proporcionan rapidez y accesibilidad, mientras que la transcripción humana asegura una precisión elevada para proyectos de gran importancia. Los científicos tienen la posibilidad de seleccionar el procedimiento y los instrumentos de transcripción que se ajusten de manera óptima a su metodología de investigación y recursos. Frecuentemente, la decisión conlleva una balanza entre rapidez y costo (automatizado) y exactitud y gestión de audio complejo (humano).

3.1.7.1. GCP Google Cloud SPEECH-TO-TEXT

Google Cloud Speech-to-Text es un servicio fundamentado en inteligencia artificial que facilita la conversión de sonido en texto a través de una API sencilla de manejar. Es capaz de transcribir

archivos de sonido o audio en tiempo real y soporta más de 125 lenguajes. Este servicio emplea modelos avanzados de aprendizaje automático entrenados en millones de horas de datos sonoros para alcanzar una precisión elevada en el reconocimiento y la transcripción. Proporciona modelos previamente capacitados optimizados para diversas situaciones de uso, tales como control por voz, transcripción de llamadas telefónicas y transcripción de video. Los usuarios también tienen la posibilidad de personalizar los modelos para incrementar la exactitud de términos particulares o palabras poco comunes a través de la modificación del modelo. Google Cloud Speech-to-Text facilita la transmisión de sonido en directo y archivos de audio guardados en Google Cloud Storage [36]. Es un recurso útil para incorporar habilidades de reconocimiento de voz en diferentes aplicaciones y procesos de trabajo de investigación [37].

3.1.7.2. GCP Google Cloud Storage

Google Cloud Storage es una aplicación que permite guardar objetos en la nube de Google. Facilita el almacenamiento de información no estructurada, como archivos de audio, en recipientes conocidos como buckets. Cada cubo está vinculado a un proyecto y los proyectos pueden ser reunidos en una cuenta [38]. Cloud Storage proporciona diversas variantes de almacenamiento diseñadas para satisfacer diferentes requerimientos de acceso y costos. Es útil para diversos contextos, tales como el almacenamiento de conjuntos de datos para Inteligencia Artificial/ML y análisis, el almacenamiento de datos para archivado y recuperación de desastres, y la distribución de contenido a usuarios a nivel global. Los científicos tienen la posibilidad de emplear Google Cloud Storage para guardar de manera segura grandes cantidades de datos de audio transcritos y otros archivos vinculados con el estudio [39]. El servicio proporciona múltiples métodos para interactuar con la información, entre ellos, las bibliotecas del cliente, la herramienta de línea de comandos `gcloud` y la herramienta `gsutil` [40].

3.1.7.3. Google Colab

Google Colaboratory, también denominado Colab, es un ambiente de Jupyter Notebook alojado en la nube que facilita la creación y ejecución de código Python a través del navegador. Es una aplicación sin costo que no necesita configuración y ofrece acceso a recursos de computación como GPUs y TPUs sin un gasto extra. Colab es particularmente idóneo para el aprendizaje automático, la ciencia de los datos y la enseñanza. Los cuadernos de Colab permiten fusionar código ejecutable y texto enriquecido en un único documento, lo que los convierte en perfectos para registrar y difundir flujos de trabajo de investigación. Los investigadores pueden utilizar Colab para escribir y ejecutar código Python para tareas como el preprocesamiento de datos de audio transcritos, el análisis de sentimiento utilizando bibliotecas de aprendizaje automático y la visualización de resultados. La integración con Google Drive permite un fácil almacenamiento y acceso a los datos y cuadernos [41].

3.1.7.4. Python

Python es un lenguaje de programación versátil y de alto nivel que se utiliza ampliamente en el procesamiento del lenguaje natural (PNL) y el aprendizaje automático. Su sintaxis simple y su rica colección de bibliotecas lo convierten en una opción popular para tareas relacionadas con el análisis de sentimiento. Bibliotecas como NLTK (Natural Language Toolkit), spaCy, Transformers (de Hugging Face) y scikit-learn proporcionan una amplia gama de herramientas y algoritmos para el preprocesamiento de texto, la extracción de características, el modelado de sentimientos y la evaluación [42]. Python se puede utilizar en combinación con plataformas como Google Colab para desarrollar y ejecutar código para el análisis de sentimiento, aprovechando la potencia de las bibliotecas de Python y los recursos informáticos proporcionados por Colab [41].

3.2. Antecedentes

Buscando en la literatura referencias para el actual proyecto de investigación, se encontró lo siguiente:

En la Universidad Complutense de Madrid se realizó el desarrollo de un Sistema híbrido para el análisis de sentimiento en llamadas, en donde se ha desarrollado un proyecto que combina varias tecnologías avanzadas como son técnicas de Machine Learning y Deep Learning, aquí se procesan los datos de redes sociales y comercio electrónico para ofrecer experiencias personalizadas a los usuarios. Este proyecto se divide en dos partes principales: el Reconocimiento de Emociones en Audio y el Procesamiento de Lenguaje Natural (PLN). Además, se ha implementado la herramienta de Speaker Diarization, que segmenta el audio en función de los hablantes, permitiendo un análisis detallado de las emociones reflejadas. Estos métodos, aplicados conjuntamente, proporcionan un análisis exhaustivo de los datos, obteniendo resultados fiables y detallados sobre los sentimientos y emociones de los usuarios [43].

Este estudio proporciona una comprensión profunda de cómo integrar tecnologías avanzadas como Machine Learning y Deep Learning para el análisis de sentimiento en llamadas de Call Center. La combinación de técnicas de Reconocimiento de Emociones en Audio y Procesamiento de Lenguaje Natural (PLN) puede ser muy útil para la actual investigación, ya que ambos enfoques pueden ser adaptados para analizar y mejorar la calidad de las llamadas. Además, la implementación de herramientas como Speaker Diarization puede ayudar en la segmentación de audio por hablantes, facilitando un análisis más detallado y preciso de las interacciones en las llamadas del Call Center.

Continuando, en España se realizó un estudio denominado: “Agent Productivity Measurement in Call Center Using Machine Learning”, este estudio se centra en la medición de la productividad de agentes en centros de llamadas mediante el uso de aprendizaje automático. Este estudio propone un sistema que utiliza herramientas de análisis de sentimiento y procesamiento de lenguaje natural (NLTK) para evaluar la productividad de representantes de servicio al cliente en centros de llamadas inmobiliarios en Egipto. El estudio detalla el proceso de construcción de un sistema de evaluación en árabe, incluyendo la transcripción de llamadas, la extracción de características, el entrenamiento de modelos y el análisis de resultados. Se utilizó un corpus de 7 horas de conversaciones reales

y se aplicaron técnicas de reconocimiento de voz para segmentar y transcribir las llamadas. El análisis de sentimiento se empleó para identificar características productivas y no productivas en el discurso de los agentes, utilizando un clasificador Naïve Bayes. Los resultados mostraron una precisión del 67 % en la clasificación de llamadas como productivas o no productivas, destacando la importancia de un corpus más grande para mejorar la precisión [44]. Este estudio ofrece un marco de referencia para medir la productividad de los agentes de un Call Center utilizando técnicas de Machine Learning y análisis de sentimiento. La metodología de transcripción de llamadas y la posterior extracción de características son relevantes para la actual investigación, especialmente en la fase de preprocesamiento de datos. Además, la utilización de un clasificador Naïve Bayes para identificar características productivas y no productivas puede ser adaptada para clasificar la calidad de las llamadas. Los desafíos y soluciones presentados en cuanto a la necesidad de un corpus grande también son importantes para asegurar la precisión de los modelos de clustering.

En Chile, específicamente en Santiago de Chile se realizó un estudio denominado: "Predicting Self-reported Customer Satisfaction of Interactions with a Corporate Call Center" en donde se describe un sistema desarrollado para predecir la satisfacción del cliente tras llamadas telefónicas en un centro de atención al cliente una compañía de seguros estadounidense. El sistema utiliza señales de múltiples fuentes de datos heterogéneas, como transcripciones de llamadas, metadatos de 29 llamadas, perfiles de clientes e información de pólizas de seguros. Los desafíos del modelado incluyen la naturaleza ordinal, subjetiva y a menudo sesgada de las puntuaciones de las encuestas. Para abordar estos problemas, el sistema emplea un flujo de trabajo que consiste en: primero, un modelo de clasificación se entrena con la fusión de datos de llamadas y luego, una función de ajuste convolucional se optimiza para mapear las puntuaciones de clasificación a las puntuaciones de satisfacción de las encuestas. Este enfoque produce predicciones más precisas que las técnicas estándar de regresión y clasificación. El sistema permite identificar a los clientes insatisfechos de manera oportuna, facilitando intervenciones significativas y mejorando la experiencia del cliente y el rendimiento del personal del centro de atención. El código fuente y los datos están disponibles en este repositorio de GitHub: <https://github.com/cyberyu/ecml2017> [45].

Este trabajo proporciona un enfoque estructurado para predecir la satisfacción del cliente mediante el análisis de transcripciones de llamadas y otros metadatos. Los desafíos relacionados con la naturaleza subjetiva y sesgada de las puntuaciones de satisfacción del cliente son relevantes para la actual investigación, ya que destacan la importancia de tener un enfoque robusto para el modelado de datos. La metodología de fusionar múltiples fuentes de datos y optimizar una función de ajuste convolucional puede inspirar el desarrollo de técnicas avanzadas de clustering para clasificar la calidad de las llamadas. Además, la disponibilidad del código fuente y los datos en un repositorio público puede servir como recurso valioso para comparar y validar mis propios resultados.

En la Universidad de Educación, Winneba se realizó un análisis de sentimiento para evaluar respuestas cualitativas de estudiantes en el contexto educativo. Se enfoca en implementar algoritmos de aprendizaje automático para predecir y clasificar los sentimientos expresados en los comentarios de los estudiantes sobre las lecciones recibidas. Se utilizó una metodología de recolección de datos en donde se recopilieron 280 comentarios de estudiantes de la Universidad de Educación, Winneba. Luego se continuó con el preprocesamiento de texto, aquí se limpiaron los datos eliminando valores

faltantes y términos no identificados. Se tokenizaron las oraciones usando la técnica de "String to Word Vector" en Weka. Teniendo en cuenta lo anterior se realizará un modelado de sentimientos utilizando cuatro algoritmos de aprendizaje supervisado: Naïve Bayes (NB), Support Vector Machine (SVM), Árbol de Decisión J48 (DT) y Bosque Aleatorio (RF) en donde se entrenaron y probaron los modelos usando la técnica de validación cruzada k-fold. De la aplicación de estos modelos se obtuvieron los siguientes resultados: El algoritmo SVM obtuvo el mejor desempeño con una precisión del 63.79%. Los comentarios se clasificaron en positivos y negativos y se subcategorizaron en 'Excelente', 'Bueno' y 'Pobre' [46].

