

# Análisis del comportamiento de cliente (NPS) usando Inteligencia Artificial Generativa

julio 2025



**Universidad**  
Internacional  
de Valencia

Titulación:

Máster en Inteligencia Artificial

Curso académico

2024 – 2025

Alumno/a:

Martín Cruz, Eva María

D.N.I.: 27536897D

Director de TFM:

Mario de Felipe Pérez

Convocatoria:

Cuarto período

De:

 Planeta Formación y Universidades



## Índice

Resumen .....	7
Abstract .....	8
1. El problema .....	9
1.1. Situación actual .....	9
1.2. Formulación del problema .....	9
1.3. Objetivos .....	10
1.3.1. Objetivo general .....	10
1.3.2. Objetivos específicos .....	10
1.4. Justificación .....	10
1.5. Delimitación del problema .....	10
2. Antecedentes .....	12
2.1. El Net Promoter Score en sectores sensibles .....	12
2.2. Limitaciones técnicas .....	13
2.3. Limitaciones de las fuentes de datos .....	14
3. Estado del arte .....	15
3.1. Análisis de Sentimientos .....	16
3.2. Técnicas de Análisis de Sentimientos .....	16
3.3. ¿Por qué necesitábamos Transformers? .....	17
3.3.1. Problemas con RNNs y LSTMs: .....	17
3.3.2. Cómo funciona la Autoatención (ejemplo) .....	18
3.3.3. Cálculo de las puntuaciones de atención .....	20
3.3.4. Normalización - Aplicar Softmax a los pesos .....	21
3.3.5. Cálculo de la suma ponderada de los valores .....	22
3.3.6. Atención Multi-Head: Entendiendo el lenguaje desde distintas perspectivas .....	23
3.3.7. Codificación Posicional .....	24
3.3.8. Cabeceras de Transformer: .....	24
3.4. Los modelos de transformers más utilizados para el análisis de sentimiento son: .....	24
3.5. ¿Cómo se entrenan los modelos de transformers para el análisis de sentimiento en diferentes idiomas? .....	25
3.6. Transformers en español para análisis de sentimiento de secuencias de texto .....	26

3.7.	Transformer más adecuado para una secuencia de 600 caracteres teniendo en cuenta que es en español .....	26
3.8.	Carencias y oportunidades de avance.....	27
4.	Análisis y Estructura del Conjunto de Datos .....	30
4.1.	Caracterización del Corpus de Transcripciones.....	30
4.2.	Composición y distribución del corpus.....	30
4.3.	Métricas descriptivas del corpus.....	30
4.4.	Análisis lingüístico y conversacional .....	31
4.5.	Estructura y dinámica conversacional .....	31
4.6.	Particularidades léxico-semánticas.....	32
4.6.1.	Terminología especializada del dominio .....	32
4.6.2.	Fórmulas protocolarias y rutinas conversacionales.....	32
4.6.3.	Marcadores pragmáticos y de cortesía .....	32
4.7.	Rasgos derivados de la transcripción automática.....	32
4.8.	Patrones Discursivos por Categoría de Sentimiento.....	33
4.8.1.	Indicadores de sentimiento positivo .....	33
4.8.2.	Marcadores de sentimiento negativo .....	33
4.8.3.	Características de las transcripciones neutras.....	34
4.9.	Problemas relacionados con la calidad de los datos .....	34
4.10.	Retos lingüísticos específicos.....	34
4.11.	Desafíos para la modelización con Deep Learning.....	35
4.12.	Estrategia de Preprocesamiento Propuesta .....	35
4.12.1.	Normalización textual adaptativa .....	35
4.12.2.	Segmentación y estructuración conversacional .....	36
4.12.3.	Extracción de características específicas para análisis de sentimiento.....	36
4.12.4.	Consideraciones para el Modelado con Deep Learning.....	36
4.12.5.	Arquitecturas prometedoras para el corpus .....	36
4.13.	Estrategias de representación textual.....	37
5.	Materiales y métodos.....	38
5.1.	Fases asociadas al proceso (CRISP-DM) .....	39
5.1.1.	Entendimiento del negocio.....	39
5.1.2.	Entendimiento de los Datos .....	40
5.1.3.	Preparación de los Datos.....	41
5.1.4.	Modelado.....	42

5.1.5.	Despliegue.....	42
5.2.	Planificación del TFM .....	43
5.3.	Flujo del proceso .....	44
5.4.	Estructura común de los módulos de preprocesamiento .....	45
5.4.1.	MÓDULO 01: Carga de datos.....	45
5.4.2.	MÓDULO 02: Identificación de interlocutores .....	46
5.4.3.	MÓDULO 03: Creación y gestión del glosario de términos .....	48
5.4.4.	MÓDULO 04: Limpieza avanzada .....	49
5.4.5.	MÓDULO 05: Generación y ajuste de embeddings .....	50
6.	Hardware y Software .....	52
7.	Conclusiones generales.....	55
8.	Futuras líneas de trabajo .....	57
8.1.	Entrenamiento Multidominio o Multicompañía .....	57
8.2.	Análisis de sentimiento más profundo de las transcripciones .....	58
8.3.	Evolución del Glosario Técnico hacia una ontología formal.....	58
8.4.	Optimización computacional y sostenibilidad del flujo de procesamiento.....	59
9.	Enlaces de interés .....	60
10.	Referencias.....	61

# Índice de ilustraciones

Ilustración 1. Mecanismo de Autoatención Neural .....	18
Ilustración 2. Autoatención en detalle .....	19
Ilustración 3. Cálculo de la matriz de autoatención .....	20
Ilustración 4. Forma de cálculo de la matriz de autoatención .....	21
Ilustración 5. Cálculo de la suma ponderada de valores .....	22
Ilustración 6. Mecanismo de atención de un modelo de transformación .....	23
Ilustración 7. CRISP DM .....	39
Ilustración 8. Flujo del proceso .....	44
Ilustración 9. Estructura común de los módulos .....	45
Ilustración 10. Generación y ajuste de embeddings .....	51
Ilustración 11. Configuración local y colab .....	52

## Resumen

Este trabajo explora cómo la inteligencia artificial generativa puede automatizar el estudio del comportamiento del cliente en el ámbito de seguros de decesos, mediante el análisis de sentimiento en transcripciones de llamadas telefónicas gestionadas desde el centro de atención al cliente. En este contexto, comprender el lenguaje emocional adquiere una importancia estratégica para la optimización del servicio.

La metodología adoptada ha sido de carácter experimental y se ha guiado por el modelo CRISP-DM, lo que ha permitido estructurar el proceso en fases claramente delimitadas, orientadas a la reutilización y la mejora continua. El desarrollo se ha llevado a cabo con una arquitectura modular, con especial atención a la independencia y facilidad de mantenimiento de cada componente.

El proceso comienza con la identificación de interlocutores mediante modelos LLM (Llama3), seguida de la extracción de términos con spaCy y TF-IDF, a partir de los cuales se construyó un glosario especializado del dominio. Este recurso, concebido como base para una futura ontología sectorial, se utilizó para expandir el tokenizer y ajustar los vectores de características lingüísticas antes de la clasificación. El componente principal, un transformer Longformer adaptado al español, fue ajustado en dos etapas: una inicial centrada en la adecuación semántica mediante embeddings, y otra posterior de aprendizaje completo.

Los elementos desarrollados permiten anticipar que el ajuste previo con terminología del sector puede mejorar la capacidad del sistema para identificar sentimientos en conversaciones sensibles. Asimismo, se plantea la viabilidad del uso de modelos generativos como herramienta de etiquetado en tareas no supervisadas, en ausencia de datos previamente registrados.

Como conclusión, este trabajo presenta una estrategia replicable en otros entornos sensibles, que combina inteligencia artificial explicable y un tratamiento lingüístico adaptado al contexto. Esta aproximación permite preparar los datos textuales de forma más precisa antes de entrenar los modelos, lo que resulta especialmente relevante cuando lenguaje profesional y afectivo se entrelazan. A partir de esta base, se abre la posibilidad de construir una ontología especializada que sirva de apoyo al desarrollo de sistemas más robustos y contextualizados.

# Abstract

This work explores how generative artificial intelligence can automate the study of customer behavior in the field of funeral insurances, through sentiment analysis applied to transcripts of telephone calls managed from the customer service center. In this context, understanding emotional language becomes strategically important for service optimization.

The adopted methodology was experimental in nature and followed the CRISP-DM model, which enabled the process to be structured into clearly defined phases aimed at reusability and continuous improvement. The development was carried out using modular architecture, with special attention to the independence and maintainability of each component.

The process begins with speaker identification using LLM models (Llama3), followed by the extraction of terms with spaCy and TF-IDF, from which a domain-specific glossary was constructed. This resource, conceived as the foundation for a future sectoral ontology, was used to expand the tokenizer and adjust the linguistic feature vectors prior to classification. The main component, a Longformer transformer adapted to Spanish, was fine-tuned in two stages: an initial one focused on semantic alignment via embeddings, and a subsequent stage of full training.

The developed components suggest that prior adjustment using sector-specific terminology can enhance the system's ability to detect sentiment in sensitive conversations. Furthermore, the use of generative models is proposed as a viable tool for labeling in unsupervised tasks, in the absence of previously annotated data.

In conclusion, this work presents a replicable strategy for other sensitive domains, combining explainable artificial intelligence with linguistically contextualized text processing. This approach enables more accurate preparation of textual data prior to model training, which is particularly relevant when professional and emotional language are linked. On this basis, it opens the possibility of constructing a specialized ontology to support the development of more robust and context-aware systems.



# 1. El problema

## 1.1. Situación actual

Grupo ASV es una organización familiar de servicios con más de 90 años de trayectoria, que opera principalmente en transporte sanitario, seguros y servicios funerarios con sede en Alicante y presencia internacional en Alemania, Estados Unidos y Colombia.

Su misión es proporcionar a sus asegurados ayuda, confianza y tranquilidad mediante servicios y productos de calidad, ejecutados por profesionales comprometidos con la atención al cliente. Su visión estratégica se centra en alcanzar el mayor índice de satisfacción del cliente.

Actualmente, la empresa evalúa la calidad percibida por sus clientes mediante el indicador Net Promoter Score (NPS), una métrica ampliamente utilizada en el sector para valorar la experiencia del cliente y su disposición a recomendar los servicios recibidos.

Para llevar a cabo esta medición, colabora con una firma especializada en estudios de satisfacción que se encarga de todo el proceso, desde la realización de las encuestas hasta el análisis de los resultados. Esta firma realiza anualmente un volumen significativo de llamadas telefónicas, que oscila entre las diez mil y las quince mil, dirigidas a clientes seleccionados en momentos clave del ciclo de relación, como pueden ser la contratación de un producto, la atención recibida tras la resolución de un siniestro o cualquier otra interacción relevante.

Este enfoque permite obtener una visión representativa del grado de satisfacción de los asegurados en diferentes fases del servicio.

## 1.2. Formulación del problema

El actual método presenta dos problemas fundamentales:

- La medición de la satisfacción no abarca todos los puntos de contacto con el cliente, sino únicamente momentos específicos.
- La externalización del proceso implica un coste económico significativo y limita la capacidad de realizar un análisis inmediato y ágil de la información obtenida.

Además, existe un desaprovechamiento de información cualitativa valiosa, presente en las transcripciones de llamadas telefónicas al Centro de Atención Telefónica (CAT), ya que no se dispone de herramientas internas para un análisis automático del sentimiento y satisfacción del cliente.

En este contexto, surge la necesidad de explorar alternativas basadas en tecnología avanzada que permitan una medición más completa, precisa y económica del NPS.

## 1.3. Objetivos

### 1.3.1. Objetivo general

Evaluar la viabilidad del uso de técnicas de Deep Learning para estimar automáticamente el NPS de los clientes, a partir de las transcripciones de llamadas gestionadas en el Centro de Atención Telefónica.

### 1.3.2. Objetivos específicos

- Analizar el estado del arte actual en técnicas de Procesamiento de Lenguaje Natural (PLN) aplicadas a casos de uso similares de estimación del NPS.
- Desarrollar un modelo basado en Deep Learning que permita calcular el NPS utilizando un conjunto de datos previamente etiquetados a partir de transcripciones.
- Facilitar un análisis más ágil, preciso y económico del sentimiento del cliente mediante un sistema interno de análisis automatizado.

## 1.4. Justificación

La implementación de técnicas avanzadas de Deep Learning en la estimación automática del NPS justifica este estudio por varias razones:

- La necesidad de obtener una medición integral del grado de satisfacción del cliente, cubriendo todas las interacciones posibles, no solo las experiencias positivas destacadas.
- La reducción de costes asociados a servicios externos y encuestas telefónicas masivas.
- La mejora significativa en la capacidad de reacción de la empresa, al poder realizar análisis de forma ágil y tomar decisiones con más celeridad basadas en información actualizada y cualitativa.
- El aprovechamiento eficiente del volumen considerable de datos cualitativos que actualmente están infrautilizados.

## 1.5. Delimitación del problema

El presente trabajo se circunscribe al análisis del sentimiento del cliente en el contexto específico del sector de atención telefónica del Grupo ASV, dentro del ramo de seguros de decesos. La investigación se centrará exclusivamente en el estudio de las transcripciones textuales de llamadas realizadas al Centro de Atención Telefónica, sin abordar el tratamiento directo de las grabaciones de audio originales.

Las transcripciones utilizadas presentan las siguientes características:

- Han sido generadas previamente mediante técnicas de reconocimiento automático del habla, concretamente con el modelo Whisper, una herramienta basada en inteligencia artificial desarrollada por OpenAI.
- Han sido sometidas a un proceso de anonimización que garantiza el cumplimiento de la normativa vigente en materia de protección de datos personales.
- Se derivan de conversaciones reales entre clientes y agentes, con una duración media de entre 2 y 3 minutos por llamada, lo que supone un volumen aproximado de 600 palabras por transcripción.

Esta delimitación implica que el estudio se enfoca exclusivamente en el procesamiento, limpieza y análisis de los textos ya generados, descartando aspectos como la conversión de audio a texto, y se limita al entorno y a los datos de una única empresa del sector asegurador, lo que acota el alcance de las conclusiones y resultados obtenidos.

## 2. Antecedentes

### 2.1. El Net Promoter Score en sectores sensibles

Tradicionalmente, Grupo ASV ha medido la calidad percibida por sus clientes a través del Net Promoter Score (NPS). Para esta medición, la empresa ha recurrido a servicios externos que anualmente realizan entre 10.000 y 15.000 encuestas telefónicas en momentos específicos de interacción con los clientes, tales como contratación de productos o seguimiento tras siniestros.

Este método ha presentado limitaciones como:

- Un coste elevado.
- Potencial sesgo hacia experiencias más positivas.
- Escasa capacidad de análisis dinámico.

El Net Promoter Score (NPS), introducido por Fred Reichheld en 2003, ha demostrado ser una métrica eficaz para medir la lealtad del cliente en diversos sectores comerciales. Sin embargo, su aplicación en servicios relacionados con el duelo tiene unas características especiales que han sido escasamente exploradas.

A diferencia de campos como el minorista, telecomunicaciones o servicios financieros, donde las interacciones con clientes se producen en contextos comerciales neutros o positivos, los seguros de decesos operan en un entorno emocionalmente complejo. Las conversaciones telefónicas están impregnadas de dolor y vulnerabilidad emocional, factores que introducen dificultades únicas en la medición de la satisfacción del cliente.

La polaridad emocional en conversaciones sobre seguros de decesos no sigue los patrones binarios tradicionales (positivo/negativo). Una conversación puede contener simultáneamente expresiones de gratitud hacia el servicio recibido y manifestaciones profundas de dolor por la pérdida, situaciones que los modelos convencionales no están preparados para interpretar adecuadamente.

La literatura existente sobre NPS se ha centrado en contextos donde la satisfacción se correlaciona directamente con aspectos tangibles del servicio (velocidad, precio, calidad del producto). En el ámbito de los seguros de decesos, dicho concepto trasciende estos parámetros tradicionales, incorporando dimensiones de empatía, o soporte emocional que no han sido abordadas en investigaciones previas (según fuentes consultadas).

Además, el potencial informativo de las transcripciones de las llamadas recibidas en su Centro de Atención Telefónica se encuentra infrautilizado. Aunque estas transcripciones se almacenan de forma anonimizada y en formato textual, hasta el momento no han sido sometidas a análisis automatizados avanzados que podrían aportar información valiosa sobre la satisfacción real de los clientes.

Cabe destacar que el NPS no solo es un indicador relevante a nivel operativo, sino que también influye directamente en la retribución variable de los empleados, afectando por tanto a sus honorarios. La medición de este indicador se utiliza internamente como herramienta de evaluación del desempeño, lo que subraya aún más la importancia estratégica de una medición precisa y eficiente.

El Grupo ASV ha explorado previamente otras vías para automatizar y mejorar este proceso, encontrándose con desafíos técnicos significativos, especialmente relacionados con la calidad de las transcripciones automáticas, el manejo de terminología especializada del sector, y la complejidad técnica inherente al desarrollo de modelos predictivos robustos y precisos.

## 2.2. Limitaciones técnicas

Los modelos tradicionales de análisis de sentimientos han sido entrenados principalmente con datos provenientes de reseñas de productos, redes sociales, noticias y textos comerciales. Estos conjuntos de datos presentan patrones emocionales y lingüísticos que difieren sustancialmente de las conversaciones que se producen en contextos de duelo.

Las interacciones telefónicas en seguros de decesos presentan particularidades lingüísticas que no han sido objeto de estudio sistemático en la literatura de PLN:

- **Eufemismos relacionados con la muerte:** El uso frecuente de términos indirectos para referirse al fallecimiento ("nos ha dejado", "ha pasado a mejor vida", "ya no está con nosotros") que pueden no ser capturados correctamente por modelos entrenados en corpus generales.
- **Terminología técnica sensible:** La intersección entre jerga técnica aseguradora y lenguaje relacionado con la muerte crea un vocabulario específico inexplorado en estudios previos.

La búsqueda realizada en las principales bases de datos académicas revela una ausencia notable de investigaciones que aborden específicamente el análisis automático de sentimientos en conversaciones de seguros de decesos. Los estudios existentes en el sector asegurador se han centrado en:

- Detección de fraude en reclamaciones generales.
- Análisis de satisfacción en seguros de vida tradicionales.
- Procesamiento de quejas en seguros de automóvil y hogar.
- Optimización de procesos en seguros de salud.

## 2.3. Limitaciones de las fuentes de datos

Los datasets públicos utilizados para entrenar modelos de análisis de sentimientos (como IMDB Movie Reviews, Stanford Sentiment Treebank, Amazon Product Reviews) carecen completamente de ejemplos representativos del lenguaje utilizado en contextos de duelo.

