

Grupo 10

myMind: Plataforma móvil para la
detección y seguimiento de emociones

Juan David Castillo Laverde
Juan José Gómez Arenas
Luis Fernando Lee Rodríguez
Natalia Mejía Estrada

PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTÁ, D.C.

2025

CISxxxxCPxx

myMind: Plataforma móvil para la detección y seguimiento de emociones

Autores:

Juan David Castillo Laverde
Juan José Gómez Arenas
Luis Fernando Lee Rodríguez
Natalia Mejía Estrada

MEMORIA DE PROYECTO DE GRADO REALIZADO PARA CUMPLIR UNO
DE LOS REQUISITOS PARA EL TÍTULO EN INGENIERÍA DE SISTEMAS

Director

Ing. Luis Gabriel Moreno Sandoval

Jurados del Trabajo Final de Grado

Ing. «Nombre Jurado 1»
Ing. «Nombre Jurado 2»

PONTIFICIA UNIVERSIDAD JAVERIANA

FACULTAD DE INGENIERIA

PROGRAMA DE INGENIERÍA DE SISTEMAS

BOGOTÁ, D.C.

Mayo, 2025

**PONTIFICIA UNIVERSIDAD JAVERIANA
FACULTAD DE INGENIERIA
PROGRAMA DE INGENIERÍA DE SISTEMAS**

Rector de la Pontificia Universidad Javeriana

Luis Fernando Múnera Congote, S.J.

Decano de la Facultad de Ingeniería

Ing. Diego Alejandro Patiño Guevara

Director de Carrera de Ingeniería de Sistemas

Ing. Carlos Andrés Parra Acevedo

Director del Departamento de Ingeniería de Sistemas

Ing. César Julio Bustacara Medina

Artículo 23 de la Resolución No. 1 de Junio de 1946

“La Universidad no se hace responsable de los conceptos emitidos por sus alumnos en sus proyectos de grado. Sólo velará porque no se publique nada contrario al dogma y la moral católica y porque no contengan ataques o polémicas puramente personales. Antes bien, que se vean en ellos el anhelo de buscar la verdad y la Justicia”

AGRADECIMIENTOS

TABLA DE CONTENIDO

I. INTRODUCCIÓN	1
II. DESCRIPCIÓN GENERAL.	2
1 Oportunidad, Problema	2
1.1 Contexto del Problema	2
1.2 Formulación del Problema	2
1.3 Propuesta de Solución	3
1.4 Justificación de la Solución	3
2 Descripción del Proyecto	4
2.1 Objetivo General	4
2.2 Objetivos Específicos	4
2.3 Entregables, Estándares y Justificación	4
III. CONTEXTO DEL PROYECTO	6
1 Transfondo	6
1.1 Procesamiento de Lenguaje Natural (NLP)	6
1.2 Conversión de Voz a Texto (STT)	6
1.3 Salud Mental	6
1.4 Modelos de Aprendizaje	7
2 Análisis del Contexto	7
2.1 Aplicaciones móviles centradas en la salud mental	7
2.2 Estudios e investigaciones sobre PLN aplicado a la salud mental	8
2.3 Valor añadido de la solución propuesta	8
IV. ANÁLISIS DEL PROBLEMA	9
1 Requerimientos	9
1.1 Módulo de Autenticación y Gestión de Usuarios	9
1.2 Módulo de Grabación y Envío de Notas de Voz	9
1.3 Módulo de Transcripción de Audio	9
1.4 Módulo de Procesamiento de Lenguaje Natural (NLP)	10
1.5 Módulo de Almacenamiento y Data Warehouse	10
2 Restricciones	10
2.1 Restricciones Generales	10
2.2 Restricciones de Software	10
2.3 Restricciones de Hardware	11
2.4 Limitaciones	11
3 Especificación Funcional	11
3.1 E1: Sistema de gestión de usuarios	11
3.2 E2: Captura y procesamiento de audio	12
3.3 E3: Análisis emocional mediante NLP	12
3.4 E4: Gestión de datos y almacenamiento	12
3.5 E5: Visualización y reportes	13
V. DISEÑO DE LA SOLUCIÓN	14
1 Antecedentes teóricos	14
1.1 Modelos de Reconocimiento de Voz	14
1.2 Modelos de Reconocimiento de Sentimientos	16

1.3	Modelos de Reconocimiento de Emociones	18
1.4	Modelos Discretos de Emociones (DEMs)	18
1.5	Modelos Dimensionales de Emociones (DiEMs)	19
2	Pruebas de modelos	20
2.1	Conjunto de Datos	20
2.2	Construcción del conjunto de datos	20
2.3	Comparación de Modelos	21
3	Arquitectura	28
3.1	Selección del estilo arquitectónico	29
4	Herramientas y tecnologías	31
VI.	DESARROLLO DE LA SOLUCIÓN	32
1	Desarrollo del Back-end	32
1.1	Modulo de autenticación de usuario	32
1.2	Módulo de enrutamiento API Gateway	32
1.3	Servicio de usuarios	33
1.4	Servicio de modelos	34
1.5	Servicio de visualizaciones	35
2	Aplicación Móvil	37
2.1	Logo de la aplicación	37
2.2	Psicología del color	37
2.3	Mockups iniciales	37
2.4	Elección de Android Studio	37
2.5	Proceso de desarrollo	38
2.6	Resultado final	38
3	Despliegue	41
3.1	Implementación del ETL (Extract, Transform, Load)	41
3.2	Implementación del Sistema de Failover para MySQL	42
VII.	RESULTADOS	45
VIII.	CONCLUSIONES	48
4	Análisis de Impacto del Proyecto	48
5	Conclusiones y Trabajo Futuro	48
IX.	REFERENCIAS	51
X.	ANEXOS	55

LISTA DE FIGURAS

1	Arquitectura a alto nivel del sistema <i>myMind</i>	30
2	Diagrama de flujo de la aplicación	40
3	Capturas principales de la aplicación móvil <i>myMind</i>	41

LISTA DE TABLAS

1	Entregables, estándares asociados y justificación	5
2	Estructura del conjunto de datos de emociones y sentimientos	21
3	Comparación de modelos de reconocimiento de voz (ASR)	23
4	Muestra del conjunto de prueba para modelos de emoción y sentimiento.	24
5	Comparación de modelos de análisis de sentimiento evaluados con métricas estándar .	24
6	Comparación de resultados obtenidos entre modelos base y modelos <i>fine-tuned</i> . Las filas resaltadas corresponden a modelos ajustados de elaboración propia.	25
7	Resultados de evaluación de modelos de análisis de emociones sobre el conjunto etiquetado con emociones primarias de Plutchik	26
8	Resultados de evaluación de modelos de análisis de emociones sobre el conjunto etiquetado con emociones primarias de Plutchik. (La fila resaltada corresponde al modelo ajustado)	27
9	Muestra del conjunto de prueba adicional para modelos de emoción.	27
10	Resultados de evaluación de modelos de análisis de emociones sobre un conjunto multi-clase en español y conjunto propio. La fila resaltada corresponde a un modelo ajustado de elaboración propia.	28

ABSTRACT

This project presents the design and implementation of *myMind*, a mobile application that supports emotional well-being in university students through voice-based self-expression. Users can record voice notes that are automatically transcribed and analyzed using natural language processing techniques to detect emotions and sentiments. The application is built on a functional architecture that integrates NLP models to generate personalized emotional reports over time. By offering a secure, private, and intuitive space for reflection, *myMind* promotes emotional awareness and provides meaningful feedback, helping students better understand their mental well-being.

I INTRODUCCIÓN

En los últimos años, el bienestar emocional y la salud mental de los estudiantes universitarios han cobrado una relevancia creciente en los entornos académicos, especialmente ante el aumento de síntomas de ansiedad, estrés y sentimientos de aislamiento. Sin embargo, muchas de las estrategias institucionales para atender estas problemáticas suelen requerir una activación voluntaria por parte del estudiante, lo que limita su alcance cuando aún no se ha identificado plenamente la necesidad de buscar ayuda. En este contexto, la tecnología ofrece oportunidades significativas para complementar los mecanismos tradicionales de apoyo psicoemocional.

Este proyecto se centra en el desarrollo de una aplicación móvil funcional, llamada *myMind*, diseñada para ofrecer a los estudiantes un espacio seguro, privado y accesible para registrar su estado emocional mediante notas de voz. A partir de dichas notas, se emplean técnicas de procesamiento de lenguaje natural (NLP) para analizar emociones y sentimientos, generar reportes personalizados y brindar un seguimiento continuo que permita al usuario observar cambios y patrones en su estado de ánimo a lo largo del tiempo. De esta manera, la aplicación no solo fomenta el autoconocimiento, sino que también actúa como un complemento a los servicios de apoyo existentes.

El presente documento detalla el proceso completo de diseño, desarrollo e implementación de *myMind*, abordando tanto sus componentes técnicos como conceptuales. A lo largo de sus secciones, se presentan los antecedentes técnicos, los objetivos del proyecto, los componentes del sistema, las decisiones metodológicas adoptadas, y los resultados obtenidos. Este documento constituye así una guía comprensiva del trabajo realizado y de su contribución a la intersección entre tecnología, salud mental y experiencia estudiantil.

II DESCRIPCIÓN GENERAL

1 Oportunidad, Problema

1.1 Contexto del Problema

La Pontificia Universidad Javeriana es una institución educativa cuya misión es “la formación integral de personas que sobresalgan por su alta calidad humana, ética, académica, profesional y por su responsabilidad social” [1]. Para lograr esta misión, la universidad tiene un compromiso institucional por el cuidado de la salud y el bienestar, lo cual incluye el bienestar emocional y la salud mental [2]. Según el Boletín Estadístico, en el 2022 tan solo 1943 de los 23968 estudiantes reportados hicieron uso del Centro de Asesoría Psicológica y Salud, en relación con la línea de salud mental [3]. Esto es especialmente relevante, ya que la tasa de suicidios en 2023 se incrementó en 15,73 % con respecto a 2022 [4], y según cifras del Ministerio de Salud, en Colombia aproximadamente el 44,7 % de los niños muestran indicios de afectaciones en su salud mental [5]. Esta cifra incluye jóvenes entre los 17 y 24 años, lo cual es el rango de edad promedio donde los estudiantes cursan grados universitarios [6].

A pesar de los esfuerzos y de las oportunidades brindadas por la universidad, la gran mayoría de los estudiantes prefieren no utilizar el Centro de Asesoría Psicológica y Salud. No obstante, la situación se agrava cuando consideramos que muchos estudiantes no son plenamente conscientes de su propio estado mental y emocional. Esta falta de autoconocimiento puede acumular estrés, ansiedad y otros problemas de salud mental que no se tratan adecuadamente a tiempo [7]. Esto es especialmente relevante en la población de estudiantes universitarios, ya que en una investigación que tenía como objetivo establecer la relación entre autoconocimiento y autorregulación emocional en universitarios colombianos, se encontró que sí existe una fuerte relación entre estos [8]. El estudio demuestra que hay una asociación estadísticamente significativa entre los niveles de autoconocimiento emocional y los de autorregulación emocional, donde entre más alta fuera la percepción e identificación de emociones, mayor era la probabilidad de desarrollar tolerancia a la frustración. Adicionalmente, este estudio encontró una correlación entre el autoconocimiento emocional y la capacidad de hablar abiertamente de las emociones, lo cual genera una mayor autorregulación y mejor manejo de problemas cotidianos [9].