Este estudio aporta un enfoque práctico para el uso de algoritmos de aprendizaje automático en la clasificación de sentimientos expresados en texto, lo cual es directamente aplicable al análisis de transcripciones de llamadas en un Call Center. La metodología de preprocesamiento de texto y tokenización puede ser adaptada para limpiar y preparar los datos de las transcripciones de llamadas de la actual investigación. Además, la comparación de diferentes algoritmos supervisados (Naïve Bayes, SVM, Árbol de Decisión, Bosque Aleatorio) y la identificación de SVM como el modelo más preciso proporciona una base sólida para seleccionar y probar varios modelos de clustering en el análisis de la calidad de llamadas.

Preparación de los datos

En este capítulo dedicado a la preparación de los datos, es importante destacar el desarrollo y la ejecución del proceso de transcripción de los audios. Una vez convertidos en texto, se lleva a cabo la limpieza y normalización del contenido. Posteriormente, se prueban diversas técnicas de vectorización, las cuales se combinan con diferentes modelos para identificar la combinación que ofrece los mejores resultados. Con esta selección óptima, es posible proceder al etiquetado completo del conjunto de datos.

4.1. Prueba de concepto de los servicios de transcripción

Teniendo en cuenta la necesidad de realizar la transcripción de los audios y dando cumplimiento al primer objetivo que hace referencia al preprocesamiento de los datos de las grabaciones de las llamadas, se llevó a cabo un estudio sobre que servicios son capaces de construir una transcripción de calidad y eficiente, entre muchos servicios existentes hoy en día, por características tanto funcionales como de procesos elegimos realizar unas pruebas de concepto sobre 3 de estos, los cuales encontramos a continuación.

4.1.1. Librería Whisper

Whisper transcribe audios mediante inteligencia artificial, contiene un modelo entrenado con un volumen de alrededor de 680.000 horas de audio.

Actualmente Whisper tiene una tasa de error de menos del 5% en el proceso de transcripción en idioma español, dentro de sus ventajas destacables, es que automáticamente detecta cambios de idioma durante la conversación, detecta pausas transcribiéndolas con signos de puntuación dependiendo de la duración.

Aun cuando Whisper tiene estas características positivas, en el proceso de prueba de concepto el cual fue hecho con 5 audios de diferentes tamaños, no funcionó bien, algunos audios cambiaba las palabras y en otros se quedaba pegado en unas palabras o frases, es decir las repetía sin esto coincidir con el audio, la prueba luego de la implementación técnica que se hacía consistió en validar la transcripción versus el contenido del audio [47].

Teniendo en cuenta este resultado llegamos a la conclusión que no era la herramienta que requeríamos para nuestro proyecto.

4.1.2. AWS Transcribe

Amazon transcribe es un servicio de voz de reconocimiento automático (ASR) totalmente provisionado. Esta impulsado por un modelo de voz propietario de AWS que ofrece transcripciones de alta precisión, este proceso al igual que Whisper tiene un modelo entrenado con millones de horas de audio en diferentes idiomas.

Tiene la capacidad de realizar el proceso de transcripción en mas de 100 idiomas, incluyendo la generación de los signos de puntuación luego de identificar las pausas en el audio, también tienen detección automática de idioma, esto en caso que existan audios con más de un idioma [48].

Esta prueba de concepto se efectuó con los mismo 5 audios usados en el servicio anterior, pero para este caso no se realizaron desarrollos. La prueba se realizó directamente sobre la consola de AWS, donde luego de comparar la transcripción realizada con el audio, tenía algunas deficiencias, como la no identificación de palabras, es decir no las transcribía o cambiaba la palabra. En esta fase, se evidenció que, a pesar de que los dos servicios previamente evaluados eran competentes, no alcanzaban un rendimiento apropiado debido a la mala calidad de algunas comunicaciones. En conclusión, se determinó que era necesario continuar con las pruebas

4.1.3. GCP Speech To Text

Speech to text, es el servicio de transcripción de Google el cual es un modelo básico de Google Cloud para la voz entrenado con millones de horas de datos de audio y miles de millones de frases de texto. Esto contrasta con las técnicas tradicionales de reconocimiento de voz, que se centran en grandes cantidades de datos supervisados específicos de cada idioma.

Este servicio cuenta con un modelo denominado por Google como Chirp, el codificador no solo fue entrenado como se mencionó anteriormente, luego de entrenado con este volumen de horas de audio y los textos, se implementó un ajuste para la transcripción de cada idioma específico, con pequeñas cantidades de datos supervisados, es decir ajustes manuales, los cuales ayudaron a su eficiencia, este modelo tiene la capacidad de hacer transcripciones en 125 idiomas [49].

Para la prueba de concepto se usaron lo mismos 5 audios y mediante la consola de GCP, se encontró una ganancia significativa comparando con los dos servicios anteriores, es claro que esto no es 100% perfecto pero se logró una asertividad bastante relevante de cara a nuestro proceso y necesidad, esto pese a que el nivel de la voz de algunos audios es muy bajo o de mala calidad.[50].

Teniendo en cuenta esto se decidió que la mejor herramienta ó servicio para el proceso de transcripción para el proyecto es Speech to text de GCP (Google).

4.2. Recolección y cargue de los datos

De cara a obtener los archivos audio fuente para el proceso, se realizó una gestión mediante nuestros tutores, los cuales realizaron la gestión interna en la universidad para poder tener acceso a estos datos.

Luego de esta gestión anteriormente mencionada, el proceso de recolección de los archivos de audio del Call Center fue mediante un archivo en drive proporcionado por la universidad, fueron

alrededor de 897 grabaciones de llamadas de diferentes tamaños, estas fueron descargadas y almacenadas en una máquina de uno de los miembros del proyecto.

Ya teniendo disponibles los datos de las llamadas, se procedió a cargarlos a un bucket del servicio Cloud Storage de GCP, este proceso de cargue se implementó de manera manual usando la consola de GCP, mediante una funcionalidad que tiene el mismo servicio, ya con esto se obtiene los datos crudos para iniciar el procesamiento de los mismos.

4.3. Procesamiento de los datos

4.3.1. Transcripción

Luego de cargar y tener en el Cloud Storage disponible la totalidad de los audios se inicia el proceso de transcripción, para este fin se desarrolló en Google Colab, un código Python que invoca el servicio de transcripción Speech To Text, incluyendo dentro del mismo código la parametrización del servicio usada en el prueba de concepto, estos parámetros están relacionados con el tipo de transcripción, el idioma y otros parámetros relevantes para contar con una transcripción idónea. Este proceso toma el o los archivos desde el Cloud Storage, los procesa y el resultado de la transcripción lo almacena en el drive de Google, en una estructura donde queda en una columna, el ID (Nombre del archivo) y en otra columna el texto con la transcripción.

Ya teniendo el desarrollo implementado, se procedió a ejecutar el script desarrollado y luego de aproximadamente 1 hora de procesamiento se obtuvo el 100 % de los audios transcritos en el drive.

4.3.2. Etiquetado manual de los datos

Ya teniendo la transcripción completa de los audios, se decidió que el total de audios a etiquetar serían 312. Esta actividad de etiquetado, se inicio escogiendo 10 audios, con los cuales cada uno de los tres integrantes del equipo del proyecto efectuó el proceso de etiquetado. En este momento, se encontró un desafío significativo: cada miembro tenía una percepción distinta de lo que podía ser un sentimiento positivo, negativo o neutral. Esta circunstancia condujo a tomar la determinación de que únicamente un integrante del equipo se ocupara del proceso de etiquetado.

Con esta definición, se eligió una primera muestra aleatoria de 100 audios para iniciar a etiquetar. No obstante, conforme se progresaba, se observó que no existía un sentimiento positivo o negativo muy marcado, esto en particular en el caso del sentimiento negativo. Esto se genera por dos motivos: en primer lugar, la mayoría de las llamadas eran salientes, es decir llamadas en las que un asesor del call center, se comunicaba con alguien que había manifestado interés en un programa a través del sitio web de la universidad; en segundo lugar, estas conversaciones siempre se desarrollaban en un clima de cordialidad y respeto. Esta circunstancia motivó la toma de otra resolución, que consistió en enmarcar las emociones de las llamadas en un grupo diverso de atributos, los cuales se describen a continuación.

- Positivo: Son llamadas donde el usuario obtiene de manera satisfactoria la información y luego de recibirla se nota conforme y con cierto interés.

- **Negativa:** Son llamadas donde el usuario luego de recibir la información, no queda conforme, no encuentra la respuesta que buscaba o deja de manera explícita su NO interés en el programa o en la universidad.
- **Neutral:** Son todas estas llamadas inconclusas, que por temas de comunicación se cortan o quedan sin una definición clara. Después de llegar a un consenso sobre esta definición, se llevó a cabo el etiquetado de los 312 audios. Este procedimiento facilitó la adquisición del recurso esencial para un entrenamiento adecuado y exitoso del modelo de clasificación.

La descripción del procedimiento de etiquetado, previamente detallada, aportó de manera significativa a la consecución de resultados excepcionales, en concordancia con los objetivos establecidos en este proyecto.

4.3.3. Preprocesamiento del texto

Ya transcritos la totalidad (897 grabaciones) de los audios y un total de 312 registros con el etiquetado manual se inicia la fase de procesamiento de texto, la cual es una de las fases más importantes en cualquier tipo de implementación de ciencia de datos, teniendo en cuenta que si no tienen los datos de una manera limpia, estandarizada y/o ajustada a la necesidad que se presenta, muy seguramente el resultado del modelo no sería el esperado, siendo datos de mala calidad o con unos indicadores bastantes bajos.

En esta situación, considerando que los datos a tratar eran textos transcritos de llamadas, las estrategias apropiadas se enfocaron en estandarizar todos los caracteres a minúsculas y suprimir artículos y palabras vacías (StopWords), entre otros cambios. En consecuencia, se logró un preprocesamiento ideal que se adecuó de manera perfecta a las exigencias del proyecto.

Para ilustrar de forma más efectiva los hallazgos, se aplicó el siguiente ejemplo, extraído de la base de datos de audios transcritos. Se llevó a cabo con el objetivo de observar la progresión del texto tras cada fase.