Frente a esta situación, surge la necesidad de implementar métodos avanzados basados en técnicas de Inteligencia Artificial, con el objetivo de realizar una medición más completa, ágil y económica del NPS, aprovechando toda la riqueza informativa contenida en las transcripciones automáticas de las que se dispone.

Este contexto y estas experiencias previas constituyen los antecedentes clave que justifican la realización del presente Trabajo Fin de Máster, proporcionando así el marco idóneo para la innovación y mejora continua en la gestión del análisis de satisfacción del cliente en el Grupo ASV.

### 3. Estado del arte

La historia del análisis de la satisfacción del cliente y la predicción de su fidelidad comienza con una métrica aparentemente simple: el Net Promoter Score (NPS).

Introducido por Fred Reichheld en 2003, el NPS se consolidó como una herramienta clave para medir la lealtad del cliente, clasificándolos en promotores, pasivos y detractores según su probabilidad de recomendar un servicio. A pesar de su amplia adopción en sectores como el retail o las telecomunicaciones, pronto se hizo evidente que, en industrias más sensibles, como los seguros de decesos, el NPS tradicional presentaba limitaciones significativas.

Pronto hubo voces críticas que las señalaban: Fisher y Kordupleski (2019)<sup>1</sup> demostraron que el NPS adolece de sesgos demográficos y culturales, y que su pregunta única (“¿Recomendaría este servicio?”) no captura toda la complejidad de la relación cliente-empresa.

Conforme el NPS se extendía a sectores sensibles (sanitarios, de emergencia o funerarios) emergió la necesidad de profundizar en su fiabilidad. En una revisión sistemática de 12 estudios entre 2005 y 2020, Adams et al. (2022)<sup>2</sup> concluyeron que, aunque fácil de usar, el NPS por sí solo aporta un valor limitado y debe combinarse con métricas cualitativas para entender por qué un cliente se convierte en promotor o detractor.

La medición convencional, basada en métodos cuantitativos en puntos de contacto específicos, deja un vacío inmenso de información cualitativa, una fuente de datos valiosa pero infrautilizada.

A partir de ahí, la comunidad empezó a explorar modelos predictivos de NPS que añaden variables contextuales a la puntuación simple. Osmanski-Zenk et al. (2023)<sup>3</sup> estudiaron la correlación entre resultados clínicos informados por el paciente (PROMs, por sus siglas en inglés: Patient-Reported Outcome Measure) y NPS tras artroplastia de cadera, estableciendo umbrales de PROMs que identifican a los detractores con notable robustez.

De modo paralelo, Carretta (2022)<sup>4</sup> desarrolló un modelo de predicción de NPS en seguros de salud y decesos que incorpora datos de retención y satisfacción, aunque señalando la escasa atención al lenguaje emocional detrás de cada respuesta.

Minocha et al. (2023)<sup>5</sup> aplicaron análisis factorial y detección de anomalías para anticipar caídas del NPS en poblaciones sanitarias, mientras que Ruffing (2024)<sup>6</sup> demostró que variables como el trabajo en equipo impactan significativamente en la puntuación, sugiriendo regresiones más complejas para modelar estas interacciones.

### 3.1. Análisis de Sentimientos

Ese mismo impulso por comprender mejor la voz del cliente llevó a desplegar técnicas de análisis de sentimiento.

Aprovechando el auge de las redes sociales y blogs, el análisis de sentimiento prometía desentrañar opiniones, actitudes y percepciones, comprendiendo el tono de las declaraciones más allá de una simple polaridad.

En el sector de seguros, ya se ha empleado para evaluar las respuestas de los clientes, combinando encuestas, transcripciones y comentarios para una visión más profunda de la experiencia. De hecho, compañías como Allianz y AXA han empezado a integrar estas herramientas para analizar opiniones sobre la gestión de siniestros, detectando sentimientos positivos o negativos y proporcionando información relevante que las puntuaciones NPS por sí mismas no pueden reflejar.

La relevancia en los Centros de Atención Telefónica es particularmente alta, permitiendo clasificar automáticamente el tono de la conversación y correlacionarlo con el NPS para ajustar estrategias de mejora.

Sin embargo, la interpretación de sentimientos en contextos sensibles no estaba exenta de desafíos. Las emociones intensas de los clientes podían llevar a clasificaciones incorrectas, y el uso de jergas o expresiones específicas relacionadas con la pérdida y el duelo no era bien capturado por los enfoques tradicionales de análisis de texto.

Este desajuste semántico se convierte en un problema central: frases como “fallecido” o “lo siento”, que en otros contextos serían negativas, son parte del lenguaje profesional y neutral en el ámbito funerario. A esto se suma la pérdida de información paralingüística (tono, entonación) en las transcripciones, crucial para el sentimiento real, y la complejidad emocional multidimensional, donde coexisten gratitud y dolor simultáneamente.

Para superar estas limitaciones, las técnicas de análisis de sentimientos tuvieron que evolucionar.

### 3.2. Técnicas de Análisis de Sentimientos

Existen varios enfoques para realizar el análisis de sentimientos en las interacciones con clientes:

- **Modelos basados en diccionarios:** Utilizan listas de palabras con connotaciones positivas o negativas para determinar el sentimiento general de un texto.
- **Modelos de aprendizaje automático:** Técnicas como las Máquinas de Soporte Vectorial (SVM), árboles de decisión y Redes Neuronales Convolucionales (CNN) pueden entrenarse para identificar patrones en el texto que indican sentimientos.



- **Aprendizaje profundo (Deep Learning):** Las Redes Neuronales Recurrentes (RNN) y, en particular, las LSTMs (Long Short Term Memory), son muy efectivas para analizar secuencias de texto como transcripciones de llamadas, ya que pueden captar la evolución del tono emocional durante una conversación.

Introducidos originalmente en el innovador artículo “*Attention Is All You Need*” (Vaswani et al., 2017)<sup>7</sup>, los Transformers revolucionaron el Procesamiento del Lenguaje Natural (PLN), y son los que impulsan hoy en día los modelos de IA más avanzados, desde BERT hasta GPT.

Pero ¿qué hace a los Transformers tan poderosos? exploremos sus componentes clave:

- Mecanismo de Autoatención.
- Atención Multi-Head.
- Codificación Posicional.
- Cabeceras de Transformer.

### 3.3. ¿Por qué necesitábamos Transformers?

Antes de que existiesen los Transformers, los modelos de NLP se basaban en RNNs y LSTMs, que procesaban el texto de forma secuencial (palabra por palabra).

#### 3.3.1. Problemas con RNNs y LSTMs:

Este procesamiento se encontraba con una serie de problemas:

- Las dependencias a largo plazo eran difíciles de capturar debido al problema del desvanecimiento del gradiente.
- Procesaban las palabras secuencialmente, lo que los hacía lentos e ineficientes.
- Las relaciones entre palabras no se aprendían de forma dinámica, lo que conllevaba una pobre gestión de estructuras lingüísticas complejas.

Los Transformers resolvieron estos problemas al reemplazar el procesamiento secuencial por el cálculo en paralelo y al introducir el mecanismo de autoatención, que modela dinámicamente las relaciones entre palabras.

La autoatención resuelve estos problemas al considerar todas las palabras al mismo tiempo y asignar distintos niveles de importancia a cada palabra en relación con las demás y capturando el contexto de manera dinámica.

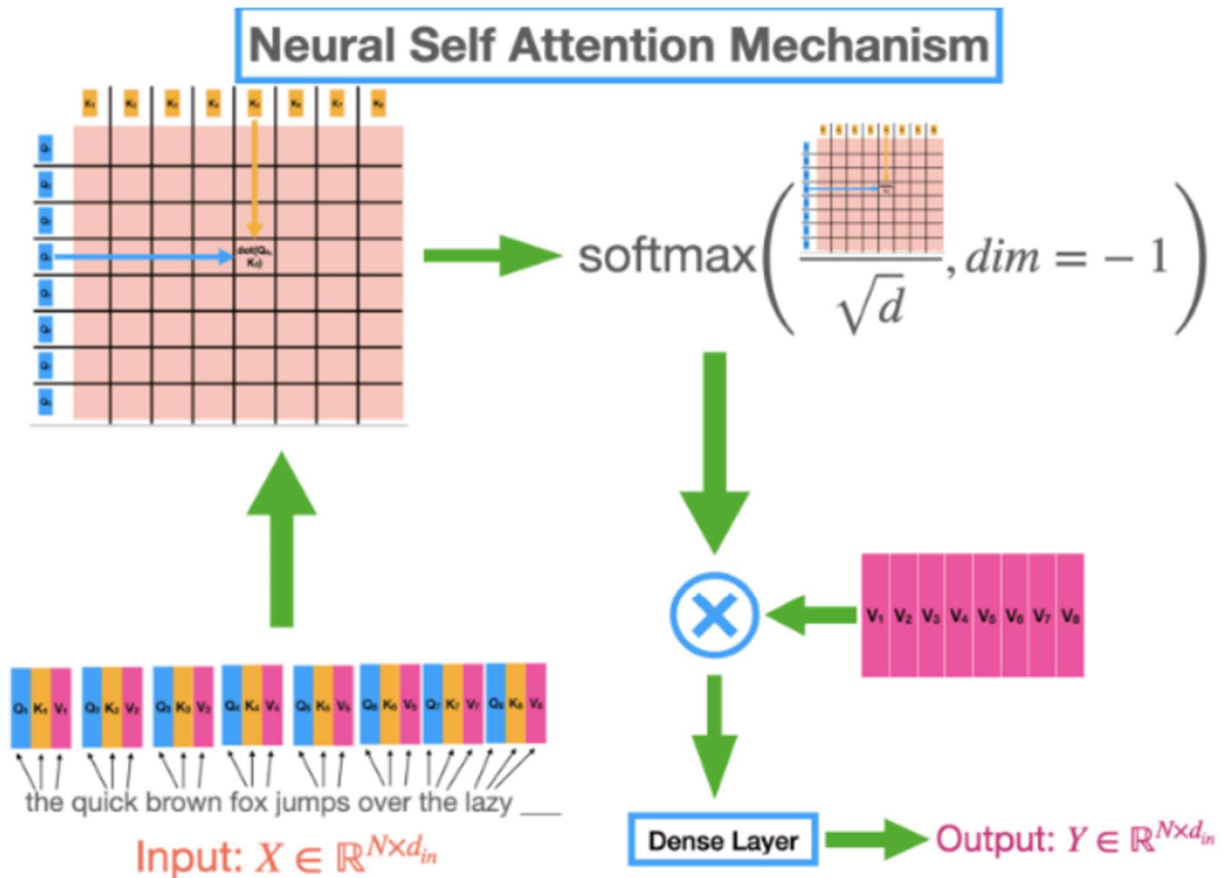


Ilustración 1. Mecanismo de Autoatención Neural

Fuente: <https://developer.aliyun.com/article/1462204>

### 3.3.2. Cómo funciona la Autoatención (ejemplo)

Consideremos la siguiente oración:

“El tigre saltó desde un árbol para beber agua porque tenía sed.”

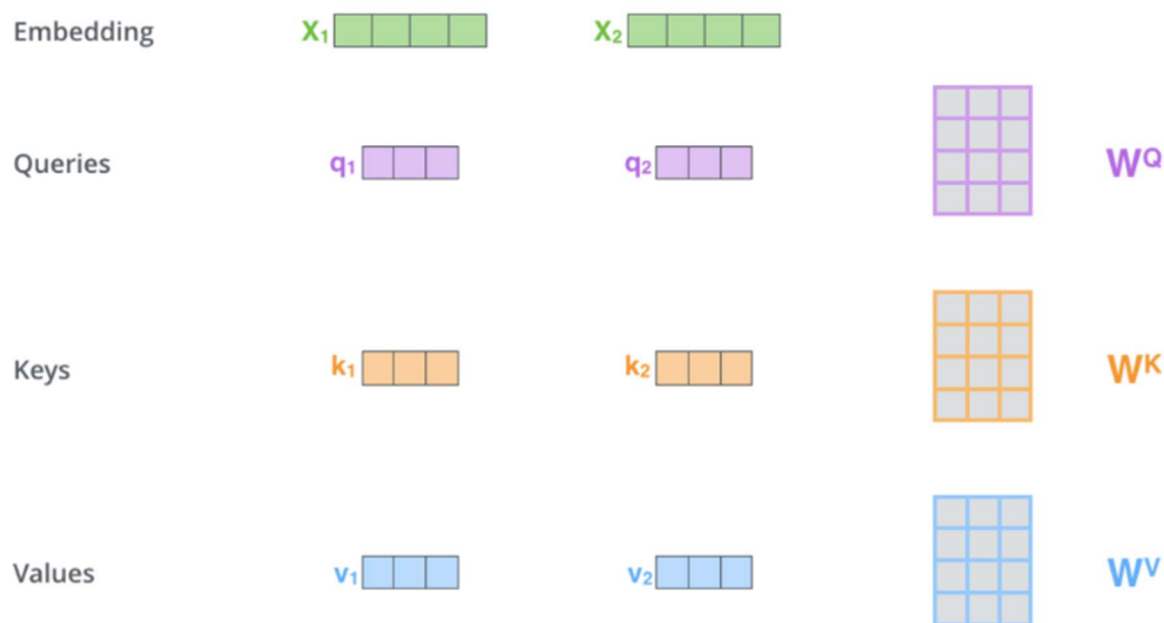
La autoatención ayuda a determinar las relaciones entre diferentes palabras y frases en las oraciones. Por ejemplo, en esta oración, “el tigre” y “tenía” (en referencia a “el tigre”) representan el mismo sujeto, por lo que esperaríamos que estas dos palabras estén fuertemente conectadas.

La autoatención logra esto a través de los siguientes pasos:

Cada embedding (vector de representación) de entrada se multiplica por tres matrices de pesos aprendidas ( $W_q$ ,  $W_k$ ,  $W_v$ ) para generar los vectores de: Query (Q) – Consulta  
Key (K) – Clave Value (V) – Valor

Estos vectores actúan como representaciones especializadas de cada palabra.

- **Query** (Consulta)
  - El vector de consulta ayuda al modelo a preguntar: “¿Qué otras palabras en la secuencia son relevantes para mí?”
- **Key** (Clave)
  - El vector clave actúa como una etiqueta que ayuda al modelo a identificar qué tan relevante puede ser una palabra para otras en la secuencia.
- **Value** (Valor)
  - El vector valor contiene la información real del contenido de la palabra.



Multiplying  $x_1$  by the  $W^Q$  weight matrix produces  $q_1$ , the "query" vector associated with that word. We end up creating a "query", a "key", and a "value" projection of each word in the input sentence.

*Ilustración 2. Autoatención en detalle*

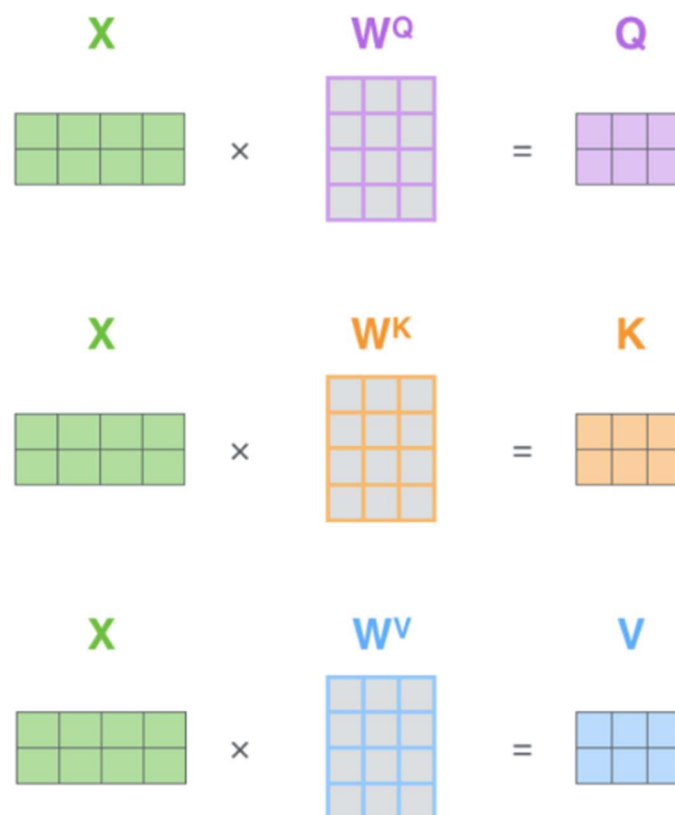
Fuente: <https://jalammar.github.io/illustrated-transformer/>

La Autoatención asigna puntuaciones dinámicas a cada par de palabras, permitiendo al modelo determinar las relaciones contextualmente.

### 3.3.3. Cálculo de las puntuaciones de atención

Las puntuaciones se calculan para determinar cuánto debe “atender” cada palabra a las demás palabras.

Esto se hace tomando el producto punto (dot product) del vector de consulta (query) de una palabra con los vectores clave (key) de todas las palabras en la secuencia.



Every row in the  $X$  matrix corresponds to a word in the input sentence. We again see the difference in size of the embedding vector (512, or 4 boxes in the figure), and the q/k/v vectors (64, or 3 boxes in the figure)

*Ilustración 3. Cálculo de la matriz de autoatención*

Fuente: <https://jalammar.github.io/illustrated-transformer/>

Impacto: La autoatención elimina la necesidad del procesamiento secuencial, haciendo que los Transformers sean significativamente más rápidos y eficientes que las RNNs.

### 3.3.4. Normalización - Aplicar Softmax a los pesos

Las puntuaciones se dividen por la raíz cuadrada de la dimensión del vector clave ( $d_k$ ) para mejorar la estabilidad numérica, y luego se pasan por una función softmax para obtener los pesos de atención. Estos pesos indican cómo de fuerte es la conexión de cada palabra con las demás.

$$\text{softmax} \left( \frac{
 \begin{matrix} \text{Q} \\ \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \end{matrix}
 \times
 \begin{matrix} \text{K}^T \\ \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \end{matrix}
 \right)
 \begin{matrix} \text{V} \\ \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \end{matrix}$$

$$= \begin{matrix} \text{Z} \\ \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \end{matrix}$$

The self-attention calculation in matrix form

*Ilustración 4. Forma de cálculo de la matriz de autoatención*

*Fuente: <https://jalammar.github.io/illustrated-transformer/>*

### 3.3.5. Cálculo de la suma ponderada de los valores

Cada vector de valor se multiplica por su correspondiente peso de atención. Los resultados se suman, generando una representación con conciencia del contexto para cada palabra.