Una de las herramientas recomendadas para el cuidado de la salud mental es el uso de diarios. Se ha demostrado que los diarios tienen un impacto positivo en la salud mental [10], ya sea por la actividad de “desahogo”, el espacio de reflexión o la documentación de las situaciones mismas [11]. Incluso, la Enciclopedia de la Salud del Centro Médico de la Universidad de Rochester avala que existen beneficios de usar un diario. Específicamente, se menciona que llevar un diario permite la organización y priorización de problemas, el seguimiento periódico de los síntomas, y provee un espacio de reflexión que permite la identificación de emociones y pensamientos tanto positivos como negativos [12]. De esta manera, los diarios pueden ayudar o facilitar la generación de la autoconciencia, con la finalidad de mejorar el bienestar de una persona [13].

1.2 Formulación del Problema

A pesar de los esfuerzos institucionales, persistía un problema: la mayoría de los estudiantes no accedía a los servicios de apoyo emocional ni contaba con herramientas personales de autoconocimiento emocional. Esta situación generaba un riesgo potencial de acumulación de malestar sin atención oportuna, dificultando la búsqueda de ayuda profesional.

Además, existía una carencia de espacios seguros, privados y accesibles para expresar pensamientos y emociones. La ausencia de medios tecnológicos adaptados a estas necesidades contribuía a que muchos estudiantes postergaran la gestión activa de su salud emocional.

1.3 Propuesta de Solución

Se propuso desarrollar una aplicación móvil que permitiera a los usuarios expresar sus emociones y sentimientos mediante una nota de voz. La aplicación analizaba los datos recogidos de estos audios y generaba reportes sobre las emociones, los sentimientos, las entidades y los tópicos identificados en ellos. El objetivo frente a la problemática fue ofrecer a los usuarios un espacio seguro y privado donde pudieran desahogarse libremente, brindándoles la oportunidad de expresar sus emociones y pensamientos sin restricciones.

Esta solución se enmarcó en el área del desarrollo de software, al combinar la creación de una plataforma digital accesible y fácil de usar con un análisis de las emociones encontradas en las notas de voz generadas por los usuarios. A través de esta herramienta, se facilitó el seguimiento de cómo los usuarios se sentían a lo largo del tiempo, permitiendo un monitoreo periódico de su estado emocional.

1.4 Justificación de la Solución

La salud mental de los estudiantes universitarios fue identificada como un aspecto crítico para su bienestar y desempeño académico. Sin embargo, la falta de autoconocimiento emocional, sumada a la ausencia de medios accesibles para expresar emociones y pensamientos, continuó siendo un desafío significativo en la Pontificia Universidad Javeriana. El problema radicaba no solo en la dificultad de los estudiantes para identificar y nombrar sus emociones, sino también en la falta de un entorno seguro y privado donde pudieran realizar esta reflexión personal. Esta situación retrasaba el acceso al apoyo profesional, como el ofrecido por el Centro de Asesoría Psicológica y Salud, aumentando el riesgo de nuevos problemas emocionales.

Frente a esta situación, se desarrolló una herramienta que permitió a los estudiantes gestionar su bienestar emocional de manera proactiva. La solución desarrollada respondió a la necesidad de contar con herramientas accesibles que fomentaran el autoconocimiento emocional y la reflexión personal, sin pretender sustituir los canales formales de atención psicológica. La propuesta consistió en una aplicación móvil que utilizó procesamiento de lenguaje natural (NLP) para analizar las notas de voz de los usuarios, respondiendo directamente a esta necesidad. La aplicación ofreció un espacio privado donde los estudiantes pudieron expresar libremente sus pensamientos y emociones, fomentando el autoconocimiento de su estado mental. Además, al analizar los audios y generar reportes sobre emociones, sentimientos y otros aspectos relevantes, la herramienta proporcionó a los usuarios una visión clara de su estado emocional a lo largo del tiempo. Este enfoque buscó facilitar el reconocimiento temprano de problemas emocionales y promovió una gestión más óptima de la salud mental.

Este enfoque combinó los beneficios comprobados de los diarios emocionales digitales con tecnologías emergentes como el Procesamiento de Lenguaje Natural (PLN), lo cual permitió analizar sentimientos y emociones específicas a partir del lenguaje expresado en las notas de voz. Gracias a esta capacidad, fue posible generar reportes personalizados y útiles para el usuario. El uso del PLN en el ámbito de la salud ha cobrado relevancia en los últimos años [8]; de hecho, se ha identificado que “el análisis de las emociones ha encontrado un nicho valioso en el sector de la salud, particularmente al abordar problemas de salud mental” [14]. Todo esto evidencia el potencial de estas tecnologías para apoyar la salud emocional desde enfoques innovadores y accesibles.

Finalmente, el proyecto se alineó con la misión de la Pontificia Universidad Javeriana de promover la formación integral de sus estudiantes. Si bien la aplicación fue concebida como un recurso adicional a los servicios existentes, su implementación a pequeña escala contribuyó a un entorno universitario más equilibrado y saludable, sirviendo como un complemento útil a las iniciativas ya presentes en la universidad.

2 Descripción del Proyecto

El proyecto consistió en el desarrollo de una aplicación móvil diseñada para apoyar y mejorar el bienestar emocional y la salud mental de los estudiantes. La aplicación permitió a los usuarios registrar sus emociones mediante notas de voz en cualquier momento. Utilizando técnicas de procesamiento de lenguaje natural, la aplicación analizaba las emociones y sentimientos expresados en las grabaciones. A partir de este análisis, la aplicación generaba reportes periódicos que mostraban los resultados obtenidos, lo que facilitaba a los estudiantes comprender y gestionar su estado emocional a lo largo del tiempo.

2.1 Objetivo General

Desarrollar una aplicación móvil que, a través de notas de voz, ofreciera a los estudiantes un espacio seguro y privado para expresarse, brindando además un seguimiento periódico de su estado emocional.

2.2 Objetivos Específicos

- Diseñar una aplicación móvil que permita la recolección y transcripción a texto de notas de voz, asegurando un espacio seguro y privado para los estudiantes.
- Aplicar modelos de procesamiento de lenguaje natural (NLP) para extraer información relevante en las notas de voz.
- Implementar el sistema de adquisición y procesamiento integrado en la aplicación móvil.
- Desarrollar un sistema de generación de reportes personalizados, acerca de los datos ingresados y los resultados obtenidos por el modelo.

2.3 Entregables, Estándares y Justificación

Entregable	Estándares asociados	Justificación
Documento de diseño de software	IEEE 1016-2009: Software design descriptions (SDD). UML para la representación gráfica de la arquitectura y diseño del software.	El uso del estándar IEEE 1016-2009 asegura que el diseño de software se documente de manera coherente y detallada, facilitando la comprensión de la estructura, comportamiento y relaciones de los componentes del sistema. Adicionalmente, el uso de UML apoya la representación visual de la arquitectura y el diseño del software.
Aplicación móvil funcional		
Documentación de los Modelos de reconocimiento de emociones	ISO/IEC 12207: ciclo de vida del desarrollo de los modelos y su integración en la aplicación	La ISO/IEC 12207 asegura que el desarrollo de los modelos de reconocimiento de emociones siga un enfoque estructurado y bien definido, cubriendo todas las fases del ciclo de vida.
Documento de política de tratamiento de datos	Ley de protección de datos personales o ley 1581 del 2012	El marco legal como referencia asegura que la aplicación cumpla con los requerimientos legales.
Memoria del trabajo de grado	Estándar definido por la Pontificia Universidad Javeriana - Ingeniería de Sistemas PUJ	
Documento de especificación de requerimientos de software	Estándar definido por la Pontificia Universidad Javeriana - Ingeniería de Sistemas PUJ	

Cuadro 1: Entregables, estándares asociados y justificación

III CONTEXTO DEL PROYECTO

1 Transfondo

1.1 Procesamiento de Lenguaje Natural (NLP)

Para comprender lo que es el procesamiento de lenguaje natural primeramente hay que entender lo que es el lenguaje.

El lenguaje se define como un medio a través del cual los seres humanos se comunican y expresan sus pensamientos, basado en la asociación de signos con significados específicos. Este proceso se apoya en diversas herramientas como la escritura, las señales y la voz para lograr la comunicación. Dentro de este contexto, existe el lenguaje natural, que incluye los idiomas como el español o inglés. Estos lenguajes evolucionan constantemente sin seguir estrictamente reglas fijas.

El Procesamiento del Lenguaje Natural es un área de estudio que se enfoca en comprender cómo funciona el lenguaje, su estructura, y la creación de nuevo lenguaje, abarcando todas las actividades relacionadas con su tratamiento. Algunas de estas actividades incluyen la generación de texto, traducción entre distintos idiomas, sistemas de preguntas y respuestas, el desarrollo de chatbots, entre otras aplicaciones [15].

Las principales tareas del Procesamiento del Lenguaje Natural incluyen descomponer y analizar datos de texto y voz para que los ordenadores puedan interpretarlos. Entre estas tareas se destacan el reconocimiento de voz, que convierte datos de voz en texto, y el etiquetado de partes del discurso, que identifica la función gramatical de las palabras según su contexto. También está la desambiguación del sentido de las palabras, que selecciona el significado adecuado de términos con múltiples acepciones, y el reconocimiento de entidades nombradas (NER), que identifica lugares, personas u objetos mencionados en el texto. Además, la resolución de correferencias determina cuándo dos palabras se refieren a la misma entidad, y el análisis de sentimientos extrae emociones y actitudes de un texto. Por último, la generación de lenguaje natural convierte datos estructurados en lenguaje comprensible para los humanos [16].

1.2 Conversión de Voz a Texto (STT)

La conversión de voz a texto es una tecnología que se encarga de transformar el habla a un formato de texto por medio de algoritmos de reconocimiento de voz. Este proceso comienza capturando los sonidos de una grabación, los cuales serán convertidos posteriormente a un formato digital. Esto se logra gracias a los algoritmos lingüísticos que son capaces de clasificar las señales emitidas por el habla humana para convertirlos en caracteres interpretables por los computadores. Estos sonidos son segmentados en pequeñas fracciones de tiempo que luego serán combinados con fonemas (la unidad básica del sonido que distingue una palabra de otra) para finalmente compararlos con frases o palabras ya conocidas por medio de un modelo matemático basado en *machine learning* [17].