Texto original

aló Buenos días por favor la joven Juanita Arenas señora mucho gusto Mi nombre es Valentina Me estoy comunicando a la universidad jariana Cali cómo está Muy bien bueno el motivo de mi llamada es porque ha presentado interés por el programa de medicina y el día de mañana finalizamos proceso de inscripción deseamos conocer si vas a iniciar el proceso aló Hola Me escuchas aló Hola hola te ahí se escucha bien Hola buenas aló aló

4.3.3.1. Limpieza del texto

En esta sección el objetivo principal es eliminar el ruido y obtener texto que sea más homogéneo, conservando el contenido importante para su respectivo análisis, en este proceso se realizó lo siguiente:

- **Reemplazo de jergas** El procedimiento de sustituir jergas implicó reemplazar palabras o expresiones características de los regionalismos del idioma colombiano por un significado estándar. Para llevar a cabo este proceso se elaboró lo siguiente:

- Se cargó jergas desde un archivo que contiene jergas comunes y su significado correspondiente.
 - Se reemplazaron estas jergas en el texto utilizando su significado.
- **Conversión a minúsculas** Se sustituyeron caracteres para que el texto estuviese totalmente en minúsculas, lo que contribuyó a unificarlo, optimizó la calidad del preprocesamiento y facilitó que el modelo adquiriera un mejor aprendizaje con menos ruido. Esto permitió reducir la dimensionalidad y evitar errores semánticos, lo cual a su vez simplificó la elaboración de un texto homogéneo.

Teniendo en cuenta lo anterior se precedió a cambiar todo el texto a minúsculas.

- **Eliminación de tildes y caracteres especiales** Este procedimiento de supresión de tildes fue de gran relevancia, particularmente para el idioma español, donde se encontraba el corpus. Esto ayudó a conseguir un texto estandarizado o unificado, ya que el proceso de transcripción no incorporaba correctamente las tildes del idioma en todas las situaciones. Los caracteres especiales suelen causar ruido y confusión en este tipo de procesos; por esta razón, se optó por suprimirlos de todo el corpus.

Para este fin se utilizó la función `unidecode()`, dejando el texto según lo deseado.

- **Eliminación de signos de puntuación, números y símbolos**

Eliminar signos de puntuación y números en el preprocesamiento de texto es una práctica común en el análisis de sentimiento, sobre todo cuando el enfoque está en el sentido semántico de las palabras. Sin embargo, es importante tener en cuenta que esto no siempre es necesario, pero para el proyecto si lo era, esto teniendo en cuenta que se estaba procesando un texto resultante de un proceso de transcripción, mas no de algo escrito por un usuario.

4.3.3.2. Normalización

Una vez que el texto fue limpiado, es decir, libre de caracteres especiales, espacios innecesarios, signos de puntuación irrelevantes y otros elementos no lingüísticos, se llevó a cabo un proceso de estandarización. Este proceso tuvo como objetivo principal normalizar la estructura léxica del texto, reduciendo la variabilidad gramatical para facilitar su análisis posterior. La estandarización permitió homogeneizar formas distintas de una misma palabra o expresión, y garantizar una representación coherente del lenguaje en el corpus.

Posteriormente, se aplicaron las siguientes técnicas de procesamiento:

- **Tokenizar:** Separa el texto en palabras individuales.
- **Quita stopwords:** Elimina palabras comunes que no tienen importancia.
- **Lematiza:** Reduce las palabras a su forma base.

Continuando el texto limpio y normalizado se transforma en una matriz numérica utilizando un vectorizador (`vectorizer.transform`), que transforma cada texto en un vector de números.

4.3.4. Esquema del procesamiento

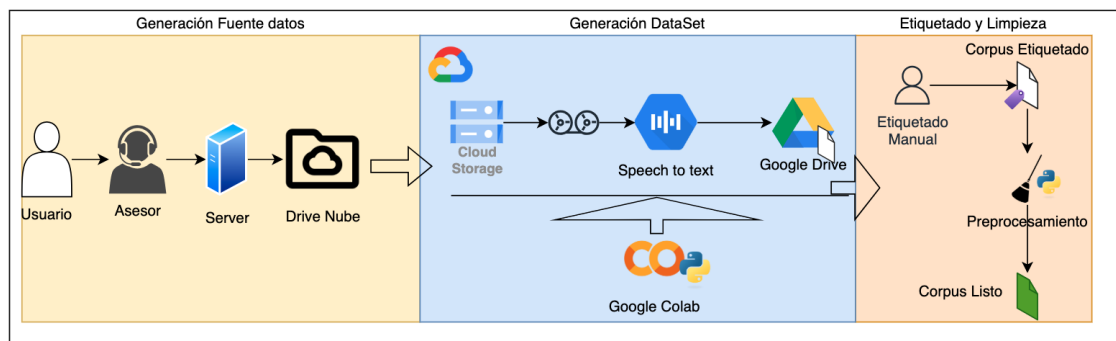


Figura 4.1: Esquema Procesamiento

Teniendo en cuenta lo mencionado en este capítulo, a continuación se detalla el esquema del proceso sobre el cual se fundamentó el procesamiento y alistamiento del Corpus de las llamadas, el cual se ilustra en la Figura 4.1: Esquema del Procesamiento. El proceso inicia con la llamada entrante o saliente de un usuario, atendida o gestionada por un asesor del call center de la universidad, estas llamadas son almacenadas en un Servidor (Server). Posteriormente, estos datos fueron compartidos mediante un Drive en nube, para el uso y objetivo del proyecto.

Estos datos fueron cargados a Cloud Storage de GCP (Google Cloud Platform) mediante un proceso soportado por el mismo servicio. El servicio Speech-to-Tex accede a los audios alojados en el cloud Storage con el fin de realizar el proceso de transcripción, el resultado de esto es almacenado en el Google Drive, en donde se crea un archivo CSV con la información, esta fase de acceder al repositorio en mención, transcribir los audios y almacenarlos es orquestada por un desarrollo hecho en Python y soportado en Google Colab.

Este DataSet resultante con 897 audios transcritos, es usado para el proceso de etiquetado que se detalla en el capítulo 4.2.1, este proceso se aplica para 312 audios, luego de esto se realiza el preprocesamiento de los registros a usar para el entrenamiento del modelo, teniendo como resultado el corpus listo para la fase de modelado, este es simbolizado en el esquema por el archivo de color verde.

Aquí es importante detallar que para la construcción y evaluación de los modelos predictivos, el corpus fue dividido estratégicamente en dos subconjuntos: un conjunto de entrenamiento (train), que comprende el 80% de los datos, y un conjunto de prueba (test), que abarca el 20% restante. Esta partición se organizó siguiendo el metodo traint test split.

4.3.5. Vectorización de los datos

Teniendo en cuenta la importancia del proceso de vectorización el cual apoya en transformar datos textuales (palabras, frases, documentos) en vectores numéricos para que puedan ser procesados por algoritmos de aprendizaje, se da inicio a una prueba de concepto utilizando tres métodos de

vectorización. Estas técnicas serán evaluadas junto con los modelos de clasificación que se presentarán en capítulos posteriores. Además, para cada método se incluirá una descripción y un ejemplo práctico utilizando el siguiente texto, el cual forma parte de la base de datos de llamadas después del preprocesamiento:

*alo buenos dia favor joven juanita arena senora gusto nombre valentina comunicando univ-
ersidad jariana cali bien bueno motivo llamada presentado interes programa medicina dia manana
finalizamos proceso inscripcion deseamos conocer si va iniciar proceso alo hola escuchas alo hola
hola ahi escucha bien hola buenas alo alo*

4.3.5.1. One Hot Encoding

Inicialmente, se utilizó el método Bag of Words con CountVectorizer, el cual representó cada documento como un vector de conteos de palabras. Los hiperparámetros utilizados fueron los siguientes:

- `ngram_range=(1, 1)` : solo unigramas.
- `min_df=1` : palabra debe aparecer en al menos un documento.
- `max_df=1.0` : se aceptan palabras que aparecen en todos los documentos.
- `lowercase=True` : convierte a minúsculas.
- `stop_words=None` : no elimina stopwords (aunque fueron eliminadas previamente en el pipeline).
- `token_pattern =` ignora tokens de menos de 2 letras

A continuación, se muestra un fragmento correspondiente a las primeras 10 posiciones de un vector perteneciente a una matriz resultante de la vectorización mediante OneHotEncoder: [1. 0. 1. 0. 1. 0. 1. 0. 1. 0.]

4.3.5.2. Term Frequency-Inverse Document Frequency (TF-IDF)

La técnica de vectorización se emplea para ponderar los términos según su importancia relativa en el corpus. Los hiperparámetros que se utilizaron son los siguientes:

- `use_idf=True` : activa el componente IDF.
- `smooth_idf=True` : suaviza el IDF para evitar división por cero.
- `sublinear_tf=False` : no aplica logaritmo a las frecuencias.
- `norm='l2'` : normaliza los vectores resultantes.
- `ngram_range=(1, 1)` : unigramas.

- `min_df=1`, `max_df=1.0` : sin filtrado de frecuencia.
- `stop_words=None` : (ya se eliminaron antes en el pipeline).

A continuación, se presenta un fragmento de los valores obtenidos al aplicar TF-IDF con los hiperparámetros seleccionados. El vector mostrado corresponde a una de las filas de la matriz resultante:

```
[0.28932015 0.0553779 0.09164106 0.09490683 0.23376964 0.26367212 0.26367212 0.0768436 0.0768436 0.12657037]
```

4.3.5.3. Word2Vec

Continuando se utilizó Word2Vec con gensim para generar vectores semánticos. Luego, cada documento es representado por el promedio de los vectores de sus palabras.

Los hiperparámetros que se utilizaron son los siguientes:

- `vector_size = 100` : Dimensión del vector por palabra.
- `window = 5` : Tamaño de contexto (ventana) para el entrenamiento.
- `min_count = 1` : Palabras deben aparecer al menos una vez.
- `workers = 4` : Número de hilos de CPU para entrenar.

En este caso, se muestra un fragmento correspondiente a las primeras 10 dimensiones de un vector de características generado mediante Word2Vec, como parte de la matriz de representaciones vectoriales:

```
[-0.34083822 0.14440799 0.13094856 0.23031265 0.32458708 -0.5641565 0.33679602 0.94886845 -0.5827702 -0.28023753]
```

4.4. Selección de modelo y etiquetado

Aquí se exploraron los diferentes modelos, combinándolos con los diferentes métodos de vectorización vistos anteriormente con el objetivo de tener el corpus 100% etiquetado, el cual se convierte en el insumo principal para entrenar el modelo final.