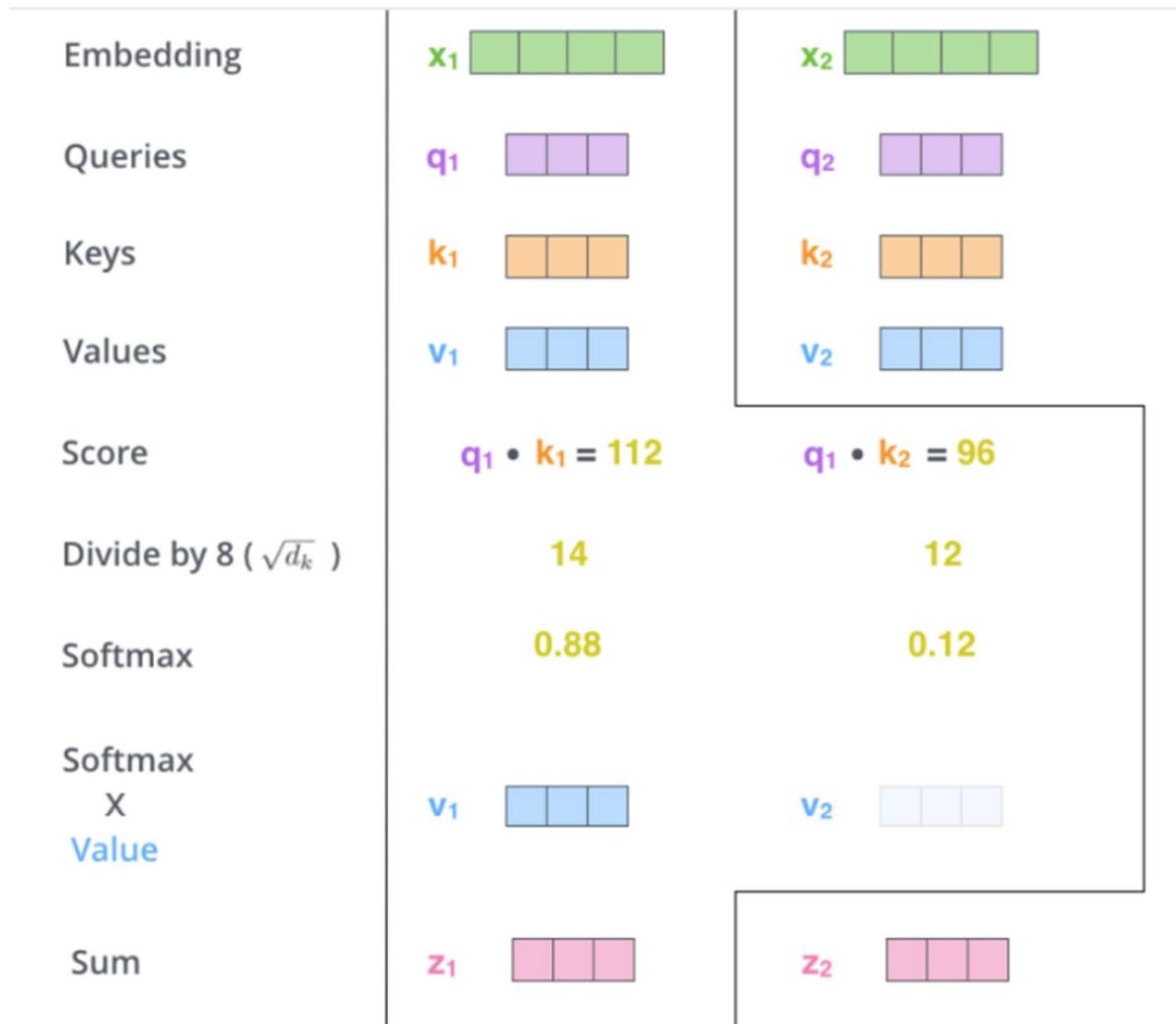


Ilustración 5. Cálculo de la suma ponderada de valores

Fuente: <https://jalammar.github.io/illustrated-transformer/>

Una forma sencilla de imaginarlo es:

- Cada palabra presta atención a todas las demás palabras.
- Las palabras importantes reciben más peso en la representación final.
- El modelo aprende dinámicamente qué palabras se relacionan entre sí.

### 3.3.6. Atención Multi-Head: Entendiendo el lenguaje desde distintas perspectivas

En lugar de tener una sola capa de atención, la atención Multi-Head divide la entrada en múltiples “cabeceras” (Heads) de atención, cada una capturando diferentes patrones lingüísticos.

Por ejemplo:

- **Head 1** → Se enfoca en la estructura sintáctica (ej., relaciones sujeto-verbo).
- **Head 2** → Captura el significado semántico (ej., sinónimos).
- **Head 3** → Detecta dependencias a largo plazo.

Cada Head procesa la entrada por separado, y sus salidas se concatenan y transforman en la representación final.

Sin la atención Multi-Head, los Transformers rendirían peor que las LSTMs.

Impacto: La atención Multi-Head permite que los Transformers aprendan múltiples relaciones entre palabras simultáneamente, mejorando la precisión y la eficiencia.

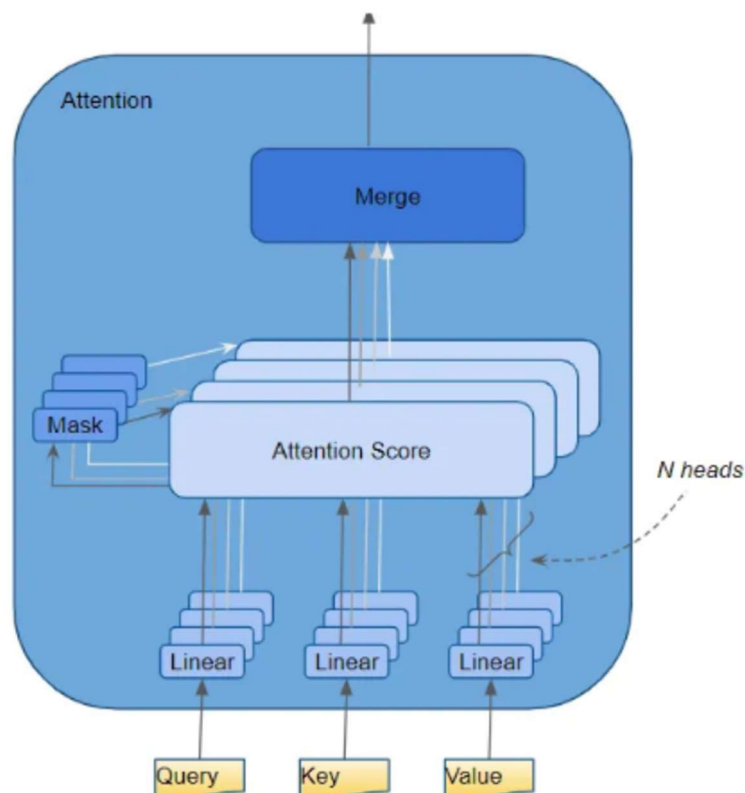


Ilustración 6. Mecanismo de atención de un modelo de transformación

Fuente: <https://towardsdatascience.com/transformers-explained-visually-part-3-multi-head-attention-deep-dive-1c1ff1024853/>

### 3.3.7. Codificación Posicional

Como los Transformers procesan palabras en paralelo, no entienden el orden de las palabras de forma inherente, por lo que se añade Codificación Posicional a cada embedding de palabra.

Cada posición de palabra (0, 1, 2, ...) recibe un patrón numérico único utilizando funciones seno y coseno.

### 3.3.8. Cabeceras de Transformer:

Los Transformers son modelos de propósito general, pero pueden ajustarse finamente para tareas específicas añadiendo cabeceras personalizadas.

**Tipos de Cabeceras de Transformer:**

- **Modelado de lenguaje enmascarado:** Predice palabras faltantes.
- **Cabecera de clasificación:** Categoriza texto.
- **Cabecera de Preguntas y Respuestas (QA):** Extrae respuestas de un texto.

Estas cabeceras permiten usar el mismo Transformer para múltiples aplicaciones de PLN.

Además de su uso en la traducción automática, los Transformers han mejorado significativamente el rendimiento en una amplia gama de tareas de PLN, como:

- El resumen de texto.
- El análisis de sentimientos.
- La respuesta a preguntas.
- La generación de lenguaje.

Estas innovaciones han llevado al desarrollo de modelos de vanguardia como BERT y GPT, que actualmente impulsan motores de búsqueda, chatbots y asistentes de IA en todo el mundo.

## 3.4. Los modelos de transformers más utilizados para el análisis de sentimiento son:

**BERT** (Bidirectional Encoder Representations from Transformers): Este modelo ha demostrado un rendimiento sobresaliente en la clasificación de sentimientos, especialmente en reseñas de productos y análisis de redes sociales.

**RoBERTa** (Robustly Optimized BERT Approach): Una versión optimizada de BERT que ha mostrado mejoras significativas en tareas de análisis de sentimiento.



**XLNet:** Un modelo que supera algunas limitaciones de BERT y ha logrado resultados destacados en varias tareas de procesamiento del lenguaje natural, incluyendo el análisis de sentimiento.

**DistilBERT:** Una versión más ligera y rápida de BERT que mantiene un buen rendimiento en tareas de clasificación de sentimiento, ideal para aplicaciones con recursos limitados.

**ALBERT** (A Lite BERT): Otra variante de BERT que reduce el número de parámetros sin sacrificar significativamente el rendimiento.

Estos modelos se destacan por su capacidad para capturar el contexto y las relaciones complejas entre palabras, lo que les permite comprender mejor las sutilezas y matices del lenguaje en el análisis de sentimiento. Su aplicación abarca desde el análisis de opiniones en redes sociales hasta la clasificación de reseñas de productos y la detección de emociones en diversos tipos de texto.

### 3.5. ¿Cómo se entrenan los modelos de transformers para el análisis de sentimiento en diferentes idiomas?

El entrenamiento de modelos Transformers para el análisis de sentimiento en diferentes idiomas implica varios pasos y consideraciones:

- **Pre-entrenamiento multilingüe:** Los modelos Transformers se pre-entrenan inicialmente con grandes corpus de texto en múltiples idiomas. Esto permite que el modelo aprenda representaciones generales del lenguaje y estructuras lingüísticas comunes.
- **Fine-tuning específico para sentimiento:** Después del pre-entrenamiento, el modelo se ajusta finamente (fine-tuning) para la tarea específica de análisis de sentimiento. Esto implica:
  - Añadir capas adicionales a la arquitectura del Transformer para adaptarlo a la clasificación de sentimientos.
  - Utilizar conjuntos de datos etiquetados con sentimientos en los idiomas objetivo.
  - Entrenar el modelo con estos datos para que aprenda a asociar patrones lingüísticos con etiquetas de sentimiento.
- **Uso de embeddings contextuales:** Los Transformers utilizan embeddings contextuales, que permiten capturar el significado de las palabras en función del contexto de la oración. Esto es crucial para entender matices y expresiones idiomáticas en diferentes lenguas.
- **Tokenización específica por idioma:** Se utilizan tokenizers adaptados a cada idioma para segmentar el texto de manera adecuada, respetando las particularidades lingüísticas de cada lengua.

- **Aumento de datos:** Para mejorar el rendimiento en idiomas con menos recursos, se pueden utilizar técnicas de aumento de datos, como la traducción automática de conjuntos de datos de sentimiento de idiomas con más recursos.
- **Evaluación multilingüe:** El modelo se evalúa en conjuntos de datos de prueba en diferentes idiomas para asegurar su rendimiento consistente a través de las lenguas.
- **Ajuste continuo:** El modelo puede seguir mejorándose con nuevos datos de sentimiento en diferentes idiomas, permitiendo una adaptación continua a nuevas expresiones y contextos lingüísticos.

Este enfoque permite crear modelos Transformers capaces de realizar análisis de sentimiento efectivo en múltiples idiomas, aprovechando su capacidad para capturar contextos y estructuras lingüísticas complejas.

### 3.6. Transformers en español para análisis de sentimiento de secuencias de texto

- **pysentimiento/robertuito-sentiment-analysis:** Este es un modelo basado en BERT, entrenado específicamente para el análisis de sentimientos en textos en español. Es una variante pequeña de BERT adaptada para documentos en español.
- **Littlejohn/analisis\_sentimientos:** Es un modelo BERT (bert-base-cased) afinado para el análisis de sentimientos en español, clasificando textos en dos categorías: positivo y negativo.
- **PySentimiento:** Es una librería de análisis de sentimiento en español que utiliza modelos de transformers para realizar la clasificación de emociones en textos.
- **BETO:** Es un modelo BERT entrenado en español por el DCC de la Universidad de Chile, que se puede adaptar para tareas de análisis de sentimiento.

Estos modelos pueden ser utilizados fácilmente con la biblioteca Transformers de Hugging Face.

### 3.7. Transformer más adecuado para una secuencia de 600 caracteres teniendo en cuenta que es en español

Para una secuencia de texto en español con una media de 600 palabras, el modelo Longformer sería el más adecuado. Longformer está diseñado específicamente para manejar secuencias largas de manera eficiente, lo que lo hace ideal para este caso.

Longformer utiliza un mecanismo de atención híbrida que combina diferentes tipos de atención, permitiéndole procesar secuencias de miles de tokens de manera eficaz. Esta característica es crucial para manejar textos largos como el manejado en este caso, que probablemente supere los 1000 tokens.

Además, Longformer reduce significativamente la carga computacional mientras mantiene la capacidad de modelar dependencias de largo alcance, lo cual es esencial para comprender el contexto completo en textos extensos.

En comparación, los transformers tradicionales y el Transformer-XL podrían enfrentar limitaciones con secuencias tan largas. Los transformers estándar tienen una complejidad computacional que crece cuadráticamente con la longitud de la secuencia, lo que los hace poco prácticos para textos muy largos.

Aunque el Transformer-XL mejora el manejo de contextos largos en comparación con los transformers estándar, el Longformer está específicamente optimizado para secuencias excepcionalmente largas, lo que lo hace más adecuado para este caso particular.

Longformer utiliza self-attention, pero de una manera modificada y más eficiente que los transformers tradicionales. Longformer emplea un mecanismo de atención local y global que le permite procesar secuencias de texto mucho más largas de manera eficiente.

Específicamente, Longformer utiliza:

- **Atención local:** Se enfoca en un contexto local alrededor de cada token, típicamente considerando una ventana de tokens a la izquierda y derecha.
- **Atención global:** Algunos tokens preseleccionados reciben atención global, permitiéndoles acceder a todos los demás tokens de la secuencia.

Esta combinación de atención local y global permite a Longformer manejar secuencias de miles de tokens de manera eficaz, reduciendo significativamente la carga computacional en comparación con la atención completa utilizada en transformers estándar.

El uso de estas matrices de atención dispersas permite que Longformer procese entradas con una longitud de secuencia mucho mayor que los transformers tradicionales, haciéndolo especialmente útil para tareas que involucran documentos largos.

### 3.8. Carencias y oportunidades de avance

A pesar de todos estos avances, la aplicación específica del análisis de sentimientos en el sector de seguros de decesos sigue siendo una frontera en gran parte inexplorada, enfrentando desafíos técnicos y metodológicos persistentes.

La mayoría de las investigaciones se han centrado en modelos genéricos, sin considerar el lenguaje técnico y emocional particular de este dominio, lo que afecta la precisión.

La coexistencia de múltiples estados emocionales (gratitud, dolor, ansiedad, frustración) genera una ambigüedad emocional multidimensional que los modelos tradicionales no pueden manejar adecuadamente. Además, un desafío fundamental y no explorado es cómo diferenciar el “ruido emocional” (emociones relacionadas con la pérdida personal) de las emociones directamente vinculadas a la calidad del servicio asegurador. Se suman otras problemáticas como el desbalanceo entre clases (la mayoría de las interacciones son neutras) y la necesidad de integrar métodos explicables de IA (XAI) para generar confianza en los resultados de estos “cajas negras”.

Las transcripciones automáticas también presentan sus propias peculiaridades: errores de reconocimiento, inconsistencias ortográficas, ausencia de puntuación y anonimización de datos.

En este punto, es evidente la necesidad de una propuesta innovadora que aborde estas brechas. Estudios previos en el sector asegurador no habían profundizado en las complejidades del análisis de sentimientos en servicios funerarios. Trabajos como los de García et al. (2021)<sup>8</sup> ya destacaban cómo el PLN y el aprendizaje automático podían identificar satisfacción y frustración en tiempo real en seguros de decesos, y Pérez y Sánchez (2020)<sup>9</sup> señalaban la mayor precisión de los enfoques híbridos en contextos de lenguaje técnico y emocional complejo. Fernández et al. (2022)<sup>10</sup> incluso desarrollaron modelos para detectar emociones en el sector funerario, facilitando intervenciones proactivas. Además, López y Ramírez (2019)<sup>11</sup> mostraron cómo el análisis de sentimientos complementa las métricas tradicionales, sugiriendo su integración en sistemas CRM, y Martínez y Gómez (2023)<sup>12</sup> demostraron el impacto estratégico en la reducción del tiempo de resolución de reclamaciones.

Este proyecto de fin de máster (TFM) se posiciona en esta intersección.

Al centrarse en la aplicación específica al sector de seguros de decesos, un dominio con un considerable vacío de investigación, desarrolla una solución adaptada a las particularidades semánticas y lingüísticas que los modelos genéricos no pueden manejar eficazmente. La metodología propuesta integra un preprocesamiento de datos riguroso (con carga de datos, limpieza textual, restauración de puntuación e identificación de interlocutores), la identificación de términos técnicos y embeddings específicos del dominio (usando spaCy, TF-IDF, Hugging Face NER y validación manual, además de aprendizaje contrastivo para distinguir usos neutrales profesionales de usos negativos reales), y un módulo central de modelado y clasificación de sentimientos basado en Longformer adaptado al español. Este enfoque no solo permite procesar eficientemente las transcripciones largas de 600 palabras, sino que también captura las sutilezas semánticas inherentes a conversaciones delicadas, donde el significado de las palabras está fuertemente ligado al contexto profesional.

De este modo, se busca transformar las transcripciones de llamadas, una fuente rica de datos cualitativos actualmente desaprovechada, en información procesable para la estimación del NPS, complementando las métricas cuantitativas tradicionales y mejorando la capacidad de reacción inmediata de la empresa.



Este estudio no solo sienta un precedente técnico y metodológico en un nicho de investigación de alta sensibilidad, sino que también abre puertas a futuras líneas de investigación, como la evolución hacia un análisis de sentimiento más profundo con detección de emociones finas y multidimensionales, la integración multimodal (combinando audio y texto), y la exploración de la explicabilidad de la IA para una mayor confianza.

## 4. Análisis y Estructura del Conjunto de Datos

### 4.1. Caracterización del Corpus de Transcripciones

El conjunto de datos analizado constituye un corpus especializado de comunicaciones orales transcritas que presenta características únicas derivadas tanto de su contexto de producción como de su proceso de obtención. A continuación, se describen sus principales características.