1.3 Salud Mental

La salud mental se refiere a un estado de bienestar emocional y psicológico que permite a las personas gestionar el estrés, desarrollar habilidades, aprender, trabajar de manera eficaz e integrarse en su entorno social. Es una dimensión clave del bienestar general, ya que influye en la capacidad de las personas para enfrentar los desafíos de la vida diaria. La salud mental está influenciada por diversos factores, tanto individuales como familiares, ambientales y estructurales, que pueden tener

un impacto positivo o negativo. Aunque muchas personas muestran resiliencia ante las dificultades, aquellas que enfrentan situaciones adversas, como la pobreza, la violencia, las discapacidades o las desigualdades, corren un mayor riesgo de experimentar problemas de salud mental [18].

1.4 Modelos de Aprendizaje

Un modelo de aprendizaje automático es un sistema entrenado para reconocer patrones específicos en los datos. El proceso de entrenamiento consiste en aplicar un algoritmo a un conjunto de datos, permitiendo al modelo aprender y extraer características clave de la información. Una vez entrenado, el modelo puede analizar datos nuevos, nunca vistos, y generar predicciones precisas basadas en lo que ha aprendido. Este enfoque es esencial para tareas como la clasificación, la regresión y la toma de decisiones automatizadas, donde la capacidad de generalizar a partir de datos pasados es fundamental [19].

2 Análisis del Contexto

El desarrollo de soluciones tecnológicas para la salud mental ha crecido significativamente en los últimos años. Existen aplicaciones móviles enfocadas en el bienestar emocional y herramientas basadas en inteligencia artificial (IA) y procesamiento de lenguaje natural (PLN) para el análisis de emociones y detección temprana de trastornos. A continuación, se presenta un análisis de algunas de estas soluciones y estudios relevantes, con el objetivo de evidenciar las fortalezas de la solución propuesta.

2.1 Aplicaciones móviles centradas en la salud mental

Diversas aplicaciones buscan apoyar el bienestar emocional mediante registros diarios, evaluaciones psicológicas o chatbots. A continuación, se destacan algunas de las más representativas:

WYSA: Utiliza un chatbot basado en IA para ofrecer apoyo emocional, ejercicios de terapia cognitivo-conductual (TCC) y técnicas de relajación. Se apoya en modelos avanzados de PLN que interpretan las emociones del usuario y adaptan las respuestas. También incorpora aprendizaje automático para personalizar recomendaciones y seguimiento emocional con visualizaciones.

MindDoc: Se enfoca en el monitoreo de la salud mental a través de autoevaluaciones, cuestionarios clínicos y seguimiento del estado emocional. Emplea algoritmos para adaptar el contenido a cada usuario, análisis de datos para generar reportes y sugerencias basadas en IA.

Daylio: Permite registrar el estado de ánimo y hábitos diarios mediante selecciones rápidas, sin necesidad de escribir. Su fortaleza radica en la visualización de datos con gráficos que muestran patrones de comportamiento a lo largo del tiempo.

Affinity: Es una app que conecta personas con experiencias similares en salud mental. Los usuarios crean un perfil, seleccionan condiciones que padecen y se emparejan con otros para conversar en privado. Utiliza IA para mejorar el emparejamiento en función de las descripciones proporcionadas. Los estudios preliminares muestran mejoras en niveles de ansiedad tras su uso [20].

RAAH.ai: Un chatbot conversacional para aliviar el estrés durante situaciones de aislamiento. Integra PLN con redes neuronales (CNN y RNN) para identificar emociones y responde utilizando síntesis de voz personalizada (voice cloning) con una tasa de similitud del 91 %. Su diseño contempla aspectos de seguridad para prevenir suplantaciones de voz [21].

Limitaciones observadas: Aunque estas aplicaciones ofrecen funciones útiles, la mayoría se limitan a interacciones textuales o registros predefinidos. No incluyen una dimensión oral (voz), lo cual limita la expresión espontánea y emocional del usuario. Además, pocas ofrecen una personalización

profunda del lenguaje o combinan múltiples modalidades (texto, voz y emociones) en un mismo entorno.

2.2 Estudios e investigaciones sobre PLN aplicado a la salud mental

El uso del procesamiento de lenguaje natural (PLN) en el ámbito médico y psicológico ha avanzado considerablemente, especialmente en la detección automática de señales de riesgo emocional.

Un ejemplo relevante es el artículo “Resumen de la Tarea MentalRiskES en IberLEF 2024: Detección Precoz del Riesgo de Trastornos Mentales en Español”. Este trabajo formó parte de la 40ª edición de la Conferencia Internacional de la Sociedad Española para el PLN, y propone tareas competitivas para identificar riesgos de depresión y ansiedad a partir de publicaciones en redes sociales. MentalRiskES ha demostrado avances significativos en la identificación automática de señales de riesgo, a través del análisis semántico, clasificación emocional y detección de patrones lingüísticos [22].

Otros estudios, como el desarrollado para RAAH.ai, exploran cómo los modelos de PLN y redes neuronales pueden combinarse con tecnologías de voz para crear experiencias conversacionales personalizadas[21]. Esto demuestra el interés creciente en integrar aspectos emocionales y contextuales al análisis del lenguaje, no solo como texto escrito, sino también hablado.

2.3 Valor añadido de la solución propuesta

La solución desarrollada en este proyecto se diferencia significativamente de las anteriores al combinar tres componentes clave:

- **Expresión hablada:** Permite que los usuarios se desahoguen mediante voz, favoreciendo una expresión emocional más natural y fluida. El sistema convierte voz a texto, preservando la intención emocional.
- **Análisis de emociones en texto y voz:** Utiliza PLN para extraer las emociones y sentimientos identificadas en los audios que graba el usuario.
- **Seguimiento continuo y personalizado:** Registra entradas del usuario a lo largo del tiempo que permite visualizar la evolución del bienestar de manera clara y segura.

Esta integración supera las limitaciones de las aplicaciones actuales, que suelen centrarse en registros de texto o formularios predefinidos. Al incorporar el lenguaje hablado, el análisis emocional y una plataforma intuitiva, la solución propuesta ofrece un entorno más humano, empático y efectivo para el monitoreo de la salud mental.

IV ANÁLISIS DEL PROBLEMA

Este capítulo presenta un análisis detallado del problema que busca resolver la aplicación. Se identifican los requerimientos funcionales fundamentales, las necesidades principales de los usuarios y las funcionalidades necesarias para asegurar un correcto desempeño del sistema. El análisis se ha estructurado en módulos que agrupan los requerimientos relacionados, permitiendo una visión ordenada y coherente de las capacidades que la plataforma debe ofrecer.

1 Requerimientos

A continuación, se resumen los principales requerimientos funcionales identificados para el desarrollo del sistema, organizados por módulo:

1.1 Módulo de Autenticación y Gestión de Usuarios

- **Inicio de sesión con credenciales (RF-001):** El sistema debe permitir que los usuarios ingresen con su nombre de usuario y contraseña válidos. En caso de error, debe mostrarse un mensaje informativo.
- **Validación de criterios de seguridad de contraseña (RF-002):** Toda contraseña ingresada debe cumplir requisitos de seguridad mínimos: longitud de al menos 8 caracteres, inclusión de una mayúscula, un número y un carácter especial.
- **Registro de nuevo usuario (RF-003):** La plataforma debe habilitar la creación de nuevas cuentas, validando que el nombre de usuario no esté registrado previamente.
- **Gestión de datos personales (RF-004):** El sistema debe permitir la edición o eliminación de los datos personales de un usuario registrado.

1.2 Módulo de Grabación y Envío de Notas de Voz

- **Grabación de notas de voz con límite de tiempo (RF-005):** El usuario debe poder grabar mensajes de voz de hasta 60 segundos de duración.
- **Pausar y reanudar grabación de notas de voz (RF-006):** Durante la grabación, debe ser posible pausar y reanudar el proceso, deteniendo el cronómetro en pausa.
- **Borrar grabación de nota de voz (RF-007):** El usuario debe poder cancelar y eliminar una grabación antes de enviarla a procesamiento.

1.3 Módulo de Transcripción de Audio

- **Transcripción automática de notas de voz (RF-008):** Una vez finalizada una grabación, el sistema debe transcribir automáticamente el audio a texto.
- **Almacenamiento de transcripción y eliminación de nota de voz (RF-009):** El texto transcrito, junto con los análisis realizados, debe almacenarse, y el archivo de audio original debe eliminarse posteriormente.

1.4 Módulo de Procesamiento de Lenguaje Natural (NLP)

- **Análisis emocional de texto transcrito (RF-010):** El sistema debe identificar las emociones principales presentes en el texto transcrito.
- **Análisis de sentimiento del texto transcrito (RF-011):** Además de emociones, se debe clasificar el sentimiento general en positivo, negativo o neutral.
- **Identificación de palabras clave en texto transcrito (RF-012):** El sistema debe extraer las palabras más relevantes del texto, omitiendo conectores y términos comunes.

1.5 Módulo de Almacenamiento y Data Warehouse

- **Almacenamiento estructurado de resultados de análisis emocional (RF-013):** Los resultados de análisis deben almacenarse de forma estructurada y asociarse correctamente a cada usuario.
- **Eliminación automática de grabaciones de voz (RF-014):** El sistema debe eliminar automáticamente las grabaciones de voz una vez completados el análisis y el almacenamiento de los resultados.

2 Restricciones

En esta sección se presentan las restricciones generales, de software y hardware, así como las limitaciones que afectan al desarrollo y operación de la aplicación.

2.1 Restricciones Generales

- Los usuarios no pueden modificar los modelos de aprendizaje ni acceder a funciones de administración.
- La aplicación solo está destinada a usuarios mayores de 18 años o con el consentimiento de un tutor legal.
- Todos los datos transmitidos (audios, reportes, etc.) deben ser cifrados para garantizar la privacidad y seguridad de la información del usuario.
- La aplicación debe ser capaz de recuperar información en caso de un fallo inesperado, como una caída del servidor o pérdida de datos.
- La aplicación solo trabajará para el idioma español.
- La aplicación debe ser capaz de procesar audios de hasta 1 minuto de duración sin afectar el rendimiento del sistema.

2.2 Restricciones de Software

- Al utilizar servicios de terceros para el análisis de audio y emociones o cualquier otro tipo de API, se deben gestionar correctamente las claves de acceso y sus límites de uso.

- Se utilizará exclusivamente lenguajes de programación para desarrollo móvil de aplicaciones Android como lo es kotlin o java y para el desarrollo en general se utilizarán lenguajes como lo son python.
- Se deben establecer los requisitos mínimos del sistema para la instalación, tanto en la máquina local como en la infraestructura de servidores.
- Al utilizar modelos de aprendizaje preentrenados o plataformas de terceros, deben obtenerse licencias adecuadas que permitan su uso dentro de la aplicación sin violar derechos de autor ni restricciones de uso.