4.4.1. Modelo de clasificación supervisada

Se inicia con el desarrollo del segundo objetivo el cual busca seleccionar el modelo supervisado que obtenga el mejor desempeño según las métricas de evaluación, y usarlo para etiquetar todo el conjunto de datos en las categorías de sentimiento: positivo, negativo y neutral, es decir al finalizar esta sección tendremos todo el corpus etiquetado.

4.4.1.1. Prueba de concepto modelo de clasificación

Se definió el uso de tres modelos de aprendizaje automático: Máquinas de Soporte Vectorial (SVM), Bosques Aleatorios (Random Forest, RDF) y Redes Neuronales Recurrentes (RNN), en combinación con diferentes técnicas de vectorización de texto, como Word2VecTransformer, TF-IDF y One-Hot Encoding. Se decidió aplicar cada modelo a todas las técnicas de vectorización para evaluar su desempeño en el análisis de sentimiento en textos transcritos.

El proceso comenzó con la importación de las bibliotecas necesarias, incluyendo pandas, nltk, scikit-learn y TensorFlow. Para la limpieza y transformación del texto, se definieron clases personalizadas, como ReemplazarJergas, encargada de sustituir términos coloquiales; LimpiarTexto, que normaliza los textos eliminando acentos y caracteres especiales; y TokenizarYLematizar, que aplica tokenización y lematización en español.

Posteriormente, se implementaron técnicas de vectorización como Word2VecTransformer, OneHotEncoder y TF-IDF para convertir los textos en representaciones numéricas. El preprocesamiento de los datos se elaboró mediante un pipeline que integró estos transformadores, seguido de la vectorización con CountVectorizer.

Una vez preprocesados los datos, se entrenaron y evaluaron tres modelos: una SVM con un kernel lineal, un Bosque Aleatorio con 100 árboles y una RNN implementada en TensorFlow. Los datos fueron divididos en conjuntos de entrenamiento y prueba y la evaluación del desempeño se realizó utilizando la métrica F1-score por clase. Finalmente, se imprimieron los resultados de cada modelo para comparar su rendimiento en la clasificación de sentimientos en llamadas del Call Center.

Esto se observa mas a detalle a continuación para cada uno de los modelos:

- **SVM:** El modelo SVM fue construido con un **kernel lineal**, este es mas adecuado para problemas de clasificación multiclase, como es nuestro caso, también obedece a su efectividad en tareas de clasificación de texto, donde las dimensiones suelen ser altas. Este modelo busca encontrar el hiperplano óptimo que maximiza la separación entre las clases, también se seleccionó un **random_state** de 42 el cual permite controlar la aleatoriedad en cada ejecución del modelo.

En este caso, se entrenó utilizando los vectores generados a partir de técnicas de vectorización textual. Los textos fueron vectorizados utilizando diferentes técnicas:

- En la prueba con CountVectorizer y OneHotEncoder, los textos fueron representados como vectores binarios de alta dimensión, lo cual refuerza el comportamiento lineal del modelo.
 - En la prueba con TF-IDF, se logró una representación ponderada que mejora la discriminación de palabras clave en los textos, beneficiando al SVM en tareas de clasificación.
 - En el enfoque con Word2Vec, cada texto se representó como el promedio de los vectores semánticos de sus palabras, permitiendo capturar cierta relación semántica.
- **RDF:**El modelo RDF se construyo con un conjunto de 100 **árboles de decisión**, estos fueron entrenados con el método de bagging, esto permite mejorar la precisión del modelo al

reducir el sobre ajuste, esto permite la variabilidad en conjuntos de datos y sus características, adicionalmente se utilizó un `random_state` de 42, el cual permitió que el modelo sea reproducible.

Esto se aplicó sobre los datos vectorizados con las diferentes tecnicas nombradas. Cada árbol en el bosque se entrenó sobre representaciones vectorizadas del texto, las cuales variaron en función del experimento:

- En el caso de CountVectorizer + OneHot, el RDF aprovechó la granularidad binaria de las palabras.
 - Con TF-IDF, la variabilidad de importancia de las palabras permitió que algunos árboles fueran más sensibles a términos relevantes por clase.
 - En Word2Vec, aunque RDF no capta relaciones temporales ni semánticas profundas, sí se benefició de la compacidad de la representación vectorial promedio por texto.
- **RNN**: se construyó con una arquitectura secuencia, es decir con capas densas, una función de activación ReLu y una capa softmax, esto teniendo en cuenta que fue una clasificación multi-clase. El modelo RNN se entrenó con una función de pérdida `sparse_categorical_crossentropy` y el optimizador Adam, esto busca capturar patrones de los textos vectorizados. Los hiperparámetros utilizados en este modelo son los siguientes:

Hiperparámetro	Qué controla
Dense(128/64)	Número de neuronas por capa oculta. Controla la capacidad de aprendizaje del modelo.
input_dim	Tamaño del vector de entrada (número de características del dataset).
activation='relu'	Introduce no linealidad en la red para capturar relaciones complejas.
Dropout(0.5)	Técnica de regularización: apaga el 50% de las neuronas durante el entrenamiento para evitar sobreajuste.
activation='softmax'	Convierte la salida de la red en una distribución de probabilidad sobre las clases.
optimizer='adam'	Método de optimización que ajusta los pesos combinando momentum y tasa de aprendizaje adaptativa.
loss='sparse_categorical_crossentropy'	Función objetivo para clasificación multiclase con etiquetas como enteros.
metrics=['accuracy']	Métrica usada para evaluar el rendimiento del modelo (proporción de predicciones correctas).

Cuadro 4.1: Resumen de los hiperparámetros del modelo de red neuronal

En este caso la entrada fue una representación vectorial fija:

- En CountVectorizer + OneHot, los vectores de entrada eran extremadamente dispersos, lo que puede dificultar el aprendizaje profundo.
- Con TF-IDF, la red pudo captar mejor la importancia relativa de las palabras en cada clase.
- En el enfoque con Word2Vec, las entradas vectorizadas eran densas y más adecuadas para redes neuronales, permitiendo al modelo identificar patrones semánticos más complejos.

4.4.1.2. Resultados prueba de concepto modelos de clasificación

Teniendo en cuenta las pruebas de concepto realizadas con los diferentes modelos y combinándolas con los diferentes métodos de vectorización encontramos los siguientes resultados de F1:

Cuadro 4.2: Resultados del F1-score por modelo y técnica de vectorización

MODELO	ETIQUETAS	TÉCNICAS DE VECTORIZACIÓN		
		WordToVect	TFIDF	OneHotEncoding
SVM	Negativo	0.48	0.64	0.55
	Neutral	0.00	0.63	0.76
	Positivo	0.78	0.84	0.74
RDF	Negativo	0.48	0.56	0.66
	Neutral	0.38	0.42	0.52
	Positivo	0.78	0.84	0.83
RNN	Negativo	0.54	0.66	0.53
	Neutral	0.00	0.68	0.00
	Positivo	0.82	0.82	0.75

Se seleccionó el modelo de clasificación SVM con el proceso de vectorización OneHotEncoding en donde se evidencia unos mejores resultados respecto a otras combinaciones, tal como se representa en la anterior tabla.

De este modelo se mejoraron los hiperparámetros y se obtuvo lo siguiente:

■ SVM con ONEHOTENCODING:

Para optimizar el desempeño del modelo Máquina de Soporte Vectorial (SVM), se efectuó una búsqueda de hiperparámetros utilizando GridSearchCV, evaluando combinaciones de:

- C (0.1, 1, 10).
- gamma (scale, auto).
- kernel (linear, rbf).

La mejor configuración encontrada fue:

- C=10.

- `gamma='scale'`.
- `kernel='rbf'`.

Se utilizó esta parametrización para ajustar el modelo final. Posteriormente, se realizaron predicciones sobre el conjunto de prueba y se calculó el F1-score por clase. Los resultados indicaron un F1-score de 0.74, con un accuracy del 74.63%. Se generó un reporte de clasificación, mostrando que la clase Positivo obtuvo el mejor desempeño (F1-score = 0.83), mientras que la clase Neutral tuvo un rendimiento menor (F1-score = 0.62). Adicionalmente, se presentó la matriz de confusión, permitiendo analizar los errores de clasificación entre categorías. También se generó la siguiente curva de ROC:

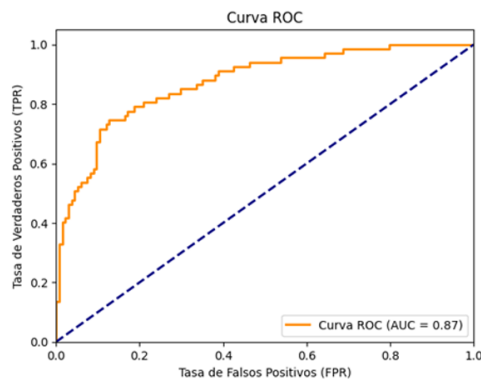


Figura 4.2: Curva de ROC

En donde se obtuvo un AUC que mide la capacidad del modelo para distinguir entre clases. En este caso, un AUC de 0.87 indica un buen desempeño del modelo.

A continuación, se indica la matriz de confusión del modelo:

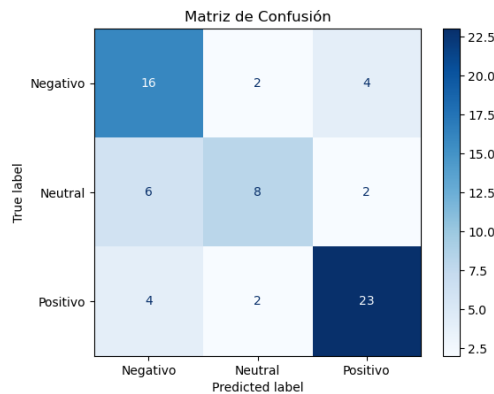


Figura 4.3: Matriz de confusión SVM.

De esta matriz se puede indicar lo siguiente:

■ **Casos Correctamente Clasificados**

- 16 casos negativos correctamente clasificados como negativos.
- 8 casos neutrales correctamente clasificados como neutrales.
- 23 casos positivos correctamente clasificados como positivos.

Total de aciertos: $16 + 8 + 23 = 47$

■ **Errores de Clasificación:**

- 2 negativos fueron clasificados como neutrales.
- 4 negativos fueron clasificados como positivos.
- 6 neutrales fueron clasificados como negativos.
- 2 neutrales fueron clasificados como positivos.
- 4 positivos fueron clasificados como negativos.
- 2 positivos fueron clasificados como neutrales.

4.4.2. Etiquetado de los datos

Teniendo en cuenta lo anterior se procede a seleccionar el modelo de SVM previamente entrenado para etiquetar la totalidad del dataset de audios transcritos.