### 4.2. Composición y distribución del corpus

El corpus está formado por 129 transcripciones de conversaciones telefónicas reales entre clientes y operadores del Centro de Atención Telefónica del Grupo ASV, aseguradora especializada en seguros de decesos. Estas transcripciones están distribuidas de manera equilibrada en tres categorías de sentimiento:

- **Positivo\_\*.txt:** 43 transcripciones (33,3%) que reflejan satisfacción del cliente.
- **Negativo\_\*.txt:** 43 transcripciones (33,3%) que denotan insatisfacción o quejas.
- **Neutro\_\*.txt:** 43 transcripciones (33,3%) de carácter principalmente informativo.

Las transcripciones han sido generadas mediante la tecnología de reconocimiento automático de voz Whisper, un sistema basado en inteligencia artificial que transforma el audio de las llamadas en texto. Este proceso de transcripción automática introduce particularidades que deben considerarse en el procesamiento a realizar.

### 4.3. Métricas descriptivas del corpus

El análisis cuantitativo del corpus revela las siguientes características:

- **Duración media de las llamadas:** Entre 2 y 3 minutos por conversación.
- **Extensión textual:** Aproximadamente 600 palabras por transcripción (con variaciones significativas).
- **Ratio operador-cliente:** Generalmente equilibrado, con ligero predominio de intervenciones del operador.
- **Vocabulario especializado:** Presencia recurrente de terminología específica del sector funerario y asegurador.

El etiquetado de las transcripciones según la polaridad del sentimiento (positivo, negativo o neutro) ha sido realizado por evaluadores humanos, lo que añade un componente de subjetividad que debe tenerse en cuenta en el desarrollo de modelos automáticos de clasificación.

## 4.4. Análisis lingüístico y conversacional

Las transcripciones presentan características propias del lenguaje oral en contexto de conversación telefónica tipo atención al cliente, con particularidades adicionales derivadas del dominio específico de seguros de decesos y del proceso de transcripción automática.

## 4.5. Estructura y dinámica conversacional

El análisis de la estructura conversacional revela aspectos críticos para su procesamiento automático:

- **Ausencia de marcadores de turno:** Las transcripciones no delimitan explícitamente quién habla en cada momento, lo que dificulta la segmentación automática y la atribución de enunciados. Por ejemplo, en la transcripción neutra: “Bienvenido a Meridiano. Por su seguridad, esta llamada podrá ser grabada. [...] Hola, buenas. Es que mi abuela había fallecido, ya lo he comunicado”, no hay indicación de cambio de interlocutor.
- **Secuencias conversacionales identificables:** A pesar de la ausencia de marcadores explícitos, las conversaciones siguen patrones estructurales recurrentes que pueden segmentarse en unidades funcionales:
  - Apertura (saludos y presentación institucional).
  - Identificación del motivo de la llamada.
  - Desarrollo (intercambio de información y/o resolución de problemas).
  - Cierre (agradecimientos y despedida).
- **Solapamientos y fragmentos incompletos:** Presencia frecuente de enunciados interrumpidos, reformulaciones y rectificaciones propias del habla espontánea, como se observa en: “¿Le van a llamar ellos para indicarle cuánto tiempo aproximadamente van a tardar en llegar? Vale.”
- **Elementos paralingüísticos no representados:** La transcripción automática no captura pausas, entonaciones, énfasis o volumen, elementos que pueden ser decisivos en la interpretación del sentimiento.

## 4.6. Particularidades léxico-semánticas

El análisis del vocabulario y expresiones recurrentes revela:

### 4.6.1. Terminología especializada del dominio

Las transcripciones contienen abundante vocabulario técnico relacionado con los seguros de decesos y servicios funerarios:

- **Términos del ámbito funerario:** “velatorio”, “incineración”, “entierro”, “sanatorio”.
- **Vocabulario asegurador:** “póliza”, “cobertura”, “servicio de gestión”.
- **Referencias a gestiones administrativas:** “trámites”, “documentación”.

### 4.6.2. Fórmulas protocolarias y rutinas conversacionales

El corpus presenta un alto grado de formalidad, con expresiones estandarizadas recurrentes:

- **Expresiones de condolencia:** “permítame trasladarle mis más sinceras condolencias”, “lamentamos la pérdida”.
- **Fórmulas de disponibilidad:** “quedamos a su disposición”, “estamos las 24 horas, por si nos necesita”.
- **Rutinas de verificación:** “¿es correcto?”, “¿es así?”.

### 4.6.3. Marcadores pragmáticos y de cortesía

Uso frecuente de elementos que gestionan la relación interpersonal:

- **Tratamientos formales:** “doña”, “don” seguidos del nombre.
- **Mitigadores:** “un momentito”, “si no le molesta”.
- **Agradecimientos ritualizados:** “muchas gracias”, “gracias a ustedes”.

## 4.7. Rasgos derivados de la transcripción automática

El uso de tecnología de reconocimiento de voz introduce características específicas que afectan a la calidad textual:

- **Errores de reconocimiento:** Palabras mal transcritas, especialmente en nombres propios, direcciones y términos técnicos menos frecuentes.
- **Inconsistencias ortográficas:** Ausencia parcial o total de signos de puntuación, acentuación irregular y errores en la representación de caracteres especiales.



- **Anonimización:** Sustitución de datos personales por secuencias “XXXXX” o “XXXXXX”, que afecta a la continuidad textual.

## 4.8. Patrones Discursivos por Categoría de Sentimiento

El análisis comparativo de las transcripciones según su clasificación de sentimiento revela patrones lingüísticos y conversacionales distintivos:

### 4.8.1. Indicadores de sentimiento positivo

Las transcripciones clasificadas como positivas presentan los siguientes rasgos distintivos:

- **Expresiones explícitas de satisfacción:** Uso de evaluadores positivos directos como “fenomenal”, “muy bien”, “todo correctamente”, “todo perfecto”.
- **Agradecimiento reiterado y enfático:** Secuencias de agradecimiento múltiple como se observa en la transcripción positiva: “Muchas gracias. Gracias a ustedes. Un saludo”.
- **Confirmación de expectativas cumplidas:** Afirmaciones que validan el servicio recibido.
- **Tono colaborativo:** Disposición a facilitar información adicional y a seguir el protocolo establecido.
- **Fluidez conversacional:** Menor presencia de interrupciones, solapamientos o reformulaciones.

### 4.8.2. Marcadores de sentimiento negativo

Las transcripciones con sentimiento negativo se caracterizan por:

- **Marcadores de disconformidad:** Expresiones de insatisfacción, generalmente relacionadas con tiempos de respuesta o gestiones incompletas.
- **Estructuras interrogativas reiteradas:** Preguntas que cuestionan procedimientos o buscan aclaraciones sobre situaciones problemáticas.
- **Repetición de problemas:** Insistencia en aspectos no resueltos o en dificultades encontradas.
- **Interrupciones y solapamientos:** Mayor frecuencia de intervenciones simultáneas y ruptura de turnos conversacionales.
- **Intensificadores negativos:** Adverbios que enfatizan problemas o deficiencias.

### 4.8.3. Características de las transcripciones neutras

Las transcripciones clasificadas como neutras muestran:

- **Predominio de contenido informativo:** Intercambio de datos factuales sin carga emocional evidente, como en la transcripción neutra donde se consulta un cambio de lugar para el velatorio.
- **Estructura pregunta-respuesta:** Secuencias organizadas de solicitud y provisión de información.
- **Verificación de datos:** Confirmación de información personal, detalles de pólizas o procedimientos.
- **Menor uso de intensificadores:** Ausencia relativa de adverbios de intensidad o expresiones enfáticas.
- **Foco en procedimientos:** Centrado en la descripción de procesos y pasos a seguir.

### 4.9. Problemas relacionados con la calidad de los datos

Las transcripciones automáticas presentan desafíos inherentes a su naturaleza:

- **Variabilidad en la calidad de transcripción:** Inconsistencias entre archivos dependiendo de factores como la calidad del audio original, acentos regionales o velocidad del habla.
- **Errores de etiquetado:** Discrepancias entre la categoría asignada y el contenido emocional real de algunas conversaciones, como se observa en la transcripción “negativa” proporcionada, que no muestra indicadores claros de negatividad.
- **Datos personales anonimizados:** Presencia de marcadores “XXXXX” que interrumpen la continuidad textual y pueden afectar al análisis contextual.
- **Pérdida de información prosódica:** Ausencia de elementos como entonación, volumen o pausas que son cruciales para la interpretación del sentimiento.

### 4.10. Retos lingüísticos específicos

El análisis lingüístico revela desafíos particulares:

- **Ambigüedad contextual:** Expresiones cuya interpretación depende de factores no explícitos en el texto.
- **Ironía y cortesía formal:** Dificultad para distinguir entre cortesía protocolar y satisfacción genuina.

- **Discontinuidad temática:** Saltos conversacionales y cambios de tema abruptos.
- **Sintaxis fragmentada:** Oraciones incompletas o sintácticamente irregulares propias del lenguaje oral.

## 4.11. Desafíos para la modelización con Deep Learning

Para la implementación de modelos de Deep Learning, se identifican los siguientes desafíos específicos:

- **Segmentación de turnos de habla:** Necesidad de identificar automáticamente quién habla en cada momento.
- **Longitud variable:** Transcripciones de extensión heterogénea que requieren estrategias de normalización.
- **Desequilibrio de participación:** Predominio de intervenciones del agente en algunas conversaciones.
- **Información contextual implícita:** Contenido emocional expresado de manera indirecta.
- **Ruido textual:** Elementos irrelevantes para el análisis de sentimiento que pueden interferir en el entrenamiento de modelos.

## 4.12. Estrategia de Preprocesamiento Propuesta

Para abordar los desafíos identificados, la propuesta es una metodología de preprocesamiento específica para este corpus, orientada a optimizar el rendimiento de los modelos de PLN y Deep Learning:

### 4.12.1. Normalización textual adaptativa

- **Corrección ortográfica contextualizada:** Implementación de algoritmos de corrección específicos para errores frecuentes en transcripciones de reconocimiento de voz, preservando la terminología especializada.
- **Homogeneización terminológica:** Normalización de variantes de términos técnicos para reducir la dispersión léxica.
- **Tratamiento de entidades nombradas:** Identificación y manejo uniforme de nombres, direcciones y otros datos personales anonimizados.
- **Restauración de signos de puntuación:** Aplicación de modelos especializados para insertar puntuación adecuada que facilite el procesamiento sintáctico posterior.

#### 4.12.2. Segmentación y estructuración conversacional

- **Identificación automática de turnos:** Desarrollo de algoritmos para delimitar las intervenciones de cada interlocutor basados en patrones lingüísticos y discursivos.
- **Detección de secuencias funcionales:** Segmentación del discurso en unidades como saludo, presentación, exposición del problema, resolución y despedida.
- **Alineación de pares adyacentes:** Identificación de secuencias pregunta-respuesta relacionadas.
- **Reconstrucción de coherencia temática:** Agrupación de fragmentos relacionados con un mismo tópico o subtópico conversacional.

#### 4.12.3. Extracción de características específicas para análisis de sentimiento

- **Marcadores explícitos de emoción:** Identificación de léxico con carga emocional directa.
- **Patrones sintácticos indicativos:** Análisis de estructuras oracionales asociadas a diferentes estados emocionales.
- **Características pragmáticas:** Extracción de indicadores de cortesía, formalidad y otros rasgos discursivos.
- **Indicadores de fluidez conversacional:** Cuantificación de interrupciones, solapamientos y reformulaciones como posibles marcadores de satisfacción/insatisfacción.

#### 4.12.4. Consideraciones para el Modelado con Deep Learning

El análisis del corpus y sus características sugiere las siguientes consideraciones para la selección y aplicación de técnicas de Deep Learning:

#### 4.12.5. Arquitecturas prometedoras para el corpus

- **Modelos basados en transformers:** Pre-entrenados en español y ajustados al dominio específico para capturar relaciones contextuales complejas.
- **Redes recurrentes bidireccionales:** Para captar dependencias de largo alcance en la estructura conversacional.
- **Modelos híbridos con atención:** Que permitan focalizar en elementos lingüísticos relevantes para la clasificación de sentimiento.

#### 4.13. Estrategias de representación textual

- **Embeddings contextuales específicos del dominio:** Adaptados a la terminología y expresiones propias del sector.
- **Representaciones secuenciales de diálogo:** Que preserven la estructura conversacional y la alternancia de turnos.
- **Incorporación de características pragmáticas:** Como complemento a las representaciones puramente léxico-sintácticas.

## 5. Materiales y métodos

Abordar un Trabajo Fin de Máster en el ámbito de la Inteligencia Artificial, requiere una metodología rigurosa. Por ello, la sección de materiales y métodos se convierte en el pilar fundamental que sustenta la validez y replicabilidad de la investigación.

A continuación, se detalla la planificación del TFM y las fases asociadas al proceso, todo ello bajo el paraguas de un enfoque sistemático y bien estructurado.

El presente trabajo fin de máster se ha diseñado siguiendo una metodología sólida, el modelo *Cross-Industry Standard Process for Data Mining (CRISP-DM)* que proporciona un marco iterativo y flexible que guía el proyecto desde la comprensión inicial del problema de negocio hasta el despliegue de la solución.

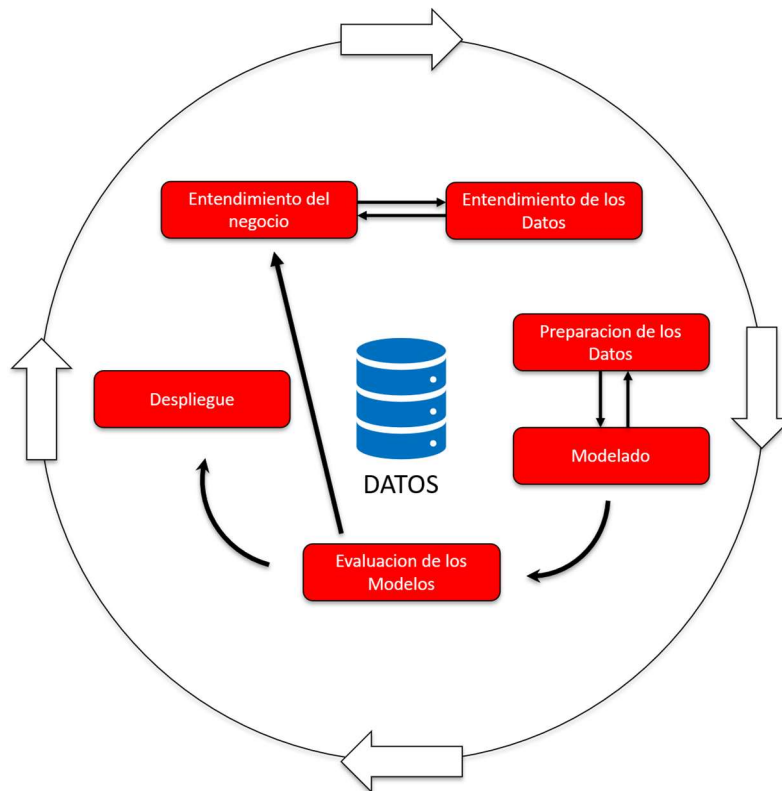
La planificación de este TFM se estructura en torno a las fases secuenciales, pero inherentemente iterativas, del modelo CRISP-DM. Esta aproximación garantiza una visión holística del proyecto, permitiendo que las lecciones aprendidas en etapas posteriores refinan las fases previas.

La elección de CRISP-DM es crucial para este proyecto por varias razones:

- **Alineación Estratégica:** Permite definir claramente los objetivos de negocio desde el inicio, asegurando que el desarrollo técnico contribuya directamente a resolver el problema.
- **Gestión de la Complejidad:** Facilita la descomposición de un problema complejo de procesamiento de lenguaje natural (PLN) en etapas manejables, desde la preparación de datos hasta el modelado y la evaluación.
- **Transparencia y Justificación:** Cada fase del TFM se documenta y justifica. Esto permite mostrar cómo se abordan los retos específicos del dominio de seguros de decesos, como la ambigüedad semántica y la naturaleza sensible de las conversaciones.
- **Orientación a Resultados:** El énfasis en la evaluación y el despliegue asegura que el TFM no solo produzca un modelo, sino una solución potencialmente aplicable que facilite un análisis más ágil, preciso y económico del sentimiento del cliente.

## 5.1. Fases asociadas al proceso (CRISP-DM)

La implementación del proyecto se ha dividido en las seis fases principales de CRISP-DM, adaptadas a las particularidades del problema planteado.



*Ilustración 7. CRISP DM*

*Fuente: Elaboración propia*

### 5.1.1. Entendimiento del negocio

En esta fase inicial, se ha definido el problema central y los objetivos del TFM. Actualmente, se utiliza una metodología externa (Net Promoter Score (NPS), mediante encuestas telefónicas anuales) para medir la satisfacción del cliente, lo que resulta costoso y no cubre todos los puntos de contacto. Existe un desaprovechamiento de la información cualitativa contenida en las transcripciones de las llamadas telefónicas al centro de atención al cliente, debido a la falta de herramientas internas para un análisis automático del sentimiento.

El objetivo general es, por tanto, evaluar la viabilidad de utilizar técnicas de deep learning para estimar automáticamente el NPS a partir de estas transcripciones. Los objetivos específicos incluyen analizar el estado del arte en PLN para casos similares, desarrollar un modelo basado en deep learning y facilitar un análisis ágil y económico del sentimiento del cliente. La justificación reside en la necesidad de una medición integral de la satisfacción, la reducción de costes, la mejora de la capacidad de reacción y el aprovechamiento de datos infrautilizados.

El alcance se circunscribe al análisis de transcripciones textuales de llamadas del Grupo ASV en el ramo de seguros de decesos. Se espera que el sistema sea más preciso, interprete correctamente términos sensibles y capture mejor los sentimientos globales.

## 5.1.2. Entendimiento de los Datos

En esta fase el foco estaba centrado en conocer a fondo el conjunto de datos de transcripciones. El corpus está compuesto por 129 transcripciones de conversaciones telefónicas reales entre clientes y agentes, etiquetadas como positivas, neutras o negativas. La longitud promedio es de 600 palabras, lo que plantea un desafío técnico para muchos modelos PLN estándar.

Los desafíos clave identificados en esta fase incluyen:

- **Naturaleza Sensible del Dominio:** Las conversaciones sobre seguros de decesos contienen términos y expresiones (“fallecido”, “lamento su pérdida”) que, fuera de este contexto, se clasificarían como negativos, pero aquí son parte del lenguaje profesional neutro.
- **Pérdida de Información Paralingüística:** Las transcripciones carecen de tono e intención, cruciales para interpretar sentimientos, y la puntuación es a menudo deficiente o ausente.
- **Subjetividad del Etiquetado Humano:** Las transcripciones han sido etiquetadas manualmente, introduciendo un componente de subjetividad que debe ser considerada.
- **Particularidades Lingüísticas:** Presencia de eufemismos, carga emocional implícita, terminología técnica sensible y patrones discursivos específicos del sector funerario y asegurador.
- **Problemas de Calidad de Transcripción:** Errores de reconocimiento, inconsistencias ortográficas y la presencia de marcadores de anonimización que interrumpen la continuidad textual.