2.3 Restricciones de Hardware

- El dispositivo debe contar con una tarjeta de red para conectarse a Internet y permitir la conexión con el servidor donde se almacenan los modelos.
- Se requiere una versión de Android 8 o superior.
- El servidor debe ser capaz de procesar un mínimo de dos modelos de aprendizaje de manera simultánea sin perder rendimiento. Los modelos específicos se definirán más adelante en el proyecto.

2.4 Limitaciones

- La seguridad está limitada al alcance de los permisos que los usuarios otorgan a la aplicación (como el acceso al micrófono). Si un usuario niega permisos, la funcionalidad puede verse afectada.
- La aplicación incorpora autenticación básica (usuario y contraseña), sin opciones avanzadas como autenticación multifactor o biométrica, debido a la complejidad de integración en una aplicación móvil ligera.
- La aplicación depende de una conexión a Internet estable para cargar audios y descargar modelos de aprendizaje, lo que puede afectar la funcionalidad en áreas con baja cobertura.
- La aplicación necesita enviar datos a servidores externos para ciertos procesos. Si el servidor falla o se encuentra en mantenimiento, la aplicación no podrá procesar audios en tiempo real, afectando la disponibilidad del servicio.

3 Especificación Funcional

Esta sección describe las principales funcionalidades que ofrecerá el sistema myMind, basándose en las historias de usuario recopiladas.

3.1 E1: Sistema de gestión de usuarios

- **HU-01: Registro de nuevo usuario**
Permite crear una cuenta solicitando nombre, correo electrónico, datos demográficos y contraseña. Se valida la unicidad del correo y la fortaleza de la contraseña. Se confirma el registro visualmente y mediante correo electrónico.

- **HU-02: Inicio de sesión**
Permite a los usuarios autenticarse mediante correo y contraseña, ofreciendo una opción de Recordarme”para mantener la sesión activa.
- **HU-03: Recuperación de contraseña**
Facilita la recuperación de acceso enviando un correo con instrucciones para restablecer la contraseña.
- **HU-04: Gestión de cuenta**
Permite la modificación de datos personales o la eliminación de la cuenta.

3.2 E2: Captura y procesamiento de audio

- **HU-05: Grabación de nota de voz**
Permite grabar notas de voz de hasta 1 minuto, mostrando un cronómetro y notificaciones al acercarse al límite de tiempo.
- **HU-06: Control de grabación**
Ofrece controles para pausar, reanudar o cancelar la grabación con confirmación previa en caso de cancelación.

3.3 E3: Análisis emocional mediante NLP

- **HU-07: Análisis de emociones**
Detecta emociones primarias en los textos transcritos, mostrando intensidad de cada emoción.
- **HU-08: Análisis de sentimiento**
Determina la polaridad general (positiva, negativa o neutral) del contenido.
- **HU-09: Identificación de palabras clave**
Extrae y muestra las palabras más relevantes asociadas al contenido emocional del usuario.

3.4 E4: Gestión de datos y almacenamiento

- **HU-10: Almacenamiento de resultados**
Guarda las transcripciones y análisis emocionales de forma asociada al perfil de usuario.
- **HU-11: Consulta de historial por fecha**
Permite visualizar análisis previos organizados cronológicamente.
- **HU-12: Consulta por tipo de emoción**
Ofrece la opción de filtrar resultados según la emoción detectada.
- **HU-13: Gestión del almacenamiento**
Optimiza el almacenamiento eliminando archivos de audio originales y gestionando políticas de retención de datos.

3.5 E5: Visualización y reportes

- **HU-14: Visualización de evolución emocional**

Genera gráficos que muestran la evolución de las emociones del usuario en el tiempo.

- **HU-15: Reporte de sentimiento general**

Presenta la tendencia general de sentimientos (positivos, negativos, neutrales) en formato gráfico.

- **HU-16: Políticas de seguridad**

Informa al usuario sobre las políticas de privacidad y manejo de datos, requiriendo aceptación explícita durante el registro.

V DISEÑO DE LA SOLUCIÓN

1 Antecedentes teóricos

En los últimos años, el desarrollo de tecnologías de inteligencia artificial ha permitido avances significativos en el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) y el análisis de emociones. Estas innovaciones han facilitado la creación de modelos capaces de interpretar, analizar y generar texto a partir de diversas fuentes, incluyendo audio, texto escrito y conversaciones en lenguaje natural.

Dentro del contexto de aplicaciones como myMind, tres componentes fundamentales han sido ampliamente estudiados en la literatura: la transcripción de voz a texto (Speech-to-Text), el análisis de sentimientos y la detección de emociones. Cada uno de estos aspectos ha evolucionado con la adopción de modelos de aprendizaje profundo, arquitecturas híbridas y técnicas de procesamiento multimodal, lo que ha permitido mejorar la precisión y aplicabilidad en distintos escenarios [23]-[25].

En las siguientes subsecciones se abordará la fundamentación teórica de cada una de estas áreas, explorando sus principales modelos, técnicas y desafíos actuales.

1.1 Modelos de Reconocimiento de Voz

Al ser la fuente principal de información que se obtendrá desde los usuarios para utilizar la aplicación, es importante que la transcripción del audio, o el “speech-to-text” que se realiza en la aplicación.

Según Acheampong et al. [26], los avances recientes en la transcripción automática del habla han mejorado significativamente la precisión mediante innovaciones en modelos acústicos, adaptación de hablantes y modelos de lenguaje. Estos avances han permitido reducir los errores en la transcripción y mejorar la portabilidad de los modelos para diferentes idiomas y estilos de habla.

Los primeros sistemas de transcripción automática se basaban en modelos ocultos de Markov y modelos de mezcla gaussiana, los cuales estimaban la probabilidad de secuencias de fonemas a partir de características acústicas extraídas del audio. Aunque estos métodos fueron fundamentales en el desarrollo de la transcripción automática, presentan limitaciones al manejar variabilidad en acentos, ruido de fondo y diferencias en la velocidad del habla.

El uso de redes neuronales profundas ha revolucionado el campo de la transcripción de voz. Modelos basados en arquitecturas como LSTMs y Transformers han permitido mejorar la precisión al capturar dependencias a largo plazo en el habla. Modelos recientes como DeepSpeech, Wav2Vec 2.0 y Whisper han demostrado ser altamente efectivos, reduciendo los errores de transcripción incluso en condiciones ruidosas [27], [28].

A pesar de estas mejoras, la implementación de estos modelos sigue teniendo desafíos en términos de rendimiento, ya que algunos sistemas pueden requerir tiempos de procesamiento significativamente superiores al tiempo real. En algunos casos optimizados, se ha logrado reducir este tiempo a valores cercanos a tres veces la duración del audio procesado, lo que representa una mejora considerable en eficiencia sin comprometer la precisión de la transcripción.

El reconocimiento de voz automático (ASR, por sus siglas en inglés) y la conversión de habla a texto (Speech-to-Text, S2T) han experimentado un avance significativo gracias a la incorporación de modelos de aprendizaje profundo y arquitecturas multimodales. La evolución de estas técnicas ha permitido mejorar la precisión de la transcripción y optimizar el procesamiento de grandes volúmenes de datos de audio en diversas aplicaciones. La literatura ha identificado distintos enfoques que han sido desarrollados para abordar los desafíos del ASR, cada uno con sus propias ventajas y limitaciones

en términos de eficiencia, precisión y aplicabilidad en distintos contextos [29].

Uno de los enfoques más recientes es la utilización de modelos de lenguaje a gran escala (Large Language Models, LLMs) combinados con arquitecturas de doble codificador (Dual Encoders, DE). Palma Gómez et al. [30] proponen un sistema de recuperación de información en habla y texto basado en LLMs, que permite alinear representaciones de voz y texto en un mismo espacio de embeddings. Este enfoque elimina la necesidad de entrenamiento previo en datos de habla y mejora la correspondencia entre ambas modalidades. La supervisión débil y el aprendizaje no supervisado han sido clave en la mejora de estos sistemas, permitiendo reducir la dependencia de grandes volúmenes de datos etiquetados.

Otro enfoque alternativo es el uso de modelos de lenguaje aumentados con capacidades de voz mediante la técnica de ajuste fino supervisado en una sola etapa (Single-Stage Supervised Fine-Tuning, SFT). Peng et al. [31] presentan el modelo VoiceTextBlender, que integra simultáneamente datos de voz y texto en un LLM base, optimizando su rendimiento en tareas multimodales sin comprometer la calidad en procesamiento de texto. Este modelo se basa en la optimización mediante adaptación de baja dimensión (Low-Rank Adaptation, LoRA), permitiendo mantener la estructura original del modelo sin modificar excesivamente sus parámetros.

Una estrategia distinta es el uso de modelos híbridos que combinan redes de reconocimiento de voz preentrenadas (Large Pre-trained Speech Models, LPSM) con motores de traducción automática neuronal (Neural Machine Translation, NMT). Avila y Crego [32] proponen un sistema que integra un LPSM con un motor NMT multilingüe, lo que permite mejorar la precisión y eficiencia en entornos industriales. Este enfoque reduce la propagación de errores y optimiza los recursos computacionales al utilizar representaciones de voz preentrenadas como entrada para la red de traducción.

En el ámbito de la transcripción de audio en largo plazo, los modelos han evolucionado para mejorar la precisión y alineación temporal. Whisper y su extensión WhisperX han introducido mejoras en la segmentación del audio y en la detección de actividad de voz (VAD), permitiendo una transcripción más precisa y organizada. Bain et al. [27] presentan WhisperX como un avance significativo al incorporar alineación fonética forzada, lo que garantiza una mayor exactitud en la asignación de marcas de tiempo a las palabras.

Por otro lado, las investigaciones en el reconocimiento de voz para idiomas con pocos recursos han demostrado que el preentrenamiento en ASR con una lengua de mayor disponibilidad de datos puede mejorar la precisión en traducción de voz a texto en idiomas menos representados. Stoian et al. (2020) encontraron que el preentrenamiento en idiomas con grandes corpus, incluso si no están directamente relacionados con la lengua objetivo, puede proporcionar mejoras sustanciales en la precisión de la transcripción [33]. Este hallazgo es relevante al considerar la adaptabilidad de modelos ASR en entornos con diversidad lingüística.

Los estudios recientes en ASR también destacan la relevancia de los modelos end-to-end, especialmente aquellos basados en Transformers y Conformers, que han demostrado superar a las arquitecturas tradicionales basadas en Hidden Markov Models (HMM) y Gaussian Mixture Models (GMM) en tareas de reconocimiento de voz [34]. Estos modelos han sido optimizados para manejar la variabilidad en los acentos, el ruido en el entorno y la escasez de datos en idiomas con pocos recursos, permitiendo una mejor adaptabilidad en entornos reales.