- Se cargaron los datos sin etiquetas.
- Se aplicó todo el preprocesamiento que con anterioridad se mencionó.
- Se aplicó la vectorización y transformación de los datos en una representación numérica.
- Los datos se utilizaron como entrada para el modelo SVM entrenado, el cual generó las predicciones de las etiquetas.
- Se crea un archivo CSV con todos los datos y sus etiquetas correspondientes.

Modelado y Evaluación

En este capítulo detallaremos el proceso de entrenamiento del modelo seleccionado ya con el corpus etiquetado al 100 % y evaluaremos los resultados del mismo, adicionalmente, se procede con el desarrollo de un modelo no supervisado orientado a identificar patrones y temáticas recurrentes dentro de los textos previamente clasificados. Para ello, se agruparán los textos según su polaridad (positiva, negativa y neutral), y se aplicaron técnicas de análisis temático como el Non-Negative Matrix Factorization.

5.1. Entrenamiento del modelo SVM

Con el dataset etiquetado se procedió a entrenar el modelo SVM con el metodo de vectorización OneHotEncoding, dicha combinación fue la que arrojó los mejores resultados y fue usada para el etiquetado total de los datos (897 registros), de lo cual el 20 % de los datos se reservaron para el conjunto de prueba, y el 80 % para entrenamiento.

5.1.1. Resultados del entrenamiento del modelo SVM

Luego de la ejecución del proceso se detalla el resultado de la evaluación del modelo SVM con el dataset anteriormente mencionado, estos resultados se ven a continuación:

- **Curva de ROC:** Como se muestra en la Figura 5.1, el modelo mostró buenos resultados, teniendo mayor capacidad de aprendizaje con respecto a las clases 1 (Neutral) y 2 (Positivo), cuyas curvas ROC están alejadas de la diagonal principal, esto refleja una alta capacidad de discriminación del modelo entre las distintas clases. El modelo diferencia entre las categorías Negativo (Clase 0), Neutral (Clase 1) y Positivo (Clase 2). También, los valores del AUC respaldan el buen comportamiento del modelo SVM.

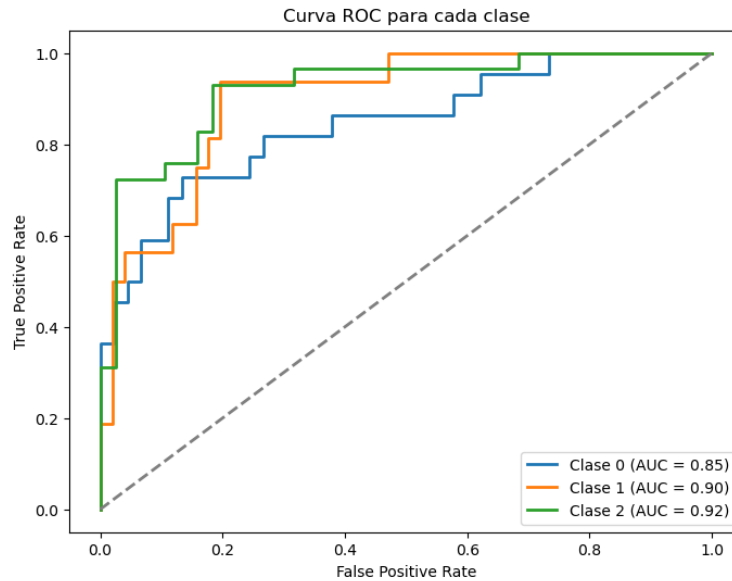


Figura 5.1: Curva de ROC SVM

▪ **Resultados por clase:**

Continuando se muestra una evaluación general del modelo:

Cuadro 5.1: Evaluación general del modelo por clase

Clase	Precision	Recall	F1-score	Soporte
Negativo	0.6154	0.7273	0.6667	22
Neutral	0.6667	0.5000	0.5714	16
Positivo	0.7931	0.7931	0.7931	29
Exactitud (Accuracy)	0.7015			
Macro promedio	0.6917	0.6735	0.6771	67
Promedio ponderado	0.7046	0.7015	0.6986	67

De la tabla anterior se puede resaltar lo siguiente:

Para clase **NEGATIVO**

- Precisión: 0.6154. De todos los casos predichos como “Negativo”, el 61.5 % eran realmente negativos.
- Recall: 0.7273. De los negativos reales, se identificaron correctamente el 72.7%.
- F1-score: 0.6667, Este fue un valor aceptable entre precisión y recall.

Para clase **NEUTRAL**

- Precisión: 0.6667. De las predicciones como “Neutral”, el 66.7 % eran correctas.
- Recall: 0.5000. Solo el 50 % de los reales “Neutral” fueron identificados correctamente.
- F1-score: 0.7931. El mejor desempeño entre las tres etiquetas.

Para clase **POSITIVO**

- Precisión y Recall: 0.7931. Buen rendimiento: casi 80 % de precisión.
- F1-score: 0.5714. Este valor indica el valor más bajo entre las clases; lo que significa que el modelo tiene más problemas con la clase neutral.

El valor de exactitud obtenido, 0.7015, refleja que el modelo logra una tasa de clasificación correcta del 70.15 % sobre el total de los casos analizados.

■ LIME (Local Interpretable Model-agnostic Explanations)

Como parte de la evaluación del modelo, se decidió implementar este modelo explicativo llamado LIME, donde se puede evidenciar de manera clara por qué el modelo en este caso SVM toma una decisión con respecto al sentimiento. La técnica fue aplicada a un audio en particular, el cual reveló la razón por la que el modelo tomó su decisión. Los detalles se muestran a continuación:



Figura 5.2: Explicabilidad con LIME

La anterior figura indica que el modelo clasificó el texto de esta llamada como positivo con una confianza del 83 %.

En donde palabras como Okay, atento, hace, puede, haciendo, hacen parte de las palabras con mayor peso que permitieron la predicción Positiva del modelo.

5.2. Modelo de clasificación no supervisada

Continuando con el desarrollo del tercer objetivo, el cual hace referencia a la elaboración de un modelo de clustering para segmentar las llamadas partiendo del análisis de sentimiento obtenido en el resultado del modelo de clasificación detallado anteriormente, se realizó la implementación de NMF, teniendo en cuenta que son textos transcritos de llamadas en un call center, esto permite identificar temas latentes de forma clara e interpretable. Esto facilita agrupar conversaciones según su contenido semántico, incluso cuando los textos son ruidosos o variados.

5.2.1. Non-Negative Matrix Factorization (NMF)

Se implementó el modelo NMF de Clustering para la identificación de topics dentro de los textos de las llamadas previamente clasificados según su sentimiento. Este modelo es útil en tareas de reducción de dimensionalidad y análisis de temas, ya que este descompone la matriz de terminos lo que permite interpretar cada componente como un topic. En este caso, se usó sobre vectores generados mediante técnicas de CountVectorizer implementado con n-gramas, esto permite capturar combinaciones relevantes de palabras. El modelo fue construido con las siguientes características:

- El texto fue previamente transformado utilizando técnicas de vectorización textual (TF-IDF).
- Se aplico eliminación de stopwords.
- Se aplicó lematización.

Cabe resaltar que el preprocesamiento es único para los dos modelos. Para cada uno de los sentimientos los hiperparámetros cambiaron teniendo en cuenta lo siguiente:

Cuadro 5.2: Hiperparámetros del modelo NMF por cada sentimiento

Sentimiento	Total de topics	min_df	max_df	Justificación
Negativo	2	60	0.7	min_df 60 conserva palabras frecuentes negativas. max_df filtra palabras muy generales.
Neutral	2	10	0.8	min_df 10 capta términos menos frecuentes para mayor variedad.
Positivo	3	40	0.6	3 topics por mayor variedad. min_df 40 filtra términos raros, max_df 0.6 descarta términos comunes.
Todas	4	180	0.6	Filtrado estricto para temas representativos generales.

5.2.1.1. Resultados modelo de NMF

De lo anterior se obtuvieron los siguientes resultados:

Para etiqueta **POSITIVO** arrojó lo siguiente:

- **Documentos Procesados:** 465
- **Shape de la Matriz Vectorizada:** (465, 366)
- **Términos por tema**
 - Tema 1: hacer, va, proceso, valor, momento, inscripción, vale, perfecto, pronto, puede
 - Tema 2: abiertas, página, pagar, osorio, oportunidad, online, número whatsapp, nuevamente
 - Tema 3: persona, ustedes, pronto, voy, curso, datos, agosto, creo, hacer, ahí

Para etiqueta **NEGATIVO** se obtuvo lo siguiente:

- **Documentos Procesados:** 322
- **Shape de la Matriz Vectorizada:** (322, 44)
- **Términos por tema**
 - Tema 1: llamada, programa, motivo, información, momento, listo, diplomado, hablar
 - Tema 2: alegre, hoy, universidad, hola, vale, motivo llamada, saber, pronto

Para etiqueta **NEUTRAL** se identificó lo siguiente:

- **Documentos Procesados:** 110

- **Shape de la Matriz Vectorizada:** (110, 125)
- **Términos por tema**
 - Tema 1: bueno, bien, información, llamada, momento, favor, vale, perfecto
 - Tema 2: ahí, mediodía, martes, lunes, llamar, lozano, interesado

Para **TODAS** las etiquetas se obtuvo lo siguiente:

- **Documentos Procesados:** 897
- **Shape de la Matriz Vectorizada:** (897, 56)
- **Términos por tema**
 - Tema 1: listo, inscripción, momento, motivo, universidad, inquietud, saber
 - Tema 2: programa, alguna duda, inscripción, interés, valor, perfecto
 - Tema 3: diplomado, interés, perfecto, pronto, saber, programa
 - Tema 4: listo, hacer, proceso, va, valor, whatsapp, correo, gusto

Análisis de resultados

En este capítulo se muestra un análisis de los resultados obtenidos de los modelos desarrollados, teniendo en cuenta la necesidad de negocio detrás de este proyecto el cual es adquirir de manera optima un modelo de análisis de sentimiento de las llamadas del call center de la universidad Javeriana de Cali y posteriormente un modelo que permite generar temas (tópicos) por cada uno de los sentimientos.

El análisis supervisado, fundamentado en un modelo de Máquinas de Soporte Vectorial (SVM), posibilitó determinar con exactitud la percepción de las llamadas: positiva, negativa o neutral. Este hallazgo es crucial para supervisar y potenciar la calidad del servicio, dando prioridad a las llamadas que necesiten respuesta inmediata e identificando los elementos que promueven la satisfacción del usuario.