### 5.1.3. Preparación de los Datos

Esta es una fase crítica, ya que la calidad de los datos de entrada impacta directamente en el rendimiento del modelo.

Para abordar las dificultades identificadas en la fase de comprensión de datos, se implementarán estrategias de preprocesamiento específicas:

- **Normalización Textual Adaptativa:**
  - **Corrección Ortográfica Contextualizada:** Se aplicarán algoritmos de corrección específicos para errores frecuentes en transcripciones automáticas, preservando la terminología especializada del dominio.
  - **Restauración de Puntuación:** Dada la importancia de la puntuación en español para delimitar el significado y los sentimientos, se implementará un sistema basado en reglas para corregir deficiencias y proporcionar señales sintácticas claras. Modelos como RoBERTa-bne, pre-entrenados con datos de la Biblioteca Nacional de España, han mostrado resultados prometedores en restauración de puntuación para el español.
  - **Eliminación de Repeticiones y Caracteres Innecesarios:** Se realizará una limpieza para eliminar repeticiones de palabras o signos y caracteres no alfanuméricos que no sean puntuación esencial.
  - **Manejo de Entidades Anonimizadas:** Identificación y tratamiento uniforme de datos personales anonimizados para mantener la continuidad textual.
- **Segmentación y Estructuración Conversacional:** Se desarrollará un “Módulo de Identificación de Interlocutores” que utilizará técnicas de PLN avanzadas, incluyendo tokenización (con la biblioteca transformers), para clasificar y etiquetar los turnos de conversación como “Cliente” o “Agente”. Esto es crucial debido a la ausencia de marcadores explícitos de turno en las transcripciones originales.
- **Creación de Embeddings Especializados:** Se entrenarán embeddings contextuales específicos del dominio funerario. Esto es fundamental para que el modelo aprenda a interpretar frases como “lamento su pérdida” como neutras en el contexto profesional. Se planifica entrenar con un corpus especializado, usar aprendizaje contrastivo y adaptación continua.
- **Ampliación del Vocabulario del Tokenizer:** Para mejorar la representación de términos técnicos del dominio (como “póliza”, “beneficiario”), se ampliará el vocabulario del tokenizer de Longformer. Esto se logrará identificando términos únicos del *dataset* que no estén ya en el vocabulario del tokenizer, utilizando enfoques como spaCy + TF-IDF para extraer frases nominales y entidades relevantes, KeyBERT para palabras clave basadas en embeddings, o pipelines de Hugging Face para Extracción de Entidades Nombradas (NER). La validación manual de estos términos con expertos del dominio será un paso esencial.

### 5.1.4. Modelado

En esta fase, se seleccionarán y entrenarán los algoritmos de Deep Learning. La tarea principal es la clasificación de sentimientos (positivo, neutro, negativo).

El modelo central propuesto es Longformer adaptado al español. La elección de Longformer se justifica por su capacidad para manejar secuencias largas (transcripciones de 600 palabras) de manera eficiente, superando las limitaciones computacionales cuadráticas de los *transformers* estándar. Longformer utiliza un mecanismo de atención híbrida (local y global) que reduce la complejidad computacional a niveles lineales, permitiendo mantener una comprensión global del texto en interacciones extensas.

El proceso de modelado incluye:

- **Fine-tuning del Modelo:** Se trata de un ajuste fino (fine-tuning) de un modelo Longformer pre-entrenado para el español. Esto implica adaptar el modelo a las características lingüísticas específicas del español y al dominio de seguros de decesos. Se busca integrar los embeddings especializados creados en la fase de preparación de datos para mejorar la representación inicial del texto.
- **Estrategia de Entrenamiento:** Para una adaptación temprana al dominio y evitar el sobreajuste inicial, se congelan todas las capas del modelo excepto los embeddings al inicio del *fine-tuning*, permitiendo que el modelo aprenda representaciones específicas de términos técnicos antes de ajustar parámetros complejos.
- **Manejo del Desbalance de Clases:** El entrenamiento aborda el problema del desbalanceo entre clases (las interacciones neutras suelen ser mayoría) mediante técnicas como funciones de pérdida ponderadas o generación de ejemplos sintéticos.
- **Enfoques de Deep Learning:** Además de Longformer, el estudio se basa en los avances de arquitecturas *Transformer* en general (como BERT, RoBERTa, T5) que han revolucionado el PLN y son adecuadas para el análisis de sentimientos.

### 5.1.5. Despliegue

Aunque el despliegue completo de un sistema en un TFM suele ser una dirección futura, esta fase conceptualiza cómo los resultados del modelo se integrarían en un entorno real y cómo se mantendría.

- **Integración del Sistema:** El objetivo es que la solución permita un análisis inmediato y ágil de la información, facilitando la toma de decisiones rápidas. El *pipeline* modular propuesto facilita la escalabilidad y adaptabilidad, permitiendo futuras incorporaciones o mejoras.

- **Monitorización y Mantenimiento:** Se prevé que el sistema desarrollado pueda ser monitorizado continuamente para medir su desempeño real y ser actualizado o reentrenado periódicamente con nuevos datos.

## 5.2. Planificación del TFM

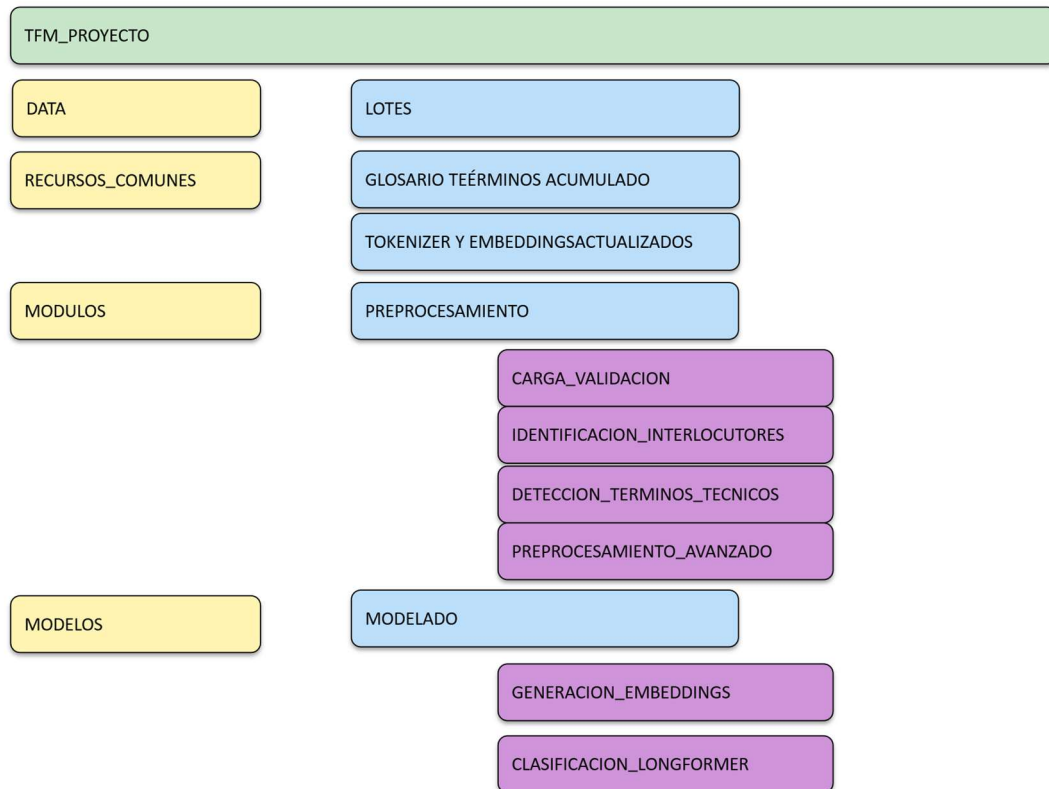
La planificación está estructurada con un enfoque iterativo, priorizando la modularidad y la adaptabilidad del sistema ante la recepción de nuevos lotes de datos, garantizando su escalabilidad a largo plazo.

La planificación general incluye:

- **Análisis inicial del dominio:** Comprensión del contexto de negocio (seguros de decesos), estudio del tipo de lenguaje usado en las conversaciones y detección de necesidades.
- **Diseño modular del flujo de trabajo:** Definición de submódulos reutilizables y auto explicativos.
- **Implementación por fases:** Cada módulo del sistema fue desarrollado, probado y validado de manera incremental.
- **Integración de módulos:** Ensamblado del sistema completo con pruebas de extremo a extremo. Cada hito se ha documentado mediante logs automáticos y versiones controladas de los datasets, lo que ha permitido trazabilidad y control de calidad durante todo el desarrollo.

### 5.3. Flujo del proceso

Este flujo, que es modular y está adaptado a necesidades reales y futuras, constituye una base sólida desde la cual se desarrollará cada módulo, que serán diseñados con el objetivo de que el proyecto sea mantenible y escalable.

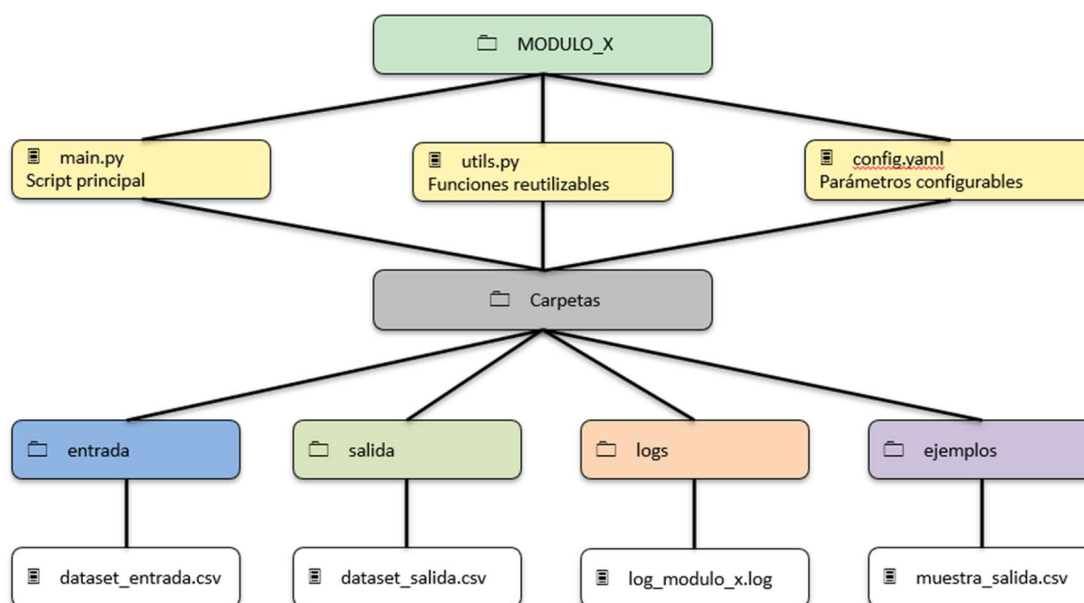


*Ilustración 8. Flujo del proceso*

*Fuente: Elaboración propia*

## 5.4. Estructura común de los módulos de preprocesamiento

Con objeto de que los módulos que componen el preprocesamiento de datos sean reutilizables y muy importante: fáciles de entender, revisar y depurar, se diseña esta estructura común:



*Ilustración 9. Estructura común de los módulos*

*Fuente: Elaboración propia*

### 5.4.1. MÓDULO 01: Carga de datos

Carga lotes de ficheros de texto que contienen las transcripciones ya anonimizadas.

Valida formatos, codificación (UTF-8), caracteres especiales, duplicados.

#### Pasos funcionales:

- **Carga y unificación de fuentes:** El módulo recibe como entrada archivos de texto. Se detecta el formato de cada fichero y se convierte todo a un único estándar interno (UTF-8 con saltos de línea “\n”).
- **Filtrado de contenido irrelevante:** Se eliminan líneas vacías o con menos de tres caracteres, evitando procesar registros sin contenido real y se descartan metadatos o marcas de tiempo que no vayan a ser usados en los siguientes pasos.

- **Normalización básica del texto:** Se pasa todo a minúsculas para garantizar que “Cliente” y “cliente” se traten igual. Se suprimen acentos y caracteres no alfabéticos (salvo puntuación mínima), de manera que el vocabulario quede reducido a letras y signos esenciales. Se ajustan espacios múltiples a uno solo, y se recortan espacios al inicio y final de cada línea.

**Salida.** El resultado es un dataset listo para usarse en el resto de los módulos, junto con un registro de versiones para reproducibilidad que está relacionado con el número de lote que procesamos.

Se garantiza que las transcripciones salientes estén libres de ruido (formatos, espacios, metadatos) y estructuradas en fragmentos de texto coherentes; lo justo para poder, en el siguiente paso, asignar quién habla sin interferencias.

## 5.4.2. MÓDULO 02: Identificación de interlocutores

Etiqueta automáticamente cada intervención como Agente o Cliente. Añade a cada fragmento de texto la etiqueta correspondiente al hablante (“cliente” vs. “agente”), proporcionando el mínimo contexto imprescindible para cualquier análisis posterior de sentimiento o término.

### **Razón de ser e importancia:**

Sin saber quién dice cada frase, las transcripciones pierden todo contexto: no podemos distinguir la voz del cliente de la del agente, ni asignar opiniones, ni extraer indicadores de satisfacción. Este paso es, por tanto, indispensable para convertir simples cadenas de texto en un diálogo estructurado.

### **Pasos funcionales y técnicos:**

Se usa un “prompt” específico, adaptado según el entorno (colab o local), que indica claramente al modelo que etiquete las frases de la conversación.

Es importante notar que incluso en el establecimiento del prompt se ha de tener cuidado y será distinto en local o en colab, ya que su longitud influye en la ejecución del modelo por su limitación en los tokens que procesa.

En primer lugar, se utilizó localmente con un modelo más ligero (“gpt2”) para verificar la viabilidad del proceso. Una vez comprobado que funcionaba correctamente para la tarea básica, se estableció un “prompt” más completo y con instrucciones más ricas que se trasladó al entorno de Google Colab para emplear con el modelo definitivo, más potente, pero con mayores requisitos computacionales.

El modelo seleccionado fue un modelo avanzado basado en redes neuronales llamado “meta-llama/Meta-Llama-3-8B-Instruct”, obtenido a través de Hugging Face. Este modelo realiza la tarea de identificar los interlocutores convirtiendo cada transcripción en una petición al modelo (pipeline de Hugging Face) y procesa cada transcripción individualmente, extrayendo el resultado etiquetado claramente separado del texto original. Este proceso incluye:

- Gestión de tokens de autenticación.
- Manejo cuidadoso de memoria GPU y CPU.
- Adaptación técnica de prompts para garantizar que el modelo entendiera perfectamente lo que se esperaba de él.

Durante esta etapa, fue fundamental la lectura detallada de la documentación oficial, tutoriales y foros especializados de Hugging Face, lo que permitió seleccionar el modelo adecuado, configurar correctamente el pipeline y solventar problemas técnicos iniciales.

#### **Justificación de la elección del modelo:**

Tras analizar diversas alternativas en Hugging Face, varios modelos de clasificación tradicionales fueron descartados, debido a su necesidad de datos etiquetados y/o sus limitaciones en contextos complejos y largos. Finalmente se seleccionó el modelo generativo “meta-llama/Meta-Llama-3-8B-Instruct” por su superior capacidad para interpretar contextos complejos, generar respuestas coherentes y adaptarse mediante prompts específicos. Este modelo se ejecutó antes de la limpieza profunda del texto para preservar el contexto original, esencial para su precisión. Además, el prompt utilizado asegura explícitamente que el modelo no inventa texto nuevo, garantizando así resultados de calidad.

**Salida.** Un dataset con las columnas:

- **Nombre fichero:** Nombre del fichero original de la transcripción anonimizada.
- **Etiqueta:** Positivo, neutro o negativo.
- **Transcripción:** Texto original de la transcripción.
- **Texto etiquetado:** Transcripción enriquecida con sus interlocutores (cliente/agente).

Este archivo, ya dotado de la identificación del interlocutor que está interviniendo, es la base imprescindible para los módulos de extracción de términos y ajuste de modelo.

Con esto se asegura que cada frase tenga el formato justo y el interlocutor correcto antes de avanzar al análisis de contenido; sin esta base, el resto del proceso perdería rigor y exactitud, incluso resultaría casi imposible de clasificar.

### 5.4.3. MÓDULO 03: Creación y gestión del glosario de términos

Recopila términos específicos del dominio de seguros de decesos que guarda en un fichero llamado “glosario de términos” y que mantendremos actualizado con nuevos términos según se vayan detectando en cada lote.

El objetivo es obtener de forma automática un listado de palabras y expresiones clave que aparecen en las transcripciones de llamadas, para construir un vocabulario específico del dominio.

#### **Pasos funcionales:**

El módulo comienza con una preparación básica del texto, incluyendo la conversión a minúsculas, la eliminación de espacios excesivos y filas vacías, asegurando una normalización adecuada para su posterior análisis.

Seguidamente, el módulo hace uso de la librería especializada “spaCy”, en su modelo de idioma español (es\_core\_news\_sm), para realizar una obtención precisa de entidades nombradas y sintagmas nominales. Esta extracción es clave para identificar términos potencialmente relevantes en un contexto altamente especializado como es el de seguros de decesos, donde se requiere precisión en la identificación de terminología específica.

Una vez realizada esta separación inicial, los términos obtenidos se someten a un análisis basado en la técnica “Term Frequency-Inverse Document Frequency (TF-IDF)” mediante la librería “scikit-learn”. TF-IDF permite evaluar la relevancia de cada término dentro del conjunto de transcripciones, asignando puntuaciones más altas a aquellos términos que aparecen con frecuencia pero que son menos comunes en el resto del corpus, destacando así los términos más distintivos y específicos.

El proceso continúa con un refinado adicional mediante la eliminación de términos considerados como “palabras vacías” (stop-words), utilizando una combinación ampliada de listas de palabras vacías de librerías como “spacy” y “stop words”. Este filtrado asegura que el glosario resultante incluya únicamente términos con alta relevancia contextual y técnica.

Además, se lleva a cabo un filtrado sintáctico basado en la categoría gramatical principal de los términos, conservando solamente aquellos cuya raíz corresponda a sustantivos comunes o propios (NOUN, PROPN). Este paso utiliza también las capacidades avanzadas de spaCy para análisis morfosintáctico.