A medida que se desarrollan estas técnicas, se han identificado varias consideraciones para la selección de un modelo óptimo. La eficiencia computacional es un factor clave, ya que algunos enfoques, como los modelos híbridos y los basados en LLMs, requieren mayores recursos de procesamiento. Asimismo, la capacidad de generalización de un modelo es crucial para garantizar un rendimiento adecuado en diversas condiciones de audio, incluyendo variaciones en el tono, el acento y el ruido de

fondo. Otro aspecto fundamental es la robustez frente a cambios en el contexto lingüístico y en la calidad del audio de entrada, lo que puede influir en la precisión de la transcripción.

A partir de estos enfoques, se pueden identificar distintos modelos que pueden adaptarse mejor a entornos específicos, ya sea optimizando la recuperación de información en voz, la transcripción precisa de grabaciones de larga duración o la integración de múltiples idiomas en un mismo sistema de ASR. La selección del modelo dependerá de las necesidades específicas del sistema y del entorno en el que se desplegará, asegurando que la precisión, eficiencia y escalabilidad sean adecuadas para la tarea en cuestión.

1.2 Modelos de Reconocimiento de Sentimientos

Para comprender el funcionamiento de los modelos de análisis de sentimientos, es fundamental definir inicialmente qué se entiende por "sentimiento". Uno de los principales desafíos en esta área de estudio es la definición de los objetos de análisis, como la opinión y la subjetividad.

Según Taboada [23], la subjetividad se refiere a un estado privado que no está abierto a la observación o verificación objetiva. Estos estados privados incluyen emociones, opiniones y especulaciones, entre otros. Debido a la naturaleza abstracta de estos conceptos, cuantificar y analizar los sentimientos resulta una tarea compleja.

A nivel práctico, como menciona Taboada [23], muchos investigadores señalan que el sentimiento suele definirse en términos de polaridad, es decir, como una opinión positiva o negativa. Esta aproximación facilita su análisis al centrarse en la clasificación del sentimiento expresado en un texto sin profundizar en su contexto subjetivo o su grado de veracidad.

Con respecto al análisis de sentimientos como campo de estudio, este se considera parte de la lingüística computacional al igual que el procesamiento de lenguaje natural; de igual manera, está estrechamente relacionado con ramas como la psicología y teorías de evaluación. Según Scherer [24], la teoría de la evaluación plantea que las emociones surgen y se diferencian a partir de la evaluación subjetiva que una persona hace sobre la importancia personal de una situación, objeto o evento. Esto refuerza la idea de que el análisis de sentimientos no solo depende del contenido del texto, sino también del contexto y la interpretación subjetiva de quien lo expresa.

En términos generales, el análisis de sentimientos puede dividirse en dos categorías: "Explícito", cuando una oración expresa directamente una opinión, por ejemplo, ".^{E1} clima es hermoso", y "Implícito", cuando el texto sugiere una opinión sin expresarla directamente, como en "Los audífonos se rompieron en dos días".

Dentro del análisis de sentimientos existen diferentes niveles en los que ha sido investigado. Según Wankhade et al. [25], estos niveles incluyen el análisis a nivel de documento, oración, frase y aspecto, cada uno con un enfoque más detallado:

- **Nivel de documento:** evalúa el sentimiento general de un documento completo, asignándole una única polaridad (positiva, negativa o neutral). Se utiliza, por ejemplo, para clasificar capítulos o páginas de libros.
- **Nivel de oración:** analiza el sentimiento de cada oración de forma independiente, permitiendo identificar la combinación de emociones dentro de un texto.
- **Nivel de frase:** examina palabras o expresiones clave dentro de una frase para identificar aspectos específicos del sentimiento expresado. Es útil en reseñas de productos donde una sola frase puede contener opiniones sobre distintos aspectos.

- **Nivel de aspecto:** asigna una polaridad a cada aspecto mencionado en una oración y luego calcula un sentimiento agregado. Esto permite un análisis más detallado cuando un texto aborda múltiples temas en una misma oración.

Para el análisis de sentimientos, una de las estrategias utilizadas es el aprendizaje automático, como menciona Taboada [23]. Este enfoque permite clasificar textos o audios según su carga emocional mediante técnicas de procesamiento del lenguaje natural y modelos de clasificación.

Con respecto al componente de procesamiento de lenguaje natural en el análisis de sentimientos, este campo ha crecido exponencialmente debido al aumento en la generación de contenido textual en plataformas digitales, redes sociales, foros y blogs. La capacidad de analizar, clasificar y extraer información relevante de estos textos permite a empresas, investigadores y gobiernos tomar decisiones informadas. En los últimos años, los modelos preentrenados (PMs) han revolucionado esta área, permitiendo mejorar significativamente el rendimiento de los modelos de aprendizaje automático en diversas tareas del NLP, incluyendo el análisis de sentimientos [35].

Sin embargo, el uso de estos modelos presenta ciertos desafíos, tales como la discrepancia entre el preentrenamiento y la adaptación a tareas específicas, así como la escasez de datos etiquetados para el entrenamiento supervisado. Además, los modelos de lenguaje a gran escala requieren una cantidad significativa de recursos computacionales para su entrenamiento y despliegue, lo que puede limitar su aplicabilidad en entornos con restricciones de infraestructura.

Esta rama ha evolucionado desde enfoques basados en diccionarios y aprendizaje automático hasta modelos de redes neuronales profundas y transformers. Inicialmente, se usaban técnicas basadas en reglas y diccionarios de palabras con carga emocional predefinida, que aunque útiles en ciertos contextos, tenían dificultades para capturar la variabilidad lingüística.

Con la introducción de redes neuronales profundas como RNN, LSTM y CNN, el análisis de sentimientos mejoró al incorporar el contexto en la interpretación de los textos. No obstante, fue con la llegada de los modelos transformers, como BERT y GPT, que se alcanzaron avances significativos en la comprensión y representación del lenguaje. Estos modelos, al aprovechar mecanismos de autoatención, han mejorado notablemente la precisión en la clasificación de sentimientos y han allanado el camino para la adopción del aprendizaje por prompts.

El aprendizaje por prompts ha surgido como una solución prometedora para abordar los desafíos del preentrenamiento y la afinación de modelos. Esta técnica permite ajustar las entradas del modelo mediante plantillas, lo que facilita la adaptación de los PMs a tareas específicas. En el contexto del análisis de sentimientos, el aprendizaje por prompts se ha aplicado con éxito en diversas sub-tareas, como el Análisis de Sentimientos Basado en Aspectos (ABSA), detección de sarcasmo y análisis de sentimientos en múltiples idiomas [35].

Una de las principales ventajas del aprendizaje por prompts es que reduce la necesidad de un entrenamiento extenso en grandes conjuntos de datos etiquetados. En su lugar, permite reutilizar modelos preentrenados sin modificaciones extensivas, utilizando estructuras de prompts adecuadas para guiar la generación de respuestas precisas. Esto es particularmente valioso en escenarios donde los datos anotados son limitados o costosos de obtener. Además, permite añadir una capa de refinamiento donde es posible ajustar un modelo ya entrenado para contextos específicos. En el contexto de este proyecto, puede ser bastante valioso considerando las posibles restricciones que se presentan a la hora de encontrar un modelo específico para texto en español.

El uso de modelos preentrenados ha demostrado mejorar significativamente la precisión en las tareas de análisis de sentimientos. No obstante, el costo computacional de la afinación de estos modelos ha incrementado, generando la necesidad de desarrollar enfoques más eficientes. El aprendizaje por prompts ofrece una solución viable al reducir la brecha entre el preentrenamiento y la tarea específica,

permitiendo una mejor utilización de los PMs sin necesidad de un costoso proceso de reentrenamiento [35].

Un aspecto clave que refuerza la aplicabilidad de estos enfoques es el estudio de Bu et al. [35], que analizó más de 200 modelos de análisis de sentimientos. Su investigación proporciona una visión clara sobre la evolución de estas tecnologías y sustenta la razón detrás del auge del aprendizaje por prompts en esta área. En myMind, donde los usuarios expresan sus emociones a través de grabaciones de voz, que luego se transcriben a texto, la selección del modelo debe considerar la subjetividad del lenguaje y los posibles matices emocionales en las transcripciones. Por ello, se buscará un modelo que capture estos matices con la mayor precisión posible, permitiendo generar reportes personalizados y realizar un seguimiento efectivo del estado emocional del usuario que sean relevantes y lo más cercano a lo que el usuario puede estar sintiendo en realidad.

1.3 Modelos de Reconocimiento de Emociones

La detección de emociones es una rama del análisis de sentimientos que se enfoca en la extracción y el análisis de emociones más específicas. A diferencia del análisis de sentimientos mencionado anteriormente, que se limita a asignar una polaridad general (positiva, negativa o neutral), su objetivo es identificar estados emocionales como felicidad, tristeza, enojo y miedo, con el fin de lograr un análisis más detallado y significativo.

Como mencionan Acheampong, Wenyu y Nunoo-Mensah [26], las emociones desempeñan un papel crucial en la comunicación humana, ya que reflejan el estado emocional de un individuo y pueden influir en su toma de decisiones. Se derivan tanto de experiencias personales como de interacciones con el entorno, lo que introduce un alto grado de subjetividad en su análisis al igual que su rama de origen.

El análisis basado únicamente en texto enfrenta desafíos importantes. A diferencia de los métodos multimodales que incluyen señales de voz o imágenes, el texto carece de indicadores visuales o auditivos que faciliten la identificación de emociones. Esto hace que detectar emociones en textos cortos, mensajes con emojis o con errores gramaticales sea un reto considerable. Además, la constante evolución del lenguaje y la falta de diccionarios de emociones especializados dificultan aún más la automatización de este proceso.

Para comprender el área de estudio del análisis de emociones, es necesario definir algunos modelos teóricos de las emociones, ya que estos determinan cómo se representan. Los modelos asumen que las emociones existen en varios estados, por lo cual es necesario hacer una distinción entre ellas. Según Acheampong et al. [26], estos modelos pueden agruparse en dos grandes categorías: modelos discretos de emociones (DEMs, por sus iniciales en inglés) y modelos dimensionales de emociones (DiEMs, por sus iniciales en inglés).

1.4 Modelos Discretos de Emociones (DEMs)

Los modelos discretos clasifican las emociones en categorías separadas y bien definidas, considerando que cada emoción es independiente de las demás. Entre los más relevantes se encuentran el modelo de Ekman, el modelo de Plutchik y el modelo OCC.

El modelo de Ekman identifica seis emociones básicas: felicidad, tristeza, enojo, miedo, sorpresa y asco. Este modelo plantea que estas emociones tienen una base biológica y surgen de sistemas neurales independientes. Además, la combinación de estas emociones puede generar otras más complejas, como culpa, vergüenza u orgullo.

Por otro lado, Plutchik amplía el enfoque de Ekman al proponer ocho emociones primarias organizadas en pares opuestos: alegría vs. tristeza, confianza vs. asco, ira vs. miedo y sorpresa vs. anticipación. Plutchik sugiere que estas emociones pueden combinarse para formar estados emocionales más complejos y que su intensidad varía según las circunstancias.