Por otra lado, el análisis no supervisado, implementado mediante Non-Negative Matrix Factorization (NMF), permitió descubrir de manera automática temas y tópicos importantes expresados en las llamadas. Este enfoque ayuda a identificar los temas relacionados a las necesidades y las áreas de oportunidad para mejorar la experiencia del usuario.

La combinación de ambos análisis posibilita a la entidad convertir datos no estructurados en información aplicable, mejorando los procesos de servicio al cliente y robusteciendo la toma de decisiones estratégicas. Esto resulta en una mayor habilidad para prever las necesidades de los usuarios, poner en marcha medidas proactivas y potenciar la satisfacción de la comunidad educativa.

En esta sección se da cumplimiento al cuarto objetivo relacionado con elaborar un informe que permita visualizar de manera tabular y resumida los resultados obtenidos por los modelos.

6.1. Análisis Clasificación Supervisada

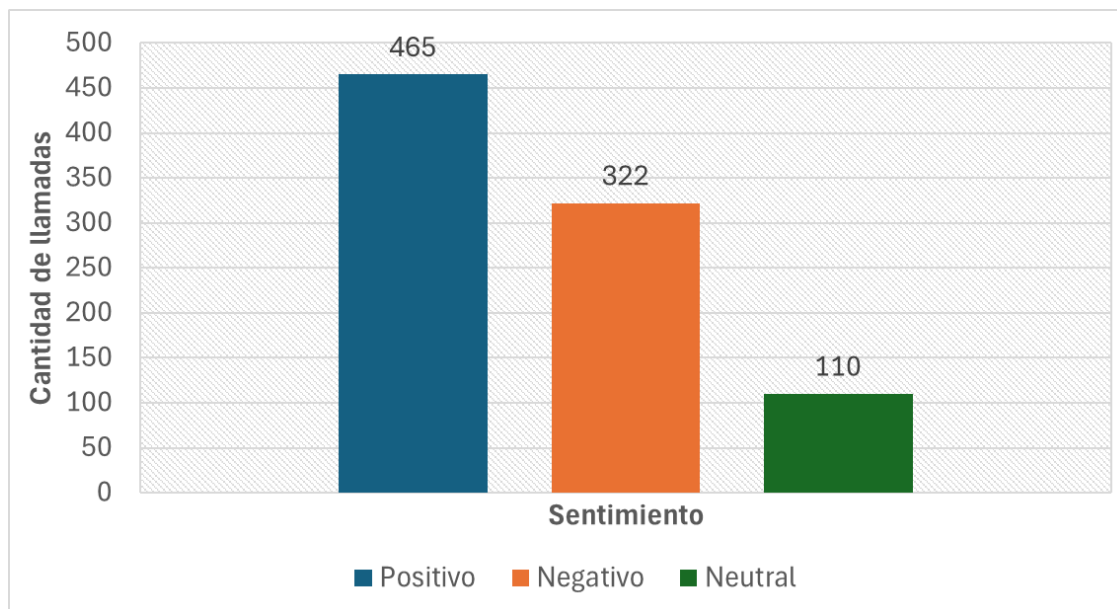


Figura 6.1: Distribución de llamadas por sentimiento

Es importante en este punto recordar los criterios definidos para cada uno de los sentimientos fueron; positiva, cuando el usuario recibe la información de forma satisfactoria y muestra conformidad o interés; negativa, cuando el usuario no queda conforme, no encuentra la respuesta esperada o manifiesta explícitamente su desinterés en el programa o la universidad; y neutral, cuando la llamada es inconclusa, se interrumpe o no tiene una definición clara debido a problemas de comunicación.

Luego de ejecutar el modelo de clasificación se encontró que de las 897 llamadas 465 casos fueron positivas, lo que permite al negocio reconocer y potenciar las buenas prácticas de servicio, así como mantener estándares de calidad en la atención al usuario. En el caso de las llamadas negativas, el modelo clasificó correctamente 322 casos, lo cual permite al negocio detectar de forma temprana posibles focos de insatisfacción, implementar planes de mejora y activar estrategias de contención y retención de usuarios. Finalmente, el modelo acertó en 110 llamadas neutras, que corresponden a aquellas interacciones en las que no se identificó un sentimiento predominante. Esto resulta clave para analizar procesos que podrían estar generando experiencias sin impacto emocional y que podrían transformarse en oportunidades de mejora.

Teniendo en cuenta los resultados anteriores, el modelo supervisado logró una exactitud general de 70,15%, lo cual indica un desempeño adecuado para un escenario de análisis de sentimiento en lenguaje natural. Para las llamadas positivas, el modelo presentó un F1-score de 0,7931, lo cual indica que aproximadamente 8 de cada 10 llamadas positivas fueron clasificadas correctamente. Este resultado destaca la capacidad del modelo para identificar con precisión las llamadas que generaron una buena experiencia, facilitando así al negocio la identificación de las fortalezas de su servicio. En

las llamadas negativas, el modelo obtuvo un F1-score de 0,6667, esto indica que en promedio 7 de cada 10 llamadas negativas fueron clasificadas correctamente. Este resultado es clave para la gestión de riesgos reputacionales y oportunidad en las mejoras del servicio, ya que permite detectar de forma oportuna posibles focos de insatisfacción y activar acciones correctivas. En el caso de las llamadas neutras, el modelo alcanzó un F1-score de 0,5714, lo que equivale aproximadamente a que 6 de cada 10 llamadas neutras correctamente identificadas. Esto evidencia un área de oportunidad para afinar el modelo y mejorar su capacidad para distinguir interacciones en llamadas de baja calidad o inconclusas, transformando estas en oportunidades de conexión con el usuario.

6.2. Análisis Clasificación No Supervisada

Teniendo en cuenta los resultados que arrojó el modelo de clasificación donde se obtuvo temas o tópicos característicos por cada uno de los sentimientos y también agrupando todas las llamadas, se encontró lo siguiente:

Etiqueta	Topic	Palabras clave	Interpretación / Tema sugerido
Positivo	1	hacer, proceso, valor, inscripción, semestre, maestría, descuento, pago, correo	Gestión de inscripción, precios y pagos (maestrías, semestre, descuentos)
	2	abiertas, página, pagar, oportunidad, online, WhatsApp, modalidad, noche	Información sobre modalidades y canales de inscripción online
	3	persona, curso, datos, agosto, tema, oportunidad, online, página	Consulta de cursos específicos y fechas disponibles
Negativo	1	llamada, programa, motivo, información, gusto, diplomado, validar	Desinterés o rechazo ante la oferta educativa o llamada inesperada
	2	alegra, hoy, hola, queremos, universidad, bien	Cordialidad sin compromiso real (interacciones sociales sin intención de continuar)
Neutral	1	buenos días, información, llamada, puedo, gusto, tarde, lunes	Interacciones informativas neutras (presentación, saludo, contacto inicial)
	2	mediodía, martes, llamar, javiana, luego	Coordinación de contacto o seguimiento posterior
Todas	1	inscripción, interés, inquietud, proceso, universidad	Interés genuino y evaluación previa al proceso de inscripción universitaria
	2	alguna duda, valor, inquietud, peso, programa	Evaluación de costos y análisis de contenidos del programa académico
	3	diplomado, inscripción, pronto, saber	Interés en diplomados y disposición inmediata para la inscripción
	4	hacer, proceso, whatsapp, correo, gusto, favor	Gestión administrativa y coordinación por medios digitales (WhatsApp, correo electrónico)

Figura 6.2: Topics generados por NMF por cada sentimiento

Como se observa en la anterior tabla el modelo de clasificación no supervisado (NMF) permitió

identificar temas o tópicos recurrentes en las llamadas del Call Center, agrupados por las etiquetas generadas previamente en el análisis supervisado (Positivo, Negativo y Neutral). Este análisis ofrece una valiosa herramienta para comprender los temas tratados en cada una de las llamadas.

■ **Etiqueta Positivo:**

Los temas detectados en las llamadas positivas muestran por parte del usuario interés con el proceso de inscripción y la disposición a continuar o tener su vínculo con la universidad:

- Tópico 1 indica el interés por el proceso de inscripción, precios y pagos.
- Tópico 2 muestra la relevancia de las modalidades de estudio y el uso de plataformas digitales (WhatsApp, páginas web) con el fin de realizar trámites.
- Tópico 3 señala la búsqueda de información sobre cursos y fechas de disponibilidad.

■ **Etiqueta Negativo:** En las llamadas clasificadas como negativas, el modelo logró identificar temas referentes a preocupaciones y posibles señales de insatisfacción:

- Topic 1 muestra interacciones donde el usuario manifiesta desinterés o rechazo ante la oferta educativa, lo que coincide con el resultado supervisado que detectó 322 llamadas negativas. Esto sugiere una oportunidad para revisar el discurso de ventas y las estrategias de captación.
- Topic 2 agrupa mensajes de cortesía iniciales que no escalan en interés real (cordialidad sin compromiso), alertando al negocio sobre la importancia de identificar estos casos y priorizar los leads con potencial de conversión.

■ **Etiqueta Neutral:**

Las llamadas neutrales representan oportunidades de conversión si se gestionan de manera proactiva:

- Topic 1 identifica llamadas de carácter informativo (saludos, contacto inicial) sin emociones predominantes, alineado con las 110 llamadas clasificadas como neutrales en el análisis supervisado.
- Topic 2 resalta temas de coordinación de horarios y seguimiento posterior.

■ **Todas**

El modelo identificó temas clave que afectan a todas las etiquetas emocionales y que deben ser considerados en la estrategia general:

- Inscripción, inquietudes y evaluación previa (Topic 1) reflejan el interés de los usuarios por comprender el proceso académico antes de comprometerse.
- Evaluación de costos y análisis de programas (Topic 2) se relacionan con preguntas frecuentes sobre la inversión académica, aspecto clave para la retención de usuarios.

- Diplomados y disposición inmediata para la inscripción (Topic 3) resaltan la necesidad de agilidad en el proceso de inscripción.
- Gestión administrativa y canales digitales (Topic 4) apuntan a la importancia de plataformas como WhatsApp y correo electrónico para mejorar la experiencia del usuario

6.3. Análisis Implementación

En esta sección se destaca como la implementación del modelo soporta y complementa los resultados obtenidos a través del análisis supervisado, generando un valor agregado para el negocio y la gestión de las llamadas en el Call Center de la Universidad Javeriana de Cali.