#### **Justificación de la elección de librerías:**

Se optó por la librería especializada spaCy por su modelo de idioma español (es\_core\_news\_sm) ya preentrenado y adaptado a particularidades lingüísticas del español, y porque ofrece una capacidad probada en la extracción automática y precisa de entidades nombradas (NER) y sintagmas nominales, garantizando la identificación correcta de términos relevantes en un corpus especializado.

Adicionalmente ofrece una integración sencilla con otras librerías, como scikit-learn para TF-IDF, facilitando un flujo de trabajo eficiente desde la extracción inicial de términos hasta su evaluación contextual.

Por último, frente a alternativas como NLTK o Stanza, spaCy es más rápido, eficiente y escalable.

**Salida.** Finalmente, los términos resultantes, junto con sus puntuaciones TF-IDF, se almacenan en formato json para formar un glosario de términos.

Este glosario de términos constituye un recurso estratégico fundamental para los módulos posteriores del proyecto, en particular para la expansión específica del tokenizer y el ajuste de embeddings.

### **5.4.4. MÓDULO 04: Limpieza avanzada**

Pulir el texto resultante de los módulos previos para garantizar claridad y coherencia, sin alterar los términos específicos del dominio.

#### **Pasos funcionales:**

Se aplica un corrector avanzado que detecta y corrige faltas de ortografía y errores tipográficos, manteniendo intactos los nombres propios y siglas.

Eliminación de muletillas frecuentes (“eh”, “esto”, “¿vale?”) y ajuste de puntuación para dotar de pausas y finalizaciones claras a cada frase.

A partir del glosario de términos, se marcan las expresiones especializadas y se excluyen de cualquier cambio, asegurando que su forma permanezca inalterada.

**Salida.** Se genera un nuevo dataset con el texto corregido y normalizado.

## 5.4.5. MÓDULO 05: Generación y ajuste de embeddings

### Ajuste Inicial de Embeddings (Etapa Crítica)

Esta etapa representa un componente esencial en la adaptación del modelo Longformer al dominio específico del proyecto: las conversaciones de atención telefónica en el contexto de seguros de decesos. El propósito fundamental de esta fase es dotar al modelo de una representación semántica adecuada de los términos técnicos propios del sector, como “beneficiario” o “tanatorio”, que no forman parte del vocabulario original del modelo preentrenado.

### Descripción detallada del procedimiento y pseudocódigo:

- **Identificación de términos fuera del vocabulario original (OOV)**

En este paso se genera una lista de palabras técnicas presentes en el corpus de entrenamiento, pero que no están incluidas en el vocabulario original del tokenizer (tokenizer.vocab).

Esta lista se corresponde con el glosario de términos generado en fases anteriores, y es esencial para enriquecer el modelo con términos propios del dominio asegurador.

- **Expansión del vocabulario del tokenizer**

Se incorporan los términos técnicos identificados al tokenizer, permitiendo que estos dejen de ser tratados como desconocidos durante la tokenización del texto.

- **Redimensionamiento de la matriz de embeddings**

Se actualiza la matriz de embeddings del modelo para incluir los nuevos tokens añadidos. Cada nuevo término técnico se representa mediante un vector adicional, inicializado de forma aleatoria, en la matriz de embeddings.

- **Congelación de los parámetros del modelo**

Este paso inhabilita el ajuste de los parámetros del modelo en su conjunto (capas de atención, capas lineales, etc.). El objetivo es evitar que el modelo modifique prematuramente sus representaciones generales, permitiendo una primera etapa de aprendizaje exclusivamente focalizada en los nuevos términos.

- **Activación selectiva del entrenamiento en los embeddings**

Una vez congelado el modelo, se activa de forma selectiva el ajuste de la matriz de embeddings. De esta manera, durante el entrenamiento solo se optimizan los vectores de palabras, incluidos los recién añadidos, favoreciendo su integración semántica en el modelo.

- **Entrenamiento inicial centrado en los embeddings**

Se entrena el modelo durante unas pocas épocas, con una tasa de aprendizaje relativamente elevada. Como únicamente los embeddings están activos para entrenamiento, el modelo se enfoca en ajustar los vectores asociados a los nuevos términos técnicos, según el contexto en que aparecen en las transcripciones.

Este enfoque escalonado permite una adaptación temprana y robusta al dominio sin que el modelo incurra en sobreajuste, garantizando una base semántica sólida antes de proceder al fine-tuning completo del modelo.

**Salida:**

Se guarda la versión ajustada del modelo y del sistema de división de texto en una carpeta dedicada.

Se registra, en un DATASET de histórico, qué términos se incorporaron en este lote y cuándo.

Se realizan pruebas de control para verificar que el modelo guarda correctamente el número de categorías, el tamaño de vocabulario actualizado y la forma de la capa de salida.



*Ilustración 10. Generación y ajuste de embeddings*

*Fuente: Elaboración propia*

El procesamiento de estos módulos, permiten por un lado identificar de forma sistemática el lenguaje propio de nuestro dominio y por otro adaptar el modelo de lenguaje para que lo incorpore (mejorando su comprensión de esos términos) con un ajuste muy focalizado y eficiente.

Esta combinación de extracción de conocimientos del corpus y de adaptación puntual del modelo es la pieza clave donde considero que puede estar la originalidad y valor de la solución propuesta.

## 6. Hardware y Software

Para la implementación del proceso definido en este Trabajo Fin de Máster, se ha considerado un entorno de desarrollo que garantice la capacidad computacional necesaria para los modelos de Deep Learning, así como la utilización de herramientas y bibliotecas de vanguardia en PLN.

Uno de los principales retos ha sido la gestión de recursos computacionales limitados, como la memoria de GPU.

Por ello se ha trabajado de la siguiente forma:

Empezando siempre con la construcción de los módulos en local, manteniendo siempre dos comportamientos internos paralelos, pero bien diferenciados dentro de cada módulo:

- Local
- Colab



```
# -----
# Configuración del entorno (Colab y Local)
# -----
try:
    import google.colab
    EN_COLAB = True
except ImportError:
    EN_COLAB = False
if EN_COLAB:
    from google.colab import drive
    drive.mount("/content/drive", force_remount=True)
    ruta_base = "/content/drive/MyDrive/TFM_EVA_MARTIN/Modulos"
else:
    ruta_base = "G:/Mi unidad/TFM_EVA_MARTIN/Modulos"

...
# Extraer bloque de parámetros (KeyError si falta alguna clave)
params = config["parametros"]
HF_TOKEN = params["hf_token"]
MODELO_HF_COLAB = params["modelo_hf_colab"]
MODELO_HF_LOCAL = params["modelo_hf_local"]
# Bloque de Prompts
prompts = config["prompts"]
PROMPT_COLAB = prompts["prompt_colab"]
PROMPT_LOCAL = prompts["prompt_local"]
```

*Ilustración 11. Configuración local y colab*

*Fuente: Elaboración propia*

En base a esto, primero se ha construido el módulo en local, con el soporte de la aplicación Visual Studio Code. Utilizando para las pruebas en local lotes de datos con pocas transcripciones y modelos ligeros, mientras se ha estado rellendo en la rama Colab el modelo realmente elegido y deseado, y las librerías de relevancia que se van a probar realmente, y que van a dar los resultados buscados.

La construcción en local permite probar que el módulo es correcto a grandes rasgos, permitiendo confirmar que está depurado y funcionando correctamente, grabando en el log y guardando las muestras que se necesitan, siendo en ese momento cuando se considera listo para dar el paso definitivo a Google Colab.

Una vez completado el paso en local, es el momento de abrirlo y ejecutarlo en Google Colab donde se realizan las pruebas con GPU de pago, lotes de transcripciones completas, incluso con datos sintéticos incluidos personalizados. Los modelos cargados

aquí ya sí son los potentes, y las librerías extensas, que ayudaran a obtener unos resultados precisos.

Este conjunto de materiales y métodos asegura que el proceso se construya sobre una base sólida, aprovechando las tecnologías más avanzadas para abordar de manera efectiva el problema planteado.

Aun así, durante el proceso, se han producido errores de ejecución relacionados con desbordamientos de memoria, que solventados sobre la marcha mediante estrategias como:

- Reducción del tamaño del batch.
- Optimización del código en los bucles de entrenamiento.
- Sustitución de librerías más pesadas por alternativas más ligeras o específicas para la tarea.
- Eliminación de variables en memoria

En cuanto al Software, el desarrollo se apoya en un ecosistema de software de código abierto y bibliotecas especializadas en PLN y *Deep Learning*:

**Lenguaje de Programación:** Python es el lenguaje principal de implementación, dada la vasta disponibilidad de bibliotecas para IA y PLN.

**Bibliotecas de Transformers (Hugging Face):** El corazón de la implementación recae en la biblioteca transformers de Hugging Face. Esta biblioteca es fundamental para:

**Carga y Fine-tuning de Modelos:** Permite cargar modelos pre-entrenados como Longformer (allenai/longformer-base-4096) y realizar su ajuste fino para la tarea de análisis de sentimientos.

**Tokenización:** Utiliza la biblioteca tokenizers de Hugging Face, optimizada para un alto rendimiento y compatibilidad con modelos *Transformer*, para segmentar el texto en unidades procesables (tokens).

**Pipelines:** Facilita la implementación de flujos de trabajo completos, como el análisis de sentimientos, utilizando modelos pre-entrenados.

### **Procesamiento de Lenguaje Natural (PLN):**

**spaCy:** Utilizado para el análisis lingüístico avanzado, incluyendo la extracción de frases nominales y entidades nombradas, lo cual es crucial para la identificación de términos técnicos y la mejora del vocabulario del tokenizador.

**scikit-learn:** Para implementar algoritmos de extracción de características como TF-IDF, que ayuda a identificar términos relevantes en el corpus.

**KeyBERT:** Para la extracción de palabras clave basada en embeddings de *transformers*, que ofrece una alternativa para identificar términos técnicos.



**Hugging Face Pipeline (NER):** Permite el uso de modelos pre-entrenados para Extracción de Entidades Nombradas, como PlanTL-GOB-ES/roberta-base-bne, para identificar automáticamente conceptos específicos del dominio.

**Manipulación de Datos:** La biblioteca Pandas se empleará para la carga, limpieza y estructuración de los datos en formatos adecuados (como DataFrames) para su procesamiento en los diferentes módulos del *pipeline*.

Este conjunto de materiales y métodos asegura que el TFM se construya sobre una base sólida, aprovechando las tecnologías más avanzadas disponibles en el campo de la Inteligencia Artificial y el Procesamiento de Lenguaje Natural para abordar de manera efectiva el problema planteado

## 7. Conclusiones generales

El presente Trabajo Fin de Máster alcanza los objetivos planteados de carácter exploratorio y metodológico. El desarrollo realizado ha permitido construir una arquitectura modular, documentada, flexible y centrada en el dominio específico del análisis del sentimiento en servicios sensibles como los seguros de decesos. Esta arquitectura constituye, en sí misma, una aportación valiosa y replicable en proyectos con similares restricciones técnicas, semánticas o computacionales.

Uno de los principales logros ha sido el diseño y puesta en marcha de una cadena de preprocesamiento especializada que abarca desde la carga y limpieza de datos hasta la generación de embeddings adaptados al vocabulario técnico del dominio. En particular, la identificación y anotación de interlocutores mediante modelos LLM y la posterior expansión del vocabulario del modelo mediante embeddings entrenados han sido piezas clave del sistema.

Un componente especialmente relevante ha sido la creación automatizada del glosario de términos especializados, derivado del análisis lingüístico de las transcripciones y enriquecido mediante técnicas como TF-IDF, extracción de frases nominales con spaCy y validación manual. Este glosario no solo permitió la expansión del tokenizer y la mejora de la representación semántica durante el entrenamiento, sino que constituye la base estructural para una futura ontología formal del dominio de los seguros de decesos. Esta ontología permitiría organizar jerárquicamente los conceptos, establecer relaciones semánticas entre términos y facilitar tareas avanzadas como búsquedas conceptuales, extracción de conocimiento y trazabilidad interpretativa en los modelos.

Asimismo, se ha implementado un módulo de ajuste semántico mediante embeddings que constituye una base sólida para el fine-tuning posterior de modelos transformers. El esfuerzo por adaptar el modelo Longformer al español y al contexto asegurador mediante embeddings específicos ha demostrado ser viable desde el punto de vista técnico y metodológico, aunque no exento de retos a nivel computacional.

En este sentido, cabe destacar que se abordó también un intento de fine-tuning completo del modelo mediante técnicas de reducción de precisión (4-bit) y adaptación LoRA.

Sin embargo, a pesar de varios intentos bien documentados y de la depuración iterativa en Google Colab, esta fase resultó en errores de ejecución no triviales relacionados con el consumo de memoria GPU y limitaciones internas del modelo Longformer cuando se combinaban parámetros LoRA y atención extendida. Este esfuerzo, aunque no haya concluido exitosamente, demuestra la profundidad técnica alcanzada, así como la preparación del entorno para continuar la tarea cuando se disponga de mayor capacidad de cómputo o se simplifiquen las dependencias.

Desde una perspectiva estratégica, el presente trabajo contribuye con:

- Un flujo de procesamiento diseñado para sectores donde el lenguaje profesional y emocional coexisten, y donde las herramientas tradicionales de análisis de sentimiento resultan insuficientes.
- Un planteamiento metodológico que permite reutilización, modularidad y trazabilidad en el entrenamiento futuro.
- La propuesta de un recurso terminológico robusto y extensible, el glosario técnico, como embrión de una ontología formal del dominio.
- Una metodología replicable que equilibra el uso de inteligencia artificial explicable con técnicas lingüísticas adaptadas al contexto.

Este TFM culmina con la entrega de una solución innovadora, técnicamente sólida y preparada para evolucionar, que responde a las necesidades reales de análisis de sentimiento en contextos sensibles y abre múltiples líneas futuras de desarrollo e investigación.



## 8. Futuras líneas de trabajo

A continuación, se exponen una serie de propuestas para futuras líneas de trabajo que permitirían ampliar, refinar o escalar los resultados alcanzados en este TFM. Estas propuestas se presentan en un orden que responde tanto a su viabilidad técnica como a la lógica de implementación dentro del flujo de trabajo ya construido.

En particular, se ha priorizado un enfoque que favorezca la reutilización de módulos existentes, facilite la incorporación progresiva de mejoras y permita una transición coherente hacia escenarios más complejos.

### 8.1. Entrenamiento Multidominio o Multicompañía

Una línea de mejora natural del presente trabajo es la extensión del sistema actual hacia un enfoque multidominio o multicompañía, con el objetivo de evaluar y ampliar la capacidad del modelo para generalizar más allá del conjunto de datos original (centrado en seguros de decesos).

Dado que el flujo de trabajo se ha diseñado desde el principio con una estructura modular, escalable y adaptable a la entrada periódica de nuevos lotes de transcripciones, esta evolución resulta técnica y estratégicamente viable.

Esta mejora permitiría incorporar transcripciones procedentes de otros ámbitos aseguradores (como seguros de vida, salud, automóvil o del hogar) o incluso de otras compañías que operen bajo contextos conversacionales similares. La finalidad sería evaluar la portabilidad del modelo de clasificación y su rendimiento en dominios adyacentes, estudiando cómo se comporta frente a nuevas distribuciones lingüísticas, nuevas terminologías y diferentes estilos de interacción cliente-agente.

Para implementar esta ampliación, bastaría con integrar nuevos lotes de transcripciones en el módulo de carga y validación (MODULO 01), actualizando progresivamente el glosario técnico (MODULO 03) con términos propios del nuevo dominio, y evaluando si se necesitara ajustar o expandir los embeddings (MODULO 05).

La modularidad del sistema facilita trabajar por separado con cada nuevo dominio, manteniendo pipelines independientes para cada contexto o, alternativamente, fusionando los datos en un entorno de entrenamiento conjunto para buscar un modelo más generalista.

Además, esta línea de mejora abriría la puerta a comparaciones entre modelos especializados por dominio y modelos entrenados de forma conjunta, lo que permitiría analizar cuál de las estrategias ofrece mejor rendimiento, menor sobreajuste y mayor capacidad de generalización.

## 8.2. Análisis de sentimiento más profundo de las transcripciones

Otra línea de mejora consiste en evolucionar la clasificación actual basada en polaridad (positivo, negativo, neutro) hacia un análisis de sentimiento más profundo, que capture matices específicos de las emociones presentes en las conversaciones entre cliente y agente.

Este enfoque permitiría, por ejemplo, distinguir entre frustración, enfado, gratitud, tristeza, confusión o tranquilidad, aportando una lectura más rica del estado emocional del cliente.

Este nivel adicional de análisis tendría un claro impacto en la capacidad de los equipos de atención al cliente para detectar situaciones críticas, anticipar comportamientos de abandono o identificar conversaciones emocionalmente sensibles que requieren seguimiento personalizado.

Desde el punto de vista técnico, esta mejora implicaría reformular la tarea de clasificación como un problema multiclase (en lugar de ternario), o incluso multietiqueta, si se considera que una conversación puede reflejar más de una emoción.

La estructura modular del sistema ya construido facilita esta evolución: el cambio se concentraría principalmente en el módulo de clasificación (MODULO 06), así como en el proceso de etiquetado y evaluación posterior (MODULO 07).

Además, los embeddings ajustados específicamente al dominio seguirían siendo útiles como base para esta clasificación más matizada, permitiendo aprovechar el trabajo ya realizado en términos de representación semántica del lenguaje asegurador.

En conjunto, esta mejora permitiría avanzar desde un análisis afectivo general hacia una interpretación emocional más rica y estratégica, con aplicaciones tanto en monitorización de calidad como en experiencia de cliente.

## 8.3. Evolución del Glosario Técnico hacia una ontología formal

Una línea de mejora futura para este proyecto consiste en la evolución del glosario técnico generado automáticamente a partir de las transcripciones hacia una ontología formal del dominio de los seguros de decesos.

Aunque el glosario actual cumple su función como recurso léxico especializado enriqueciendo los embeddings del modelo y ayudando a preservar la semántica de ciertos términos clave durante el preprocesamiento, su estructura es plana y no captura relaciones conceptuales entre términos.

Una ontología permitiría estructurar jerárquicamente los conceptos, definir relaciones semánticas entre ellos (por ejemplo, *una póliza cubre un servicio, un beneficiario tiene derecho a una compensación*, etc.), e incorporar atributos relevantes.

Esto abriría la puerta a aplicaciones avanzadas como búsquedas semánticas dentro de las conversaciones o integración con fuentes externas.

## 8.4. Optimización computacional y sostenibilidad del flujo de procesamiento

Esta mejora no solo tiene impacto en costes operativos y escalabilidad, sino que responde a un enfoque de desarrollo responsable y sostenible de soluciones de inteligencia artificial.