Finalmente, el modelo OCC (Ortony, Clore y Collins) rechaza la idea de emociones básicas y, en su lugar, propone una clasificación más extensa con 22 categorías emocionales. Este modelo incluye emociones adicionales como alivio, envidia, decepción y admiración, ofreciendo así una representación más amplia y detallada de los estados emocionales humanos. Debido a su mayor número de clases, el modelo OCC proporciona un enfoque más flexible y detallado para la detección de emociones.

1.5 Modelos Dimensionales de Emociones (DiEMs)

A diferencia de los modelos discretos, los modelos dimensionales consideran que las emociones no son entidades separadas, sino que existen en un espacio continuo con relaciones entre ellas. Estos modelos buscan representar las emociones en dimensiones medibles.

Un ejemplo representativo es el Circunflejo de Russell, el cual organiza las emociones en dos dimensiones principales: valencia, que diferencia entre emociones agradables y desagradables, y activación/excitación, que distingue entre emociones de alta o baja energía. Este modelo enfatiza que las emociones están interconectadas y no pueden ser consideradas de manera completamente independiente.

Otro enfoque dimensional es la rueda de emociones de Plutchik, que representa las emociones en un gráfico circular donde las emociones primarias se encuentran en el centro, y su combinación en los extremos genera emociones secundarias. Además, este modelo ilustra cómo la intensidad de una emoción varía dependiendo de su posición en la rueda.

El último de estos modelos es el modelo tridimensional de Russell y Mehrabian, el cual amplía el modelo bidimensional de Russell al agregar la dimensión de dominancia, la cual mide el grado de control que una persona tiene sobre su estado emocional. En este modelo, la valencia indica qué tan placentera o desagradable es una emoción, la excitación mide su nivel de activación y la dominancia representa el grado de control que se tiene sobre ella.

Aplicaciones y Modelos Computacionales para el Reconocimiento de Emociones

Como se evidenció anteriormente, una subrama del análisis de sentimientos corresponde a la detección de emociones, la cual tiene aplicaciones en salud mental, interacción humano-computadora y análisis de opiniones. En los últimos años, el desarrollo de modelos preentrenados (PMs) ha permitido avances significativos en la comprensión del lenguaje humano, facilitando el reconocimiento de emociones en textos y grabaciones de voz. Inicialmente, la detección de emociones se basaba en diccionarios de palabras con carga emocional predefinida y técnicas de aprendizaje automático tradicionales. Sin embargo, estos enfoques presentaban limitaciones en la comprensión del contexto y la ambigüedad del lenguaje. La introducción de redes neuronales profundas, como RNN y LSTM, mejoró la capacidad de modelado del lenguaje, permitiendo capturar dependencias temporales en secuencias de texto. Posteriormente, con la aparición de modelos basados en transformers, como BERT y RoBERTa, se lograron avances notables en la representación semántica, aumentando la precisión en la clasificación de emociones [36].

Ngo y Kocoń [37] identificaron las limitaciones de los modelos tradicionales y propusieron un enfoque innovador que incorpora información contextual y personalización en la clasificación de emociones. En su investigación, analizaron el dataset CLARIN-Emo y desarrollaron un modelo de clasificación de

secuencia de oraciones (SSC), que supera las metodologías previas al considerar cada oración dentro de su estructura discursiva. Además, incorporaron la estrategia de UserID, permitiendo que el modelo captara características específicas de cada usuario, mejorando así la precisión del reconocimiento de emociones.

El modelo propuesto por Ngo y Kocoń [37] utilizó transformers preentrenados combinados con una arquitectura optimizada para la clasificación en tareas de análisis de sentimientos. Implementaron una variante de DistilRoBERTa para maximizar la eficiencia computacional sin comprometer la calidad del reconocimiento de emociones. Este enfoque demostró que integrar información de contexto y personalización puede reducir las imprecisiones derivadas de la subjetividad en la anotación de emociones y mejorar la interpretabilidad de los resultados.

En el contexto de myMind, estos avances resaltan la importancia de seleccionar modelos eficientes, mitigar sesgos y garantizar interpretaciones precisas para ofrecer un análisis emocional confiable y accesible.

2 Pruebas de modelos

2.1 Conjunto de Datos

En el ámbito del procesamiento de lenguaje natural (PLN), la clasificación de emociones y sentimientos es un área fundamental para el desarrollo de aplicaciones en el análisis de redes sociales, sistemas de recomendación y asistentes virtuales. Sin embargo, para entrenar modelos efectivos, es crucial contar con conjuntos de datos estandarizados que permitan evaluar de manera uniforme el desempeño de los algoritmos.

El propósito de este trabajo es la construcción y estandarización de un conjunto de datos con el objetivo de proporcionar una base confiable para la clasificación de emociones. Este dataset servirá como base para el entrenamiento y evaluación de modelos que serán implementados en la aplicación móvil, permitiendo un análisis más preciso del estado emocional de los usuarios y mejorando la capacidad de la herramienta para ofrecer información relevante sobre su bienestar.

2.2 Construcción del conjunto de datos

Se llevó a cabo un proceso de creación y adaptación de un conjunto de datos a partir de una muestra de audios obtenidos de estudiantes principalmente de la Pontificia Universidad Javeriana. A cada estudiante se le solicitó relatar una experiencia vivida, procurando expresar una emoción predominante en el audio. Este enfoque permitió una clasificación manual más precisa, asegurando que las etiquetas emocionales asignadas respondieran a un esquema coherente y alineado con un conjunto bien definido de emociones primarias.

Las categorías emocionales utilizadas corresponden al modelo de Plutchik, el cual proporciona una organización estructurada de las emociones humanas. Dichas categorías incluyen: anger (ira), anticipation (anticipación), joy (alegría), trust (confianza), fear (miedo), surprise (sorpresa), sadness (tristeza) y disgust (asco). La adopción de este modelo como marco de referencia en el estudio permitió una categorización consistente, facilitando la interpretación y comparación de las emociones identificadas en los audios.

Los audios recopilados tuvieron una duración variable entre 30 segundos y 1 minuto con 30 segundos, y fueron grabados en diversos entornos, tanto controlados como naturales, con el fin de representar de forma más realista distintas condiciones de uso. Todos los audios se almacenaron en

formato .wav, asegurando una calidad adecuada para su posterior procesamiento mediante modelos de reconocimiento de voz y análisis emocional.

Adicionalmente, se realizó un análisis de sentimientos, clasificando cada audio en una de tres categorías: negativo, neutral o positivo. Esta clasificación permitió establecer la polaridad predominante en cada muestra, complementando así la dimensión emocional previamente descrita. Cabe mencionar que dichas clasificaciones se hicieron de la mano con una psicóloga especializada en el tema.

A continuación, se presenta una muestra representativa del conjunto de datos construido en la tabla 2

Audio	Emoción	Sentimiento	Edad	Localidad	Personalidad	Universidad	Carrera	Sexo
A001	Disgust	Negative	22	Usaquén	Introvertido	Pontificia Universidad Javeriana	Ing. Sistemas	Masculino
A002	Joy	Neutral	22	Usaquén	Introvertido	Pontificia Universidad Javeriana	Ing. Sistemas	Masculino
A003	Anger	Negative	22	Usaquén	Introvertido	Pontificia Universidad Javeriana	Ing. Sistemas	Masculino
A004	Joy	Positive	21	Suba	Introvertido	Pontificia Universidad Javeriana	Ing. Mecatrónica	Masculino
A005	Joy	Positive	21	Suba	Introvertido	Pontificia Universidad Javeriana	Ing. Mecatrónica	Masculino

Cuadro 2: Estructura del conjunto de datos de emociones y sentimientos

En este caso, el conjunto de datos incluye el identificador único de cada uno de los 88 audios recopilados, así como las emociones y sentimientos asignados, junto con los datos demográficos de los estudiantes participantes. Es importante destacar que la participación fue completamente voluntaria y se formalizó mediante la firma de un consentimiento informado por parte de cada estudiante.

Las categorías asignadas a cada audio fueron validadas por la psicóloga Laura Gaitán, quien asesoró al grupo durante este proceso. Para ello, se seleccionó aleatoriamente una muestra representativa de 20 audios, sobre los cuales se realizó una verificación inicial de las etiquetas emocionales y de sentimiento. Esta muestra sirvió como guía de referencia para llevar a cabo el etiquetado manual del resto del conjunto de datos, asegurando coherencia y consistencia en la clasificación.

2.3 Comparación de Modelos

A partir del conjunto de datos definido y los audios correspondientes, se llevó a cabo una evaluación utilizando diversos modelos, alineados con los ejes principales que myMind plantea como solución: modelos de speech to text, modelos de reconocimiento de sentimientos y modelos de reconocimiento de emociones. A continuación, se presenta la metodología empleada para la evaluación de los modelos adecuados para el sistema myMind, así como los modelos seleccionados que cumplen con los requisitos establecidos.

Modelos de Reconocimiento de Voz

Para la etapa de reconocimiento de voz y transcripción automática, se realizó una comparación entre

distintos modelos de Automatic Speech Recognition (ASR). Una de las restricciones más importantes durante este proceso fue la necesidad de contar con modelos que ofrecieran soporte adecuado para el idioma español, dado que la aplicación myMind está orientada a usuarios cuya lengua materna es el español. Esta condición excluyó varios modelos de alto rendimiento que solo ofrecen soporte optimizado en inglés u otros idiomas.

La evaluación de los modelos se basó en la comparación entre las transcripciones generadas automáticamente por cada modelo y una transcripción manual previamente elaborada por el equipo. A partir de esta comparación, se calcularon las siguientes métricas clave, ampliamente utilizadas en la evaluación de sistemas ASR:

- **WER (Word Error Rate)**: mide la proporción de palabras incorrectamente transcritas en relación con la transcripción de referencia. Un valor menor de WER indica mayor precisión.
- **WA (Word Accuracy)**: calcula la proporción de palabras correctamente transcritas respecto al total, complementando la información proporcionada por el WER.
- **CER (Character Error Rate)**: evalúa los errores a nivel de caracteres, siendo especialmente útil en textos cortos o con variaciones mínimas entre versiones.
- **WPM (Words Per Minute)**: determina la velocidad de transcripción, dividiendo el número de palabras transcritas por la duración del audio en minutos.
- **Tiempo (Segundos)**: el tiempo de ejecución promedio por transcripción.

Estas métricas permiten analizar tanto la precisión como la eficiencia de los modelos evaluados, lo que facilita una selección objetiva del modelo que ofrezca el mejor desempeño en función de los requerimientos específicos de myMind en cuanto a reconocimiento automático de voz.