El desarrollo de la implementación permitió consolidar la arquitectura necesaria para ejecutar los modelos de aprendizaje automático de manera práctica, incluyendo:

- **Procesamiento de datos y transcripción de llamadas:** Gracias a la integración con la herramienta GCP Speech-to-Text, se generó la base de datos de transcripciones de llamadas, insumo esencial para el modelado y análisis.
- **Clasificación supervisada:** La implementación técnica de este modelo permitió clasificar las llamadas según el sentimiento de la interacción.
- **Interfaz gráfica de usuario (GUI):** Este entorno ayuda a la ejecución del proceso de transcripción y visualización de resultados.

6.4. Discusión de resultados orientado al negocio

Teniendo en cuenta los resultados del desarrollo de este proyecto, es importante destacar cómo la implementación de este análisis y los resultados obtenidos pueden tener un impacto positivo en la gestión del canal call center, así como en el área de mercadeo de la Universidad Javeriana de Cali.

6.4.1. Canal Call Center

El canal del call center es el principal actor en este proceso, dado que de la gestión que se realice en las llamadas dependen factores reputacionales y la efectividad en la gestión comercial de los programas académicos. Según los resultados del proyecto:

- **Contar con un proceso automático que permite realizar la transcripción de las llamadas:** tener este insumo es muy valioso, ya que resulta útil para contar con un respaldo más económico en términos de almacenamiento de las llamadas y, además, facilita diversos análisis posteriores.
- **Contar con un modelo entrenado** para clasificar las llamadas según el sentimiento expresado (positivo, negativo y neutral): el modelo supervisado mostró un desempeño global de

70,15% de exactitud, con un F1-score de 0,7931 para llamadas positivas (465 casos), 0,6667 para llamadas negativas (322 casos) y 0,5714 para llamadas neutrales (110 casos). Este análisis permite identificar oportunidades de mejora en los procesos actuales, en la forma en que se manejan las llamadas y, de ser necesario, realizar ajustes en los diálogos o herramientas con las que cuenta el asesor al momento de atender una llamada.

- **Herramientas suficientes para identificar falencias** en los procesos o en la comunicación brindada en cada una de las llamadas: el modelo no supervisado (NMF) permitió extraer temas relevantes como gestión de inscripción, procesos administrativos, costos y modalidades de atención digital. Esto aporta insumos clave para identificar las causas de las llamadas negativas y neutrales, permitiendo la mejora continua del servicio.
- **Contar con un modelo que evalúa las llamadas por sentimiento y extrae temas o tópicos relevantes:** esto proporciona una lectura diferente, permitiendo tener un resumen ejecutivo de las preocupaciones y necesidades de los usuarios en las llamadas. Por ejemplo, se identificaron temas como gestión de inscripción y pagos en llamadas positivas, y rechazo o desinterés en llamadas negativas, lo que facilita la implementación de estrategias de mejora y priorización de casos.

6.4.2. Área de Mercadeo:

- La información resultante luego del procesamiento y modelado de estos datos se convierte en un insumo de gran importancia para plantear estrategias para mejorar la efectividad en la gestión de las llamadas realizadas para ofertar los diferentes programas académicos. Por ejemplo, el análisis no supervisado evidenció un interés transversal en procesos de inscripción, costos y programas académicos, lo cual refuerza la necesidad de campañas de comunicación claras y segmentadas que respondan a las inquietudes detectadas. Con esta información, el área de mercadeo de la universidad podría tomar decisiones basadas en datos, mejorando así la planificación estratégica y aumentando las probabilidades de conversión de llamadas en inscripciones.

6.4.3. Implementación de la aplicación:

Esta implementación garantiza que el modelo desarrollado no solo se valide académicamente, sino que se integre en los procesos del Call Center y el área de Mercadeo, potenciando la calidad del servicio y la satisfacción de los usuarios. Además, permite la escalabilidad de la solución y el desarrollo de nuevas funcionalidades, así como también el entrenamiento continuo del modelo.

Implementación

7.1. Esquema de la aplicación

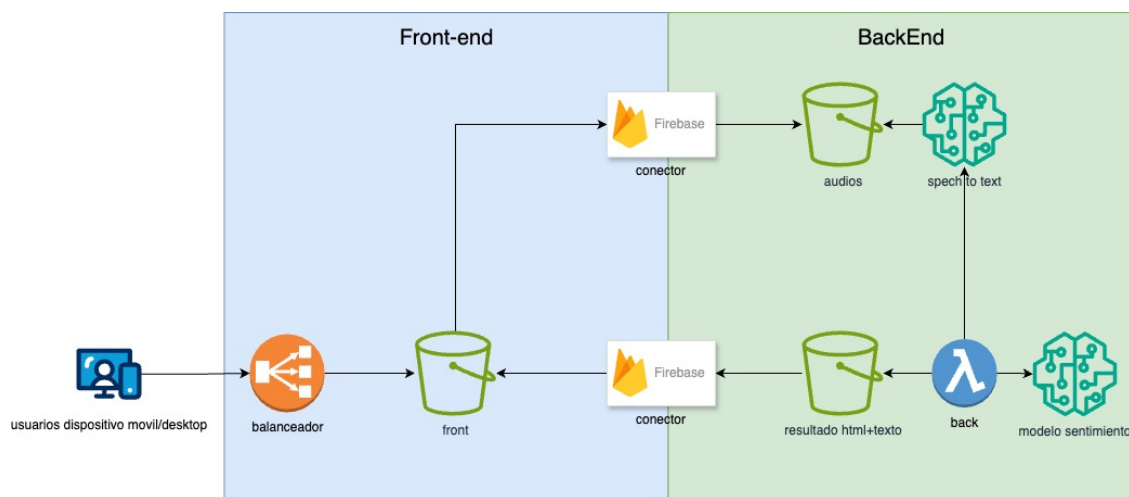


Figura 7.1: Esquema Aplicación

De cara al prototipo de implementación, en la Figura 6.1 se puede ver cómo los usuarios (representados con el ícono del computador y el teléfono) ingresan inicialmente a través de un balanceador de carga que distribuye las peticiones hacia nuestro front-end, el cual está hospedado como un sitio estático en un bucket de Cloud Storage en GCP. Desde este front, cualquier persona puede subir un archivo de audio; al hacerlo, Firebase se encarga de gestionar de forma segura[51] la autenticación y autorización, y finalmente envía esos audios a un segundo bucket de Cloud Storage dedicado exclusivamente a almacenarlos[52].

Cuando el audio llega a ese bucket de entrada, se dispara automáticamente un evento que invoca el servicio de Speech-to-Text de GCP. Mientras se lleva a cabo la transcripción, el front-end permanece en “escucha” (haciendo polling) del bucket de resultados para detectar cuándo el audio ya ha sido convertido a texto. Una vez que Speech-to-Text termina su trabajo, sube el texto transcrito a un bucket intermedio y, a su vez, dispara otro evento que activa la Cloud Function llamada “back”. Esta función, que ya tiene cargado nuestro modelo de análisis de sentimiento, toma el texto, calcula el sentimiento correspondiente y, además, genera el archivo HTML de explicación con LIME. Ambos resultados (el valor de sentimiento y el HTML de LIME) se guardan en el bucket de resultados.

Inmediatamente, Firebase notifica al front-end que hay nuevos archivos disponibles, de modo que la interfaz toma esa información y la muestra al usuario. Si alguien vuelve a subir un audio que ya existía en el bucket de entrada, el sistema detecta que el resultado ya está en el bucket de resultados y lo devuelve de inmediato, ahorrando tiempo y costos de procesamiento.

Además, esta arquitectura está diseñada para manejar cargas masivas: si se suben muchos audios a la vez, cada uno desencadena su propia ejecución de Speech-to-Text y de la Cloud Function “back”, y todos los outputs terminan en el bucket de resultados. Tanto los buckets como el servicio de Speech-to-Text y las Cloud Functions son completamente escalables, por lo que pueden soportar un alto número de peticiones por segundo sin necesidad de provisión manual de recursos. Todo este flujo de eventos está orquestado en GCP y, como se evidencia en [Figura 7.2 Arquitectura Aplicación], la conexión entre front-end y back-end se maneja mediante Firebase con integración a Cloud Storage, lo que garantiza la seguridad necesaria [46] y sigue el patrón observado en [47].

Para facilitar la modularidad, el front-end puede tratarse como un microfrontend independiente. Esto quiere decir que, en el futuro, podríamos dividir la interfaz en varios módulos, cada uno en su propio bucket y con sus reglas de seguridad en Firebase, sin alterar la lógica de orquestación. Finalmente, la aplicación ya está en producción y se puede acceder en <http://34.96.95.174/>, donde la primera pantalla permite seleccionar el archivo a procesar y poner en marcha todo el flujo descrito.

Finalmente, la aplicación fue alojada en el siguiente enlace: <http://34.96.95.174/>.

Esta primera pantalla permitió seleccionar el archivo a procesar.



Figura 7.2: Pantalla Cargue

La segunda pantalla luego de confirmar que se cargó el archivo, permite ejecutar el proceso.



Figura 7.3: Pantalla Procesamiento

Ya aquí se puede ver el resultado del análisis de sentimiento, usando el modelo entrenado.