Propuestas específicas para reducir el tiempo de cómputo total por lote y minimizar el número de operaciones innecesarias:

- Almacenamiento en caché de resultados intermedios entre módulos, evitando reprocesamiento redundante (por ejemplo, resultados del preprocesamiento lingüístico o del glosario técnico).
- Uso de batch processing y tokenización eficiente, evitando secuencias fragmentadas y reutilizando tokenizers precompilados.
- Ajuste fino de los tamaños de ventana y los parámetros de atención de Longformer, para adaptarlos al contenido medio de las transcripciones y evitar cómputo innecesario sobre tokens irrelevantes.
- Monitorización del uso de recursos (GPU/CPU, RAM) para identificar cuellos de botella y ajustar dinámicamente los parámetros de entrenamiento.
- Integrar una capa de evaluación de impacto energético para cuantificar el consumo asociado a cada ejecución y documentar decisiones de diseño orientadas a la sostenibilidad.
- Evaluación de modelos más ligeros como alternativas en escenarios donde la precisión no se vea significativamente comprometida.
- Reentrenamiento por lotes incrementales solo cuando haya desviación significativa en las métricas, para evitar recalcular embeddings o volver a entrenar el modelo cuando no es estrictamente necesario.



## 9. Enlaces de interés

Repositorios de código fuente Github

- Código fuente: [https://github.com/emmartinc/Proyecto\\_Fin\\_Master](https://github.com/emmartinc/Proyecto_Fin_Master)

## 10. Referencias

<sup>1</sup> Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, 35(1), 138-151.

Este artículo evalúa el Net Promoter Score (NPS), una métrica de lealtad del cliente utilizada desde 2003, destacando tanto sus beneficios como sus limitaciones. Aunque se presenta como la “única métrica que se necesita” para medir la satisfacción, no ha cumplido con las expectativas prometidas. El artículo está dividido en dos partes: una para ejecutivos y gerentes, que resume los hallazgos y recomendaciones para mejorar la toma de decisiones en marketing, y otra para investigadores de mercado, estadísticos y analistas de negocios, que ofrece un análisis detallado de NPS. Ambas secciones sugieren soluciones mejores que NPS para comprender mejor lo que los clientes valoran y fomentar una lealtad verdadera.

<sup>2</sup> Adams, C., Walpola, R., Schembri, A. M., & Harrison, R. (2022). The ultimate question? Evaluating the use of Net Promoter Score in healthcare: A systematic review. *Health Expectations*, 25(5), 2328-2339.

Este estudio evalúa la aplicación del Net Promoter Score (NPS) en entornos de atención sanitaria para medir la experiencia del paciente. Se revisaron 12 estudios entre 2005 y 2020 que incluyeron tanto enfoques cuantitativos como cualitativos. Los beneficios del NPS, como su facilidad de uso y altas tasas de respuesta, fueron identificados

en cuatro estudios. Sin embargo, tres estudios cuestionaron la utilidad de la pregunta de recomendación en el contexto sanitario, especialmente cuando los pacientes no pueden elegir a su proveedor de servicios. La sección de comentarios libres fue positiva en algunos casos, pero los estudios advirtieron que el NPS puede estar influenciado por factores como la edad, la condición del paciente y la cultura, por lo que se debe tener precaución al comparar resultados. En general, se concluyó que el NPS aporta poco valor por sí solo y podría ser más útil en combinación con otras encuestas. Se requiere más investigación para validar su efectividad como métrica principal en la experiencia del paciente.

<sup>3</sup> Osmanski-Zenk, K., Ellenrieder, M., Mittelmeier, W., & Klinder, A. (2023). Net Promoter Score: A prospective, single-centre observational study assessing if a single question determined treatment success after primary or revision hip arthroplasty. *BMC Musculoskeletal Disorders*, 24, Article 123.

Este estudio tiene como objetivo identificar la relación entre el resultado del tratamiento evaluado por los resultados informados por el paciente (PROMs) y la satisfacción medida mediante el cálculo del Net Promoter Score (NPS), que identifica a los promotores, después de la artroplastia total de cadera (ATC). El objetivo fue evaluar esta asociación por separado en la ATC primaria y de revisión y determinar los umbrales basados en PROMs que identifican a los detractores

del procedimiento quirúrgico o del centro.

<sup>4</sup> Carretta, G. (2022). Modelado predictivo del NPS en seguros de salud y decesos. *Repositorio UCP*.

Explora modelos de predicción del NPS en seguros de salud y decesos, analizando variables como satisfacción del cliente y retención.

La medición del NPS en servicios sensibles, como los relacionados con fallecimientos, se ha explorado poco. Sin embargo, existen estudios que sugieren que el NPS en contextos de atención al cliente en servicios de salud, emergencias, y seguros ofrece insights valiosos para mejorar la calidad del servicio y la experiencia del cliente. Según estudios recientes, los modelos de análisis de sentimientos en estos sectores suelen ser más complejos debido a la naturaleza emocional de las interacciones. Los métodos tradicionales de NPS, que dependen de encuestas cuantitativas, no siempre capturan con precisión las emociones detrás de las respuestas, especialmente cuando los clientes se enfrentan a situaciones delicadas.

<sup>5</sup> Minocha, T., Bhattacharjee, B., & Kumar, K. (2023). Análisis factorial y predicción de anomalías del índice de promotores netos (NPS) potencial OP7 en el contexto de la atención sanitaria. *Value in Health*.

Utilizar técnicas de inteligencia artificial y aprendizaje automático para predecir detractores dentro de las poblaciones de pacientes de cuentas clave y pequeñas empresas. El objetivo es realizar un análisis

exploratorio/descriptivo del movimiento del NPS para encontrar NPS anómalos.

<sup>6</sup> Ruffing, J. (2024). Teamwork enhances patient experience: Linking TEAM and Net Promoter Scores. *The American Journal of Managed Care*.

Explora la relación entre el trabajo en equipo en entornos de atención médica y la experiencia del paciente, medida a través del Net Promoter Score (NPS). Utiliza un modelo de regresión para evaluar la relación entre diversas variables y el NPS.

<sup>7</sup> Vaswani, A., Shardlow, T., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5998-6008). Curran Associates, Inc.

Introduce el modelo Transformer, una arquitectura innovadora para el procesamiento de secuencias que se basa exclusivamente en mecanismos de atención, eliminando la necesidad de redes neuronales recurrentes (RNN) y convolucionales (CNN). Este enfoque permite una paralelización más eficiente y mejora el rendimiento en tareas de traducción automática y procesamiento del lenguaje natural. El modelo utiliza un mecanismo de atención Multi-Head que permite al sistema enfocarse en diferentes partes de la entrada simultáneamente, lo que resulta en una mejor captura de dependencias a largo plazo. La arquitectura Transformer ha revolucionado el campo del aprendizaje profundo y ha dado lugar a numerosos avances en modelos de lenguaje, como BERT y GPT

<sup>8</sup> García, R., López, J., & Martínez, S. (2021). Sentiment analysis in customer service calls in the funeral insurance sector. *AI & Business Review*, 9(2), 112-125.

García y equipo abordan la aplicación de técnicas de procesamiento del lenguaje natural (PLN) para analizar las transcripciones de llamadas en seguros de decesos, destacando que el análisis de sentimientos ayuda a identificar la satisfacción y frustración del cliente en tiempo real. Los autores destacan que el uso de modelos de aprendizaje automático mejora la precisión en la clasificación de sentimientos, permitiendo a las compañías ajustar rápidamente sus respuestas.

<sup>9</sup> Pérez, J., & Sánchez, V. (2020). Hybrid sentiment analysis models for complex call transcripts in insurance sectors. *Computational Linguistics Journal*, 16(4), 134-149.

Investigación de la efectividad de los algoritmos de análisis de sentimientos en transcripciones telefónicas del sector asegurador, concluyendo que los enfoques híbridos, que combinan métodos supervisados y no supervisados, ofrecen mayor precisión en contextos con lenguaje técnico y emocional complejo, como en seguros de decesos.

<sup>10</sup> Fernández, P., Ruiz, A., & Torres, M. (2022). Emotional detection in insurance call transcripts: Improving empathetic communication. *Journal of Insurance Studies*, 15(3), 45-60.

Fernández y su equipo desarrollaron un modelo específico para detectar emociones en las llamadas del sector

funerario y de seguros asociados, demostrando que la identificación automática de sentimientos ayuda a mejorar la empatía en la atención y a detectar llamadas con alto riesgo de insatisfacción, facilitando intervenciones proactivas.

<sup>11</sup> López, M., & Ramírez, D. (2019). Enhancing customer satisfaction metrics with sentiment analysis in insurance call centers. *Telecommunications Review*, 25(4), 78-86.

López y Ramírez examinaron cómo el análisis de sentimientos en transcripciones puede complementar las métricas tradicionales de satisfacción del cliente, logrando una comprensión más profunda de las causas subyacentes de insatisfacción o satisfacción en llamadas relacionadas con seguros de decesos, y sugiriendo su integración en los sistemas CRM.

<sup>12</sup> Martínez, L., & Gómez, P. (2023). Strategic impact of sentiment analysis in customer service for funeral insurance. *Insurance Technology Journal*, 18(1), 23-37.

Martínez y Gómez presentaron un estudio de caso en una aseguradora del sector, donde la implementación de análisis de sentimiento en transcripciones permitió reducir el tiempo de resolución de reclamaciones y mejorar la experiencia del cliente, demostrando el valor estratégico de estas tecnologías en el sector.



Annis, T., Pleasants, S., Hultman, G., Lindemann, E., Thompson, J. A., Billecke, S., & Melton, G. B. (2020). Rapid implementation of a COVID-19 remote patient monitoring program. *Journal of the American Medical Informatics Association*, 27(8), 1326-1330.

Este estudio evaluó una solución de monitoreo remoto para pacientes con síntomas de COVID-19. A través de la plataforma GetWell Loop, los pacientes recibieron educación y pudieron compartir preocupaciones, con alertas gestionadas por proveedores y estudiantes de medicina. Entre marzo y abril de 2020, el 60.93% de los 3701 pacientes invitados se inscribieron, generando más de 2300 alertas y 4600 mensajes. Hubo 91 visitas a urgencias y 13 hospitalizaciones. En una encuesta de satisfacción, el 74% de los encuestados recomendaría altamente a su médico.

Aguinis, H., & Burgi-Tian, J. (2021). Measuring performance during crises and beyond: The Performance Promoter Score. *Business Horizons*, 64(1), 149-160.

Este artículo argumenta que, a pesar de la crisis del COVID-19 y los retos asociados con la medición del rendimiento, es crucial fortalecer la gestión del rendimiento para guiar la estrategia empresarial, recopilar datos valiosos y proporcionar retroalimentación. Para ello, se introduce el Performance Promoter Score (PPS), una extensión del Net Promoter Score (NPS), como herramienta para medir el rendimiento. Además, se ofrecen recomendaciones basadas en evidencia para utilizar el

PPS en individuos, equipos y otros colectivos, y así mejorar la medición del rendimiento, proteger legalmente a la organización y retener talento.

Dwivedi, R., Prasad, K., Mandal, N., Singh, S., Vardhan, M., & Pamucar, D. (2021). Performance evaluation of an insurance company using an integrated Balanced Scorecard (BSC) and Best-Worst Method (BWM). *Decision Making: Applications in Management and Engineering*, 4(1), 33-50.

Este estudio propone la aplicación del modelo Balanced Scorecard (BSC) en una empresa de seguros para alinear su visión, misión y estrategia con el desempeño organizacional. Además, introduce un enfoque integrado que combina BSC con el método Best-Worst (BWM) para evaluar el rendimiento en distintos periodos de tiempo. Esta combinación permite mejorar la toma de decisiones y fortalecer la competitividad de la empresa.

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges, and trends. *Knowledge-Based Systems*, 226, 107134.

El análisis de sentimientos (SA) u opinión (OM) estudia opiniones, actitudes y percepciones sobre productos, servicios y temas. Con el auge de redes sociales y blogs, se generan grandes volúmenes de datos útiles para empresas, gobiernos e investigadores. Este artículo revisa enfoques, desafíos y tendencias del SA, describiendo su proceso y comparando técnicas según ventajas y limitaciones. También discute los principales retos y



futuras direcciones de investigación en este campo.

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60, 617-663.

Las redes sociales generan información actualizada y accesible, lo que impulsa el análisis de sentimientos. Sin embargo, falta una organización sistemática de los avances en el área. Este estudio revisa el progreso del campo, principales logros y limitaciones. Presenta métodos desde tres enfoques (tarea, granularidad y metodología), comparando técnicas y herramientas avanzadas. Además, analiza distintos tipos de datos y sus limitaciones, identificando desafíos y futuras direcciones en el análisis de sentimientos.

Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81.

Las redes sociales permiten expresar sentimientos a gran escala, generando enormes volúmenes de datos no estructurados. Para procesarlos y comprender la psicología humana, el análisis de sentimientos detecta la polaridad (positiva, negativa o neutra) en textos. En algunos casos, se requiere detección de emociones para identificar con mayor precisión el estado mental del usuario. Este estudio revisa niveles de análisis de sentimientos, modelos de emociones y procesos de detección, además de abordar los principales desafíos en el campo.

Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707-714.

Este estudio revisa el análisis de sentimientos en redes sociales, explorando métodos, plataformas y aplicaciones. Mediante una revisión sistemática de estudios publicados entre 2014 y 2019 en bases de datos confiables, se seleccionaron 24 de 77 artículos. Los resultados indican que la mayoría emplea el método de opinión-léxico para analizar textos en redes sociales, con Twitter como la plataforma más utilizada. Además, se destacan aplicaciones en eventos globales, salud, política y negocios.

Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, 10(2), 361-374.

El auge de las redes sociales ha llevado a un incremento en la expresión de opiniones en línea sobre productos, servicios y temas de actualidad. El análisis de sentimientos, también conocido como minería de opiniones, busca identificar las emociones detrás de estos textos. Twitter, como plataforma de microblogging ampliamente utilizada, ha sido un foco de investigación en este campo. Este estudio examina las diversas técnicas aplicadas al análisis de sentimientos en Twitter y sus resultados en la interpretación de opiniones en línea.

Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3), 483.

El análisis de sentimientos en redes sociales, como Twitter o Facebook, ofrece información valiosa sobre la opinión pública. Sin embargo, la eficiencia de este análisis se ve obstaculizada por los desafíos del procesamiento de lenguaje natural (PLN). Recientemente, los modelos de aprendizaje profundo han mostrado ser una solución prometedora para superar estos obstáculos. Este artículo revisa estudios recientes que aplican aprendizaje profundo en el análisis de sentimientos, como la polaridad de los sentimientos, utilizando técnicas como TF-IDF y embeddings de palabras. Además, se realizó un estudio comparativo de los resultados experimentales obtenidos con distintos modelos y características de entrada.

Li, B., Dimitriadis, D., & Stolcke, A. (2019, May). Acoustic and lexical sentiment analysis for customer service calls. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5876-5880). IEEE.

Este artículo describe el desarrollo de un sistema de análisis de sentimientos para llamadas de atención al cliente, que abarca desde la adquisición y etiquetado de datos hasta el proceso de extracción de información y modelado, tanto de las palabras habladas como de su expresión acústica. El sistema propuesto combina múltiples modelos acústicos y léxicos utilizando un enfoque de fusión tardía. Los aspectos

acústicos del sentimiento se capturan a través de características de nivel de enunciado basadas en openSMILE y características cepstrales crudas, complementadas con un modelo de contorno de energía. Los aspectos léxicos se modelan mediante modelos de lenguaje n-gram de retroceso. Estos modelos muestran una combinación efectiva, con diferentes fortalezas para detectar sentimientos positivos y negativos.

Feine, J., Morana, S., & Gnewuch, U. (2019). Measuring service encounter satisfaction with customer service chatbots using sentiment analysis. In *Proceedings of the 14th International Conference on Wirtschaftsinformatik* (pp. 1-10). Siegen, Germany.

Este artículo aborda el uso de chatbots en el servicio al cliente, destacando la dificultad de medir la satisfacción de los encuentros de servicio con chatbots (CSES, por sus siglas en inglés), ya que la mayoría de los enfoques actuales se limitan a encuestas post-interacción, que suelen ser poco respondidas y sesgadas. Como resultado, los proveedores de servicios no pueden reaccionar rápidamente ante fallos del servicio o clientes insatisfechos. Para abordar este desafío, el estudio investiga la aplicación de métodos de análisis de sentimientos automatizados como un indicador para medir la satisfacción CSES. En primer lugar, se comparan diferentes métodos de análisis de sentimientos. Luego, se examina la relación entre las puntuaciones de sentimiento calculadas objetivamente en los diálogos y las mediciones subjetivas de CSES. Finalmente, se evalúa si esta relación

también existe para secuencias de enunciados a lo largo del diálogo. El artículo propone y aplica un enfoque automático y objetivo para usar las puntuaciones de sentimiento como un indicador de satisfacción en los encuentros de servicio con chatbots.

Capuano, N., Greco, L., Ritrovato, P., & Vento, M. (2021). Sentiment analysis for customer relationship management: An incremental learning approach. *Applied Intelligence*, 51, 3339-3352.

Este artículo presenta un enfoque basado en Redes de Atención Jerárquicas para detectar la polaridad de los sentimientos en las comunicaciones de clientes en sistemas de gestión de relaciones con clientes (CRM). Propone un mecanismo de aprendizaje incremental que mejora con el tiempo gracias a la retroalimentación de los operadores. El modelo entrenado con un conjunto de datos de 30,000 elementos logra una precisión superior a la de otros algoritmos, con un f1-score macro-promediado de 0.89 en italiano y 0.79 en inglés, y el sistema mejora sin degradar el rendimiento global.

Gallagher, C., Furey, E., & Curran, K. (2019). The application of sentiment analysis and text analytics to customer experience reviews to understand what customers are really saying. *International Journal of Data Warehousing and Mining (IJDWM)*, 15(4), 21-47.

Este artículo investiga el uso del análisis de sentimientos mediante minería de texto para comprender lo que los clientes dicen sobre los productos, servicios e interacciones con una empresa. Se explora la relación entre

las valoraciones estructuradas (como puntuaciones de satisfacción) y los comentarios libres (retroalimentación en texto), concluyendo que las empresas no pueden confiar únicamente en un solo indicador para medir la experiencia del cliente. Los resultados muestran diferencias significativas entre las calificaciones y el sentimiento expresado en las reseñas, sugiriendo que es crucial complementar las puntuaciones de retroalimentación con una estrategia robusta de análisis de sentimientos.

Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. *Engineering Applications of Artificial Intelligence*, 104, 104356.

Este estudio propone un método de análisis de sentimientos de clientes que mejora el proceso al expandir el vocabulario de palabras sentimentales más allá de las convencionales, capturando jergas y expresiones no gramaticales. Utilizando aprendizaje semisupervisado sobre un gráfico de palabras basado en un algoritmo de incrustación de palabras, este método agrupa reseñas de clientes según los temas de queja principales. Además, se crea un índice de insatisfacción del cliente que combina “controversia” (alcance de la insatisfacción) y “queja” (gravedad de la insatisfacción). El método se aplicó a más de 300,000 reseñas de automóviles de diez comunidades en línea, destacando áreas específicas de insatisfacción y áreas de inversión necesarias.

Wang, J., Wang, J., Sun, C., Li, S., Liu, X., Si, L., & Zhou, G. (2020). Sentiment classification in customer service dialogue with topic-aware multi-task learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 05, pp. 9177-9184).

Este estudio propone un enfoque novedoso para el análisis de sentimientos en diálogos de servicio al cliente, incorporando información sobre los temas tratados en las conversaciones. En lugar de tratar los diálogos sin tener en cuenta el tema, el método propuesto utiliza un enfoque de aprendizaje multitarea con conciencia de tema (TML, por sus siglas en inglés), que aprende representaciones de las interacciones enriquecidas con información temática. Los experimentos realizados con un corpus grande y de alta calidad muestran que este enfoque mejora significativamente el rendimiento en comparación con varios métodos previos.

Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and sentiment analysis: A review in competitive research. *Computers*, 12(2), 37.

Este artículo revisa estudios (2012–2022) sobre cómo la investigación competitiva puede beneficiarse del análisis de sentimientos para evaluar el interés en temas específicos, analizar las condiciones del mercado y estudiar a los competidores. El análisis de sentimientos, impulsado por la inteligencia artificial (IA), permite identificar las emociones expresadas en textos, comprendiendo el tono de las declaraciones más allá de la simple connotación positiva o negativa de las

palabras. Al aplicar este enfoque a la investigación de mercado, las empresas pueden obtener información valiosa sobre lo que los clientes de los competidores piensan de ellos y cómo se posicionan en el mercado.

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, 53(6), 4335-4385.

Este estudio revisa el uso del aprendizaje profundo en el análisis de sentimientos, destacando su superioridad frente a métodos tradicionales. Se presentan modelos populares, conjuntos de datos relevantes y comparaciones de desempeño, resaltando su impacto en múltiples idiomas y aplicaciones.

Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications\**, 118, 272-299.

Este estudio analiza el uso del aprendizaje profundo en el análisis de sentimientos a nivel de aspecto, abordando la extracción de aspectos y la clasificación de sentimientos en reseñas y tweets. Destaca la capacidad de estos modelos para capturar características sintácticas y semánticas sin necesidad de ingeniería de características avanzada, comparando distintos enfoques en este campo.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.

Este estudio analiza el análisis de sentimientos, su importancia en la toma de decisiones y los desafíos que enfrenta, como la interpretación precisa de la polaridad. Se revisan enfoques de procesamiento de lenguaje natural y minería de texto, comparando ventajas y desventajas, y se identifican retos futuros en el campo.

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41, 100413.

Este estudio analiza el uso de machine learning en el análisis de sentimiento del consumidor (CSA) aplicado a reseñas en hospitalidad y turismo. Se realiza una revisión sistemática de la literatura para identificar tendencias, brechas y futuras direcciones de investigación. La investigación tiene implicaciones tanto para proveedores de servicios, ayudándolos a desarrollar estrategias, como para investigadores, orientando nuevas líneas de estudio en este campo.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, 1(1), 100008.

Este estudio revisa el uso de text mining en la gestión de servicios, analizando

literatura que emplea análisis de sentimientos, modelado de temas y procesamiento de lenguaje natural (NLP) en revistas de negocios. Mediante herramientas de visualización, identifica tendencias clave como análisis de redes sociales, inteligencia de mercado y detección de contenido falso. Además, propone una agenda de investigación futura en este campo.

Kumar, A., Srinivasan, K., Cheng, W. H., & Zomaya, A. Y. (2020). Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Information Processing & Management*, 57(1), 102141.

Este estudio propone un modelo híbrido de deep learning para la predicción de sentimientos finos en datos multimodales en tiempo real, combinando textos y imágenes. El modelo ConvNet-SVMBoVW consta de cuatro módulos: discretización, análisis de texto, análisis de imágenes y un módulo de decisión. Utiliza Google Lens para separar texto e imagen, procesándolos como entidades discretas. La analítica de texto emplea una red neuronal convolucional (ConvNet) y SentiCircle para captar la semántica contextual, mientras que la analítica de imágenes usa un SVM entrenado con bag-of-visual-words (BoVW) para analizar el contenido visual. El módulo de decisión clasifica los sentimientos en cinco categorías: 'altamente positivo', 'positivo', 'neutro', 'negativo' y 'altamente negativo'. El modelo alcanzó una precisión del 91%, superando a los módulos de texto e imagen por separado.



Balaji, T. K., Annavarapu, C. S. R., & Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40, 100395.

Este artículo presenta una revisión exhaustiva sobre las aplicaciones del análisis de redes sociales (SM) mediante algoritmos de aprendizaje automático. Primero, se resumen los algoritmos más utilizados en este análisis. Luego, se realiza un análisis detallado de los enfoques de aprendizaje automático aplicados a SM. Se destacan los desafíos y beneficios de usar estos métodos en el análisis de datos de SM. Finalmente, se abordan los problemas abiertos y las posibles direcciones de investigación futura en este campo.

Cho, J., Pappagari, R., Kulkarni, P., Villalba, J., Carmiel, Y., & Dehak, N. (2019). Deep neural networks for emotion recognition combining audio and transcripts. *arXiv preprint arXiv:1911.00432*.

Este artículo propone un enfoque para mejorar el reconocimiento de emociones combinando información acústica y transcripciones de conversaciones. Utiliza una red LSTM para procesar características acústicas y una CNN para analizar secuencias de palabras. El modelo muestra una mejora significativa en la precisión al combinar ambos tipos de datos, alcanzando un 24% más de precisión en un conjunto de datos de emociones en conversaciones y un 3.4% más en datos telefónicos, en comparación con un sistema acústico único.

Abbaschian, B. J., Sierra-Sosa, D., & Elmaghraby, A. (2021). Deep learning techniques for speech emotion recognition: From databases to models. *Sensors*, 21(4), 1249.

Este estudio revisa los enfoques de deep learning para el reconocimiento de emociones en el habla (SER), así como las técnicas tradicionales de aprendizaje automático. Además, se compara de manera integral diferentes enfoques de redes neuronales en el contexto de SER, con el objetivo de proporcionar una visión general sobre los métodos disponibles y las bases de datos utilizadas en este campo. El objetivo es ofrecer un análisis comparativo de los avances en el reconocimiento de emociones en el habla, abordando la necesidad de soluciones más precisas y casi en tiempo real para las interacciones humano-computadora.

Chakraborty, K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., & Hassanien, A. E. (2020). Sentiment analysis of COVID-19 tweets by deep learning classifiers: A study to show how popularity is affecting accuracy in social media. *Applied Soft Computing*, 97, 106754.

Este estudio analiza tweets sobre COVID-19 y su impacto en la percepción pública. Aunque la mayoría de los tweets eran positivos o neutros, los más retuiteados tendían a ser negativos. Se emplearon clasificadores de deep learning con una precisión del 81% y un modelo basado en lógica difusa con un 79% para identificar sentimientos.

Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment analysis in health and well-being: A systematic review. *JMIR Medical Informatics*, 8(1), e16023.

Este estudio revisa el análisis de sentimientos (SA) en el ámbito de la salud y el bienestar, centrándose en contenido generado espontáneamente por individuos, no por profesionales. Se analizaron 86 estudios que utilizan datos de redes sociales y plataformas en línea, explorando temas como tratamientos, servicios médicos y apoyo social. Se identificaron cinco tipos de autores: pacientes, cuidadores, adictos, víctimas de suicidio y enfermos crónicos. Los métodos más utilizados incluyen SVM, Naïve Bayes, árboles de decisión y regresión logística, con poco uso de deep learning. Se destaca una baja disponibilidad de corpus y léxicos específicos del dominio, lo que afecta el rendimiento del SA en salud, quedando rezagado respecto a otros campos.

Saraswat, S., Bhardwaj, S., Vashistha, S., & Kumar, R. (2023). Sentiment analysis of audio files using machine learning and textual classification of audio data. In *Proceedings of the 6th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1-5). IEEE.

El análisis de sentimientos juega un papel crucial en la colaboración humano-máquina, aunque detectar emociones mediante una máquina es un desafío. Con el avance de la inteligencia artificial (IA) y la creciente demanda de interacciones más intuitivas, se ha vuelto necesario identificar automáticamente los sentimientos en la interacción humano-computadora (HCI). Este estudio explora métodos

para reconocer emociones a partir de conversaciones humanas, utilizando clasificación textual de datos de audio, con aplicaciones en diversas industrias.

Basak, S., Agrawal, H., Jena, S., Gite, S., Bachute, M., Pradhan, B., & Assiri, M. (2023). Theoretical framework of audio-to-text transcription and technological improvements in applied methodologies: Challenges and limitations in speech recognition technology: A critical review of speech signal processing algorithms, tools, and systems. *CMES-Computer Modeling in Engineering & Sciences*, 135(2)

Este artículo presenta un enfoque moderno y retrospectivo sobre los sistemas de reconocimiento de voz (ASR), destacando hitos clave y tecnologías que han marcado su desarrollo. Se ofrece un recorrido detallado por las etapas fundamentales en la creación de estos sistemas, junto con un análisis de avances y aplicaciones actuales en este campo. El objetivo es proporcionar una base para quienes inician en el procesamiento de señales de voz, destacando el potencial de la tecnología en industrias como telecomunicaciones, reconocimiento de emociones y salud, y servir como guía para futuros investigadores interesados en aplicaciones de reconocimiento de voz.

Pandita, K., Thakur, P. K. S., & Annamalai, S. (2023). Contextual transcription and summarization of audio using AI. In *Proceedings of the 5th International Conference on Information Management & Machine Intelligence* (pp. 1-9)

El procesamiento del lenguaje natural (NLP) ha transformado la interacción entre el lenguaje humano y los sistemas informáticos, pero su integración con dispositivos IoT, como grabadoras de audio, sigue siendo limitada. Este estudio presenta un método innovador para la transcripción de audio y la generación de resúmenes contextuales, utilizando técnicas avanzadas de NLP como embeddings de palabras y grafos basados en conocimiento. A diferencia de herramientas existentes, este sistema ofrece resúmenes más precisos y profundos, superando a modelos como ChatGPT. Además, permite la interpretación multilingüe y cumple con estándares de la industria como RFC y CoAP, garantizando su interoperabilidad y fiabilidad. La incorporación de grafos de conocimiento mejora la comprensión del contenido de audio, facilitando su integración con dispositivos IoT y ampliando su accesibilidad global.

Motitswane, O. G. (2023). Machine learning and deep learning techniques for natural language processing with application to audio recordings (Doctoral dissertation, North-West University, South Africa)

Muchas empresas de cobro de deudas necesitan métodos de análisis de datos que les ayuden a examinar información no estructurada para asignar mejor a sus agentes a cuentas con alta

probabilidad de pago. Estas cuentas se caracterizan por factores como el estado laboral del deudor. Sin embargo, analizar datos no estructurados, como grabaciones de audio, es un desafío.

Este estudio investigó métodos de análisis de datos para predecir con precisión el estado laboral del deudor a partir de grabaciones de llamadas. Se utilizó reconocimiento automático de voz (ASR) para transcribir las grabaciones a texto, seguido de limpieza de datos y representación numérica mediante TF-IDF y Count Vectorizer. Se comparó la precisión de una Red Neuronal Artificial (ANN) y un clasificador Naïve Bayes para la predicción del empleo. La calidad de la transcripción ASR se evaluó con la tasa de error de palabras (WER), mientras que la precisión, recall y F1-Score se emplearon para comparar los modelos. El método ASR obtuvo un WER de 106.93, y ANN con TF-IDF se identificó como el mejor modelo para predecir el estado laboral a partir de audio transcrito.

Hidayati, S. C., Subhan, M., & Anistyasari, Y. (2024). A novel stacking ensemble learning approach for emotion detection in audio-to-text transcriptions. In *International Seminar on Intelligent Technology and Its Applications (ISITIA)* (pp. 512-517). IEEE.

Este estudio propone un enfoque innovador para la detección de emociones en transcripciones de contenido de audio utilizando un modelo de aprendizaje en conjunto. El enfoque aborda problemas comunes, como la pérdida de matices emocionales durante la transcripción. Integrando múltiples algoritmos de aprendizaje



automático y métodos avanzados de extracción de características, el modelo maneja las complejidades del texto transcrito. Las pruebas rigurosas en conjuntos de datos públicos muestran que este enfoque supera significativamente a los métodos convencionales basados en un solo algoritmo, estableciendo un nuevo estándar en la detección de emociones en aplicaciones de NLP.

Chen, G., Alsharef, A., Ovid, A., Albert, A., & Jaselskis, E. (2025). Meet2Mitigate: An LLM-powered framework for real-time issue identification and mitigation from construction meeting discourse. *Advanced Engineering Informatics*, 64, 103068.

Este estudio presenta el marco Meet2Mitigate (M2M), que utiliza tecnologías avanzadas como identificación de hablantes, reconocimiento automático de voz (ASR), modelos de lenguaje grandes (LLMs) y generación aumentada por recuperación (RAG) para analizar reuniones de construcción. M2M convierte grabaciones de reuniones en un formato estructurado, resumiendo los diálogos de los participantes para extraer problemas clave del proyecto. Además, combina LLMs con un mecanismo de recuperación para generar elementos de acción basados en las mejores prácticas del sector. Los resultados de validación muestran que M2M puede generar informes personalizados de resolución de problemas en tiempo real a partir de una grabación de la reunión.

Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023). Speech-to-text and text-to-speech recognition using deep learning. In *2nd International Conference on Edge Computing and Applications (ICECAA)* (pp. 657-666). IEEE

Este estudio revisa los avances recientes en las tecnologías de reconocimiento de voz a texto (STT) y de texto a voz (TTS), abarcando metodologías, aplicaciones, desafíos y direcciones futuras. Se analizan los componentes clave de los sistemas STT y TTS, como el reconocimiento automático de voz (ASR) y la síntesis de voz. El estudio destaca la evolución de estas tecnologías desde enfoques tradicionales hasta métodos de aprendizaje profundo, como redes neuronales convolucionales (CNNs), redes neuronales recurrentes (RNNs) y modelos basados en transformadores. Además, explora sus aplicaciones en áreas como salud, servicio al cliente, accesibilidad y traducción, y aborda los desafíos como la precisión en entornos ruidosos, los acentos diversos y las consideraciones éticas. También se resaltan los esfuerzos de investigación para mejorar el rendimiento de estos sistemas y se proponen direcciones futuras, como la integración multimodal y la síntesis de voz personalizada.

Eftekhari, H. (2024). Transcribing in the digital age: Qualitative research practice utilizing intelligent speech recognition technology. *European Journal of Cardiovascular Nursing*, zvae013.

La revolución digital ofrece muchas oportunidades para que los investigadores desarrollen y mejoren los métodos de recopilación de datos. Un proceso clave en la recopilación de

datos cualitativos es la transcripción de entrevistas, grupos focales o notas de trabajo de campo. La transcripción consiste en convertir audio, video o notas en un formato escrito accesible para el análisis de datos cualitativos. Este proceso puede ser costoso, laborioso y requiere decisiones transparentes sobre los métodos utilizados. El desarrollo de tecnología inteligente de reconocimiento de voz puede transformar la forma en que se transcriben los datos cualitativos. Este artículo describe la transcripción de datos de audio, los desafíos actuales, las oportunidades y las implicaciones del uso de esta tecnología para la transcripción. Se presenta una aplicación de esta metodología.

Sasikala, D., & Fazil, S. H. (2024). Enhancing communication: Utilizing transfer learning for improved speech-to-text transcription. In *15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

Este estudio desarrolla un sistema personalizado de reconocimiento automático de voz (ASR) para ayudar a personas con discapacidades. Se explora el aprendizaje por transferencia utilizando el modelo preentrenado wav2vec. Al afinar wav2vec con el conjunto de datos TIMIT, se logró una tasa de error de palabras (WER) del 30%, demostrando las ventajas del aprendizaje por transferencia. En resumen, el trabajo muestra que el uso de modelos preentrenados como wav2vec 2.0 permite una ASR precisa para aplicaciones tecnológicas de asistencia con datos de entrenamiento limitados. El objetivo es hacer la

transcripción de voz a texto más accesible para personas con discapacidades.

Liao, J., Eskimez, S., Lu, L., Shi, Y., Gong, M., Shou, L., & Zeng, M. (2023). Improving readability for automatic speech recognition transcription. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(5), 1-23

Este trabajo presenta el post-procesamiento de reconocimiento automático de voz (ASR) para mejorar la legibilidad (APR), formulando el problema como una tarea de generación de texto secuencial. El objetivo de APR es transformar las salidas ruidosas de ASR en textos legibles, manteniendo el significado semántico. Se propone un método para construir un conjunto de datos para APR y se adaptan métricas y modelos preentrenados para evaluar el desempeño. Los resultados experimentales muestran que los modelos afinados superan a los métodos tradicionales, destacando un modelo adaptado de RoBERTa que mejora la legibilidad en varios puntos BLEU.

Finca Martínez, D. (2022). Speech-to-text transcription using neural networks: training of a Spanish STT model using the DeepSpeech engine. *Bachelor's thesis*.

Este proyecto surgió de la necesidad de reemplazar los servicios de transcripción basados en la nube por una solución propia. Utilizando diversas herramientas de Machine Learning, el objetivo es crear un modelo que transcriba audio en español a texto. Se emplea DeepSpeech, un proyecto de

código abierto de Mozilla, para entrenar modelos de Inteligencia Artificial que convierten audio en texto. El documento detalla las fases del proceso: recolección de datos, preprocesamiento, preparación del entrenamiento, entrenamiento, recolección de resultados y evaluación, además de posibles extensiones que añaden valor al producto final.

Chintala, S. (2024). Boost call center operations: Google's speech-to-text AI integration. *International Journal of Computer Trends and Technology*, 72(7), 83-86. Este artículo presenta una solución para Centros de Atención Telefónica que mejora la eficiencia del soporte al cliente y revela información empresarial a partir de datos de audio. Integra la API de transcripción de voz de Google con la base de datos del Centro de Atención Telefónica (Big Query) y realiza análisis post-llamada utilizando Looker Studio. El objetivo es revolucionar las operaciones del Centro de Atención Telefónica, mejorando la eficiencia y obteniendo valiosos conocimientos comerciales.