Adicionalmente, durante la fase de pruebas, los modelos fueron evaluados en un entorno controlado utilizando Google Colab. Este entorno ofrece por defecto una memoria RAM de 12.7 GB y un espacio de almacenamiento de 107.7 GB, lo cual representó una limitación importante al momento de ejecutar ciertos modelos. En particular, algunos de los modelos más complejos presentaron un consumo excesivo de memoria, impidiendo su ejecución en dicho entorno. Por este motivo, aquellos modelos que no pudieron ser ejecutados en las condiciones proporcionadas por Colab fueron descartados del proceso de selección, ya que esta restricción representa un criterio relevante para la elección del modelo final, teniendo en cuenta los recursos disponibles para la implementación del sistema.

A continuación se puede ver un resumen de los resultados obtenidos para los modelos de transcripción utilizados en la tabla 3:

Modelo	Autor	WER	WA	CER	WPM	Execution Time (s)
Whisper medium	OpenAI	0.2845	0.7155	0.1068	146.8859	90.2599
WhisperX medium	OpenAI + Faster-Whisper	0.2914	0.7086	0.1100	146.9938	78.1319
WhisperX large	OpenAI + Faster-Whisper	0.2932	0.7068	0.1040	148.0906	143.6904
Nemo	NVIDIA	0.3153	0.6847	0.1312	143.7664	12.3329
Whisper small	OpenAI	0.3173	0.6827	0.1239	147.3300	29.8797
WhisperX small	OpenAI + Faster-Whisper	0.3292	0.6708	0.1248	147.1370	29.9180
Wav2Vec medium	Facebook AI	0.4774	0.5226	0.1837	121.6006	45.0303
Whisper tiny	OpenAI	0.5145	0.4855	0.2350	151.3393	6.7896
Kaldi Vosk Complete	Kaldi Project	0.5216	0.4784	0.3872	100.5796	14.6752
SpeechBrain	SpeechBrain Team	0.5366	0.4634	0.2529	103.4301	44.2926
WhisperX tiny	OpenAI + Faster-Whisper	0.5406	0.4594	0.2540	156.7578	8.3030
Kaldi Vosk	Kaldi Project	0.5683	0.4317	0.4050	101.9412	6.5283

Cuadro 3: Comparación de modelos de reconocimiento de voz (ASR)

Los resultados muestran que los modelos con mejor desempeño en cuanto a calidad de transcripción lograron una **Word Accuracy (WA)** superior al 70 % y un **Character Error Rate (CER)** relativamente bajo, lo cual evidencia una alta fidelidad en la conversión del audio a texto. Sin embargo, dichos modelos presentaron tiempos de ejecución significativamente elevados, lo que representa una limitación importante al considerar su uso en el contexto de una aplicación móvil. En este tipo de sistemas, el rendimiento y la usabilidad son atributos clave de calidad, ya que los usuarios esperan respuestas ágiles y una experiencia fluida. En este sentido, el modelo **Nemo** se destaca por mantener un equilibrio óptimo entre precisión y eficiencia, ofreciendo una transcripción de calidad aceptable con un tiempo de ejecución considerablemente menor en comparación con otros modelos evaluados. Por esta razón, fue seleccionado como el modelo de reconocimiento de voz a implementar en el sistema myMind.

Modelos de Reconocimiento de Sentimientos

Dado que la calidad de las predicciones de emociones y sentimientos depende directamente de la transcripción de los audios, y considerando que el modelo seleccionado para la transcripción fue Nemo, se estableció que todas las evaluaciones de los modelos de sentimientos se realizarían utilizando como entrada las transcripciones generadas por dicho modelo.

Sin embargo, para garantizar una evaluación objetiva e independiente de posibles errores de transcripción, se optó por comparar las predicciones de los modelos afectivos con etiquetas manuales previamente definidas, es decir, etiquetas de emociones y sentimientos asignadas directamente por evaluadores miembros del equipo.

De esta forma, se construyó un conjunto de datos de prueba que incluye: la transcripción generada

por Nemo, la emoción etiquetada manualmente y el sentimiento asociado. A continuación, se presenta una muestra de dicho conjunto en la tabla 4:

Transcripción	Emoción	Sentimiento
hoy estaba caminando me había comprado mi café...	Anger	Negative
hoy dicen panadas con mi mamá estuvimos mucho ...	Joy	Neutral
hoy me peleó con mi hermano no quería hacer lo...	Sadness	Negative
pues hoy me lance temprano porque el año de mi...	Joy	Positive
en el mes de general en el mes de febrero me f...	Joy	Positive

Cuadro 4: Muestra del conjunto de prueba para modelos de emoción y sentimiento.

Para el caso del análisis de sentimientos, se utilizaron las categorías Negativo, Neutral y Positivo, definidas manualmente para cada uno de los audios para realizar la evaluación de los distintos modelos de análisis de sentimientos. La evaluación se realizó empleando métricas estándar ampliamente aceptadas en tareas de clasificación. Las métricas utilizadas fueron las siguientes:

- **Accuracy:** proporción total de predicciones correctas sobre el total de muestras evaluadas.
- **Precision:** medida de la calidad de las predicciones positivas para cada clase (cuántas predicciones positivas fueron realmente correctas).
- **Recall:** capacidad del modelo para identificar correctamente todas las instancias reales de cada clase.
- **F1-score:** media armónica entre la precisión y el recall, que proporciona un balance entre ambos indicadores.

Estas métricas se calcularon para cada una de las clases (Negativo, Neutral y Positivo) y se consideraron tanto sus valores individuales como sus promedios ponderados. Esta metodología permitió identificar con mayor claridad el desempeño de cada modelo frente a los distintos matices afectivos presentes en las transcripciones, facilitando así la selección del modelo más adecuado para su posterior integración en el sistema myMind.

Modelo	Autor	Accuracy	Precision	Recall	F1
PySentimiento	Juan Manuel Pérez et al.	0.6136	0.6012	0.6136	0.5808
VADER	C.J. Hutto	0.3864	0.3575	0.3864	0.3160
beto-sentiment-analysis	Juan Manuel Pérez et al. & José Cañete et al.	0.5341	0.4558	0.5341	0.4698
distilbert-base-multilingual-cased-sentiment	lxyuan (Google)	0.5114	0.3409	0.5114	0.3989
roberta-large-multilingual-sentiment	ClapAI	0.5909	0.5591	0.5909	0.5302
multilingual-sentiment-analysis	TabularisAI	0.3068	0.3324	0.3068	0.2539

Cuadro 5: Comparación de modelos de análisis de sentimiento evaluados con métricas estándar

A partir de los resultados obtenidos y presentados en la Tabla 5, se observa que el modelo de *PySentimiento* obtuvo los mejores valores en las métricas de evaluación general, particularmente en

Accuracy, Precision, Recall y F1-score. En consecuencia, este modelo fue tomado como candidato para ser implementado en el sistema *myMind*, dado su buen desempeño y adecuada cobertura del idioma objetivo.

Fine-tuning

Una vez obtenidos y evaluados los resultados de los modelos de análisis de sentimiento, se decidió realizar un ajuste adicional sobre aquellos que presentaron un mejor desempeño mediante un proceso conocido como fine-tuning. Este procedimiento consiste en reutilizar un modelo previamente entrenado y adaptar sus parámetros a un conjunto de datos más específico o especializado, en este caso, los audios transcritos del estudio.

El fine-tuning permite mejorar la capacidad del modelo para clasificar correctamente nuevas instancias, ya que refuerza el aprendizaje a partir de ejemplos más representativos del dominio de aplicación. A diferencia del entrenamiento desde cero, el fine-tuning parte de una base sólida de conocimientos generales ya aprendidos por el modelo, lo que reduce el tiempo de entrenamiento y mejora la precisión en contextos particulares. En este caso, se utilizó una porción del conjunto de datos anotado manualmente para ajustar los pesos del modelo, permitiéndole reconocer con mayor exactitud las emociones y sentimientos presentes en las transcripciones de los audios. Al realizar fine-tuning al modelo de "PySentimiento" "Bert-Multilingual" los resultados se pueden ver en la tabla 6

Modelo	Autor	Accuracy	Precision	Recall	F1
PySentimiento	Juan Manuel Pérez et al.	0.6136	0.6012	0.6136	0.5808
VADER	C.J. Hutto	0.3864	0.3575	0.3864	0.3160
beto-sentiment-analysis	Juan Manuel Pérez et al. & José Cañete et al.	0.5341	0.4558	0.5341	0.4698
distilbert-base-multilingual-cased-sentiment	lxuan (Google)	0.5114	0.3409	0.5114	0.3989
roberta-large-multilingual-sentiment	ClapAI	0.5909	0.5591	0.5909	0.5302
multilingual-sentiment-analysis	TabularisAI	0.3068	0.3324	0.3068	0.2539
RoBERTuito FineTuned	Juan Manuel Pérez et al. & Ajuste Propio	0.6197	0.6218	0.6197	0.5655
BERT Multilingual FineTuned	Google & Ajuste Propio	0.3889	0.1512	0.3889	0.2178

Cuadro 6: Comparación de resultados obtenidos entre modelos base y modelos *fine-tuned*. Las filas resaltadas corresponden a modelos ajustados de elaboración propia.

A partir de los resultados obtenidos tras aplicar el proceso de fine-tuning, se evidenció que el modelo ajustado a partir de RoBERTuito presentó una mejora leve en comparación con su versión base, alcanzando mayores valores en las métricas de accuracy, precision y recall, aunque con una ligera disminución en la métrica F1. Esta diferencia no fue significativa, lo que sugiere que el modelo original de **PySentimiento** ya se encontraba suficientemente optimizado para la tarea. Por otro lado, el modelo ajustado a partir de BERT Multilingüe arrojó un desempeño considerablemente inferior en todas las métricas, lo que indicó una pérdida de capacidad general de clasificación. Debido a estos resultados, se decidió mantener el modelo de PySentimiento como la alternativa preferida, ya que ofrece un buen equilibrio entre rendimiento y simplicidad sin necesidad de aumentar la complejidad del modelo base.

Modelos de Reconocimiento de Emociones

Para la tarea de análisis de emociones se empleó un conjunto de datos construido a partir de las transcripciones de los audios, etiquetado manualmente según las ocho emociones primarias propuestas en la rueda de Plutchik: Alegría, Tristeza, Miedo, Ira, Anticipación, Sorpresa, Disgusto y Confianza. Estas emociones, agrupadas en pares de opuestos (por ejemplo, Alegría-Tristeza o Miedo-Ira), permitieron establecer una estructura más clara en la clasificación y fueron especialmente útiles para estandarizar las salidas de modelos que no contemplaban explícitamente todas las emociones del conjunto. En particular, este principio de oposiciones se aplicó en modelos que requerían reconfigurar sus salidas a partir de las probabilidades de emociones opuestas, asegurando así coherencia con las categorías definidas en el corpus. Los resultados obtenidos por cada modelo se resumen en la tabla 7:

Modelo	Autor	Accuracy	Precision	Recall	F1
robertuito-emotion-analysis	Juan Manuel Pérez et al.	0.3523	0.6519	0.4429	0.4881
robertuito-emotion-analysis-personalizado	Juan Manuel Pérez et al. (ajustado con opuestos de Plutchik)	0.4545	0.3859	0.4545	0.3678
T5	Colin Raffel et al. & Manuel Romero	0.3182	0.1708	0.3182	0.1982
BETO	Juan Manuel Pérez et al. & del Arco et al.	0.1477	0.1204	0.1477	0.1321
XLM-EMO-T	Bianchi, Federico; Nozza, Debora; Hovy, Dirk	0.1364	0.0883	0.1364	0.1031

Cuadro 7: Resultados de evaluación de modelos de análisis de emociones sobre el conjunto etiquetado con emociones primarias de Plutchik

Los resultados obtenidos en la evaluación inicial de los modelos de análisis de emociones no fueron satisfactorios en términos generales, presentando en su mayoría valores bajos en las métricas de desempeño consideradas. Dada esta situación, se optó por explorar la posibilidad de ajustar uno de los modelos mediante una etapa de fine-tuning. Para ello, se seleccionó el modelo que mostró el mejor rendimiento relativo en la evaluación preliminar —**robertuito-emotion-analysis**— como base para el ajuste, al presentar un mayor equilibrio entre precisión, recuperación y puntaje F1 frente a los demás modelos evaluados.

Fine-tuning

Con base en los resultados obtenidos en la evaluación inicial, se procedió a realizar un proceso de fine-tuning sobre el modelo que mostró el mejor desempeño general: robertuito-emotion-analysis. Este ajuste se enfocó en adaptar el modelo a las características específicas del conjunto de datos utilizado, alineando el número y tipo de categorías emocionales con las definidas en la rueda de emociones de Plutchik. Esto permitió una correspondencia directa entre las predicciones del modelo y las clases presentes en los datos, lo cual es fundamental para asegurar una evaluación coherente y precisa del desempeño.

A continuación, se presentan los resultados obtenidos tras el ajuste en la tabla 8 :

Modelo	Autor	Accuracy	Precision	Recall	F1
robertuito-emotion-analysis	Juan Manuel Pérez et al.	0.3523	0.6519	0.4429	0.4881
robertuito-emotion-analysis-personalizado	Juan Manuel Pérez et al. (ajustado con opuestos de Plutchik)	0.4545	0.3859	0.4545	0.3678
T5	Colin Raffel et al. & Manuel Romero	0.3182	0.1708	0.3182	0.1982
BETO	Juan Manuel Pérez et al. & del Arco et al.	0.1477	0.1204	0.1477	0.1321
XLM-EMO-T	Bianchi, Federico; Nozza, Debora; Hovy, Dirk	0.1364	0.0883	0.1364	0.1031
RoBERTuito FineTuned	Juan Manuel Pérez et al. & Ajuste Propio	0.6111	0.5014	0.6111	0.5322

Cuadro 8: Resultados de evaluación de modelos de análisis de emociones sobre el conjunto etiquetado con emociones primarias de Plutchik. (La fila resaltada corresponde al modelo ajustado)

A pesar de que los resultados del modelo ajustado continúan siendo moderados en términos absolutos, **RoBERTuito FineTuned** se consolidó como el modelo con mejor rendimiento general en esta tarea específica. Su ventaja radica no solo en las métricas alcanzadas, sino en que todas sus predicciones corresponden directamente a las categorías definidas en el conjunto de datos, garantizando así una integración coherente en la arquitectura del sistema myMind. No obstante, al tener resultados tan limitados, se optó por realizar una capa adicional de validación.

Validación Adicional

En primer lugar, se realizó un proceso de transformación y adaptación de un dataset previamente existente, asegurando que las etiquetas emocionales siguieran las categorías previamente definidas correspondientes al modelo de Plutchik mencionado anteriormente.

El dataset original utilizado en este estudio proviene de la competencia "SemEval-2018 Task 1: Affect in Tweets"[38], la cual incluye tareas relacionadas con la clasificación de emociones y sentimientos en textos de redes sociales. Al estandarizar los datos para la clasificación de emociones en un contexto más amplio, fue necesario realizar una serie de transformaciones y adaptaciones, incluyendo limpieza y estructuración de los datos para garantizar la coherencia en los valores numéricos asociados a las emociones, descartando columnas irrelevantes, se corrigieron valores faltantes y se filtraron registros sin contenido útil, asegurando que el conjunto de datos se centrara exclusivamente en la información relevante para la clasificación emocional. De esta manera, se obtuvo un conjunto de datos con 1935 filas y con la siguiente estructura:

phrase	sentiment	emotion
Para tristeza y felicidad, Ed Sheeran al rescate	Negative	sadness
Lo lindo de sentir cerca a esa gente con la que compartiste tanto...	Positive	trust
mi enojo dura 5 minutos y después ya te extraño.	Negative	anger
En dos findes o tres voy a estar en corboda.	Positive	joy
Llevo 24 horas sintiéndome mal	Negative	sadness

Cuadro 9: Muestra del conjunto de prueba adicional para modelos de emoción.

De igual forma, se llevó a cabo el proceso de evaluación de cada uno de los modelos previamente

mencionados. Además, se reentrenó el modelo ajustado mediante fine-tuning, incorporando nuevos datos y aplicando ajustes destinados a balancear las clases, incluyendo el uso de oversampling para favorecer a las clases minoritarias. Como resultado, se obtuvieron las métricas de desempeño en la Tabla 10

Modelo	Autor	Accuracy	Precision	Recall	F1
RoBERTuito - Raw	Juan Manuel Pérez et al.	0.4103	0.5428	0.4103	0.4251
T5	Colin Raffel et al. & Manuel Romero	0.2093	0.2782	0.2093	0.1880
BETO	Juan Manuel Pérez et al. & José Cañete et al.	0.1457	0.1592	0.1457	0.1346
XLM-EMO-T	Bianchi, Federico; Nozza, Debora; Hovy, Dirk	0.1819	0.1862	0.1819	0.1796
Robertuito Finetuned (All data)	Juan Manuel Pérez et al. & Ajuste Propio	0.6917	0.6815	0.6917	0.6733

Cuadro 10: Resultados de evaluación de modelos de análisis de emociones sobre un conjunto multiclase en español y conjunto propio. La fila resaltada corresponde a un modelo ajustado de elaboración propia.

Como se puede observar, el modelo ajustado mediante fine-tuning presentó una mejora significativa en su rendimiento al utilizar ambos conjuntos de entrenamiento como referencia. Para la evaluación, se optó por considerar únicamente las clases base presentes en cada modelo; en caso de que una clase no estuviera incluida, se categorizaba como ‘other’. Esta estrategia permitió obtener resultados más específicos y contextualizados según la configuración de cada modelo, lo cual evidenció que el modelo con fine-tuning alcanzó métricas superiores en comparación con los modelos base. Teniendo en cuenta estos resultados, se decidió implementar dicho modelo como componente principal del sistema myMind.

3 Arquitectura

La arquitectura propuesta para el sistema *MyMind* se fundamenta en un enfoque basado en micro-servicios, con el objetivo de garantizar escalabilidad, mantenibilidad, independencia de componentes y facilidad de despliegue. Esta decisión arquitectónica permite una organización modular del sistema, donde cada componente se encuentra desacoplado y cumple una función específica dentro del flujo general de la aplicación.

Criterios de selección arquitectónica

La arquitectura de myMind fue diseñada considerando los siguientes atributos de calidad como criterios principales para su definición:

- **Funcionalidad:** Se priorizó que el sistema soporte funciones esenciales como el registro, transcripción y análisis de notas de voz. Para ello, se definió como criterio que cada componente tuviera responsabilidades claramente delimitadas, facilitando la implementación de funcionalidades específicas y evitando interferencias entre ellas.
- **Fiabilidad:** Se consideró fundamental que el sistema fuera capaz de seguir operando ante fallos parciales. Por esta razón, se estableció como criterio que los componentes pudieran funcionar

de forma aislada en caso de errores, promoviendo el diseño de servicios desacoplados y con tolerancia a fallos.

- **Disponibilidad:** La necesidad de garantizar un acceso continuo al sistema llevó a definir como criterio la posibilidad de replicar y distribuir los componentes más críticos. Esto permitió elegir una arquitectura que habilitara mecanismos de redundancia y balanceo de carga para reducir tiempos de inactividad.
- **Seguridad:** Dado que el sistema trata información sensible relacionada con la salud emocional de los usuarios, se definió como criterio la existencia de puntos de control claros para implementar medidas de autenticación, autorización y protección de datos, tanto en la comunicación como en el almacenamiento.
- **Mantenibilidad:** Para facilitar la evolución del sistema, se definió como principio que los componentes pudieran desarrollarse, probarse y actualizarse de forma independiente. Este criterio favorece una estructura modular que facilite la intervención localizada ante cambios o correcciones.
- **Escalabilidad:** Se estableció que el sistema debía ser capaz de adaptarse a diferentes niveles de carga sin comprometer su funcionamiento. Por tanto, se buscó una solución que permitiera escalar componentes de manera individual, en función del uso y demanda de cada uno.
- **Desempeño:** La necesidad de procesar tareas intensivas en cómputo, como el análisis de voz o sentimientos, llevó a definir como criterio la separación de dichas tareas en procesos optimizables, que pudieran ejecutarse de forma eficiente, sin afectar la experiencia general del usuario.

3.1 Selección del estilo arquitectónico

Tomando en consideración los atributos de calidad definidos previamente se optó por una arquitectura basada en servicios para el sistema *MyMind*. Este estilo arquitectónico ofrece un equilibrio pragmático entre la simplicidad de las arquitecturas monolíticas y la alta granularidad de los microservicios.

La arquitectura basada en servicios (service-based architecture) es un enfoque distribuido en el que cada servicio representa una unidad de software autónoma que encapsula una funcionalidad de negocio completa. Según [39], esta arquitectura es una variante híbrida del estilo de microservicios, caracterizada por su flexibilidad y menor complejidad operativa. A diferencia de los microservicios puros, los servicios aquí son más complejos y están organizados por dominios, lo que reduce la necesidad de coordinación entre servicios. Esto la convierte en una opción ideal para aplicaciones orientadas a dominios funcionales claramente diferenciados, como las que maneja *myMind*.

Además, este estilo permite lograr un nivel adecuado de modularidad sin comprometer la simplicidad del sistema. Su enfoque favorece una evolución progresiva hacia arquitecturas más distribuidas, si en un futuro se requiere mayor granularidad. De esta manera, se evita incurrir en los costos innecesarios asociados con arquitecturas excesivamente complejas, como las orientadas a eventos o a microservicios finamente distribuidos.

Descripción general de la arquitectura

En la Figura 1 se presenta una vista de alto nivel de la arquitectura seleccionada. Dicha arquitectura ha sido diseñada considerando principios de diseño orientados a servicios, asegurando una correcta

separación de responsabilidades y la posibilidad