Figura 7.4: Pantalla Resultado Modelo

Ya por ultimo se puede ver la evaluación de la clasificación por medio del método LIME

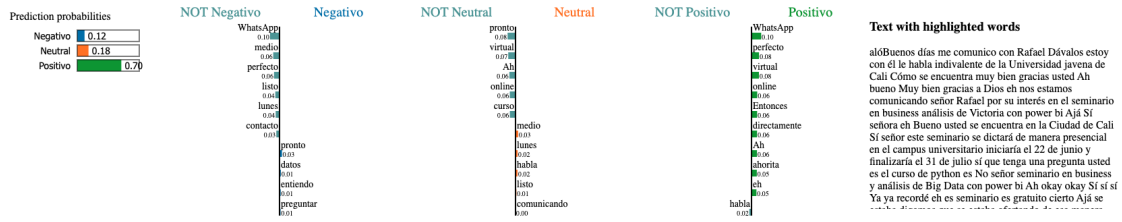


Figura 7.5: Pantalla Evaluación Modelo LIME

El repositorio correspondiente a todo el proyecto de investigación esta a continuación:

https://github.com/nelsonAndres1/Proyecto_Analisis_Sentimiento.git

Conclusiones y Trabajos futuros

8.1. Conclusiones

- El sistema de análisis de sentimiento implementado, construido con GCP Speech-to-Text y con la implementación de modelos SVM demostró su eficiencia para clasificar con notables valores de precisión los sentimientos en las llamadas del Call Center. Esta arquitectura logró automatizar la comprensión de las percepciones de los usuarios según los sentimientos Positivo, Negativo y Neutral.
- El modelo SVM con vectorización ONEHOTENCODING obtuvo resultados satisfactorios, alcanzando valores de F1 de 0.74 para el sentimiento Positivo, 0.76 para el Neutral y 0.55 para el Negativo, así como una capacidad de aprendizaje de 0.87. De igual forma, las redes neuronales recurrentes (RNN) indicaron resultados similares, aunque con valores bajos en el sentimiento Neutral. También NMF permitió extraer de buena forma temas representativos para cada topic de sentimiento, esto ayudo para interpretar de mejor manera los sentimientos.
- Algunos de los problemas o limitaciones fue la calidad de la transcripción ya que está depende de la claridad del audio original, es decir de la llamada, y de la homogeneidad de las mismas, esto representa un problema para que los modelos generalicen nuevos temas o situaciones diferentes, por lo que a futuros trabajos se deberá tener en cuenta el aumento y diversidad de los datos de entrenamiento.
- En la ejecución del proyecto, descubrió que los estudios de sentimientos no siempre manifiestan emociones en un sentido riguroso, tal como inicialmente se anticipo. Conforme se avanzó en el proceso, se entendió que en este caso, el estudio de la emoción debería centrarse en indicadores vinculados con la satisfacción y el éxito de la llamada.
- La integración de modelos supervisados y no supervisados en el análisis de las llamadas del call center permite a la Universidad Javeriana de Cali optimizar sus procesos de atención y mercadeo, alineando sus estrategias con las necesidades reales de los usuarios y potenciando su liderazgo en calidad educativa y servicio al cliente.
- Un aspecto clave del proyecto fue la implementación, que permitió desarrollar y desplegar una aplicación funcional capaz de integrar el modelo de clasificación entrenado con una interfaz de usuario. Esto garantiza la operatividad real del sistema, brindando a las áreas de Call Center y Mercadeo una herramienta accesible y efectiva para visualizar los resultados de clasificación

de las llamadas. Este enfoque práctico convierte la investigación en un producto aplicable directamente en el contexto empresarial de la Universidad.

8.2. Trabajos futuros

- **Profundización del estudio:** Es importante en un análisis y entrenamiento futuro del modelo poder contar con una base de audios mas amplia y poder tenerlas clasificadas de cuales son llamadas entrantes o salientes de las líneas del Call Center, esto ayudaría a que el modelo fuera mas efectivo y poder ser un poco mas asertivos en la predicción del modelo de clasificación y contar con mejores tópicos del modelo de clusterización.
- **Implementación sobre el área de negocio:** Sería fascinante que el departamento de atención al cliente y marketing empleara los resultados obtenidos en este proyecto como el fundamento de un proceso futuro, respaldado en la ciencia de los datos y enriquecer y profundizar en el progreso y los hallazgos obtenidos en este proyecto. Es crucial comprender que el trabajo efectuado aquí es un proceso viable y con todas las oportunidades de ser mejorado con la reentrenamiento del modelo y la optimización de los procesos existentes, tanto de atención de clientes como de ofertas de programas académicos.
- **Propuestas para mejora del prototipo:** Contando con un mayor número de datos, se podría ampliar el número de audios etiquetados, esto permitiría un mejor resultado y seguramente se podría encontrar algunos audios no tan homogéneos como con los que contamos en el proyecto, adicional a esto seguramente la aplicación tiene muchas oportunidades de mejora en tanto a la parte gráfica y que la experiencia al usuario sea mucho mejor y al final que los modelos y la aplicación se vean como una solución integral a la mejora de procesos y efectividad en los procesos de servicios al cliente y marketing.

Bibliografía

- [1] significados.com, “Sentimientos (qué son, concepto y significado) - enciclopedia,” 2025. Fecha de acceso: mayo 10, 2025.
- [2] psicologiaymente.com, “Tipos de sentimientos,” 2025. Fecha de acceso: mayo 10, 2025.
- [3] Academia Leedu, “¿qué son los sentimientos?,” 2025. Fecha de acceso: mayo 10, 2025.
- [4] alquimedez.com, “Los 16 tipos de sentimientos y su función psicológica,” 2025. Fecha de acceso: mayo 10, 2025.
- [5] Revista de Psicología Científica, “Emociones, sentimientos y salud del individuo,” 2025. Fecha de acceso: mayo 10, 2025.
- [6] J. M. F. Dols and J. E. Ortega, “Los niveles de análisis de la emoción: James, cien años después,” *Studies in Psychology*, vol. 6, no. 21, pp. 34–56, 1985.
- [7] Y. Rengifo de Vargas, “Las emociones de enfermería: una mirada desde el cuidado humano,” *Revista arbitrada del CIEG [Internet]*, vol. 52, pp. 146–61, 2021.
- [8] Á. D. Ortiz, “Teoría de las emociones,” *Innovación y experiencias educativas (29)*, 2010.
- [9] psicologiaymente.com, “Psicología emocional: principales teorías de la emoción,” 2025. Fecha de acceso: mayo 10, 2025.
- [10] C2Care, “La teoría de la emoción construida,” 2025. Fecha de acceso: mayo 10, 2025.
- [11] H. Hota, D. K. Sharma, and N. Verma, “Lexicon-based sentiment analysis using twitter data: A case of covid-19 outbreak in india and abroad,” 2021.
- [12] N. T. Grygoryev, “Transforme su empresa con la clasificación de texto automatizada,” Sep 2023.
- [13] Transkriptor, “¿qué es la transcripción de audio?,” 2023. Accedido el 18 de mayo de 2025.
- [14] PubMed Central, “Lexicon-based sentiment analysis using twitter data: a case of covid-19 outbreak in india and abroad,” 2025. Fecha de acceso: mayo 10, 2025.
- [15] Amberscript, “Transcripción de audio: Cómo transcribir audio a texto,” 2025. Fecha de acceso: mayo 10, 2025.
- [16] Google Cloud, “¿qué es el aprendizaje supervisado?,” 2025. Fecha de acceso: mayo 18, 2025.
- [17] Google Cloud, “¿qué es el aprendizaje automático?,” 2025. Fecha de acceso: mayo 18, 2025.
- [18] Wikipedia, “Aprendizaje no supervisado,” 2025. Fecha de acceso: mayo 18, 2025.

-
- [19] L. Ibero, “Técnicas de machine learning: Clasificación, regresión y clustering.”
- [20] F. Sebastiani, “Machine learning in automated text categorization,” *ACM computing surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002.
- [21] SciELO Colombia, “Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning,” 2025. Fecha de acceso: mayo 10, 2025.
- [22] Linedata, “Supervised machine learning and classification,” 2025. Fecha de acceso: mayo 18, 2025.
- [23] Innovatiana, “Análisis de sentimientos: la revolución del marketing con ia,” 2025. Fecha de acceso: mayo 10, 2025.
- [24] Zonka Feedback, “Unsupervised sentiment analysis: Extracting insights from unlabeled data,” 2025. Fecha de acceso: mayo 18, 2025.
- [25] MSMK University, “¿qué es el análisis de sentimiento?,” 2025. Fecha de acceso: mayo 15, 2025.
- [26] AWS, “¿qué es una rnn?: Explicación sobre redes neuronales,” 2025. Fecha de acceso: mayo 15, 2025.
- [27] IBM, “¿qué es una red neuronal recurrente (rnn)?,” 2025. Fecha de acceso: mayo 15, 2025.
- [28] Codificando Bits, “Introducción a las redes neuronales recurrentes,” 2025. Fecha de acceso: mayo 15, 2025.
- [29] ResearchGate, “Técnicas de análisis de sentimientos aplicadas a la valoración de opiniones en el lenguaje español,” 2025. Fecha de acceso: mayo 10, 2025.
- [30] MDPI, “Sentiment analysis of product reviews using machine learning and pre-trained llm,” 2025. Fecha de acceso: mayo 10, 2025.
- [31] Wikipedia, “Non-negative matrix factorization,” 2025. Fecha de acceso: mayo 15, 2025.
- [32] QuestionPro, “Análisis de sentimiento. qué es y cómo realizarlo,” 2025. Fecha de acceso: mayo 15, 2025.
- [33] D. M. W. Powers, “Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [34] Sonix, “¿qué es la transcripción de datos? una guía completa,” 2025. Fecha de acceso: mayo 10, 2025.
- [35] wovi.ai, “Transcripción de datos en investigación: menos tiempo transcribiendo, más descubriendo,” 2025. Fecha de acceso: mayo 10, 2025.

-
- [36] Google Cloud, “Speech-to-text ai: speech recognition and transcription,” 2025. Fecha de acceso: mayo 15, 2025.
- [37] Cloudfresh, “Google cloud speech-to-text,” 2025. Fecha de acceso: mayo 15, 2025.
- [38] Observability Cloud, “Gcp cloud storage,” 2025. Fecha de acceso: mayo 15, 2025.
- [39] Google Cloud, “Storage overview,” 2025. Fecha de acceso: mayo 15, 2025.
- [40] Google Cloud, “Cloud storage documentation,” 2025. Fecha de acceso: mayo 15, 2025.
- [41] Google Colab, “Te damos la bienvenida a colab,,” 2025. Fecha de acceso: mayo 15, 2025.
- [42] Trantor, “Natural language processing with python in 2025,” 2025. Fecha de acceso: mayo 15, 2025.
- [43] F. Quiñones Bellón, M. Sánchez de Paz, and E. P. Fernández de Pedro, “Sistema híbrido para el análisis de sentimientos en llamadas,” 2021.
- [44] A. Ahmed, S. Toral, and K. Shaalan, “Agent productivity measurement in call center using machine learning,” in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*, vol. 2, pp. 160–169, 2017.
- [45] J. Bockhorst, S. Yu, L. Polania, and G. Fung, “Predicting self-reported customer satisfaction of interactions with a corporate call center,” in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part III 10*, pp. 179–190, Springer, 2017.
- [46] D. K. Dake and E. Gyimah, “Using sentiment analysis to evaluate qualitative students’ responses,” *Education and Information Technologies*, vol. 28, no. 4, pp. 4629–4647, 2023.
- [47] Y. Fernández, “OpenAI whisper: qué es, cómo funciona y cómo puedes usar esta inteligencia artificial para transcribir audios.” Section: basics.
- [48] “Amazon transcribe – voz a texto – AWS.”
- [49] “IA de speech-to-text: reconocimiento y transcripción de voz.”
- [50] “Google cloud chirp model for speech AI.”
- [51] “Plantilla_proyaplicado.”
- [52] “Google speech-to-text API features— evonence | google cloud partner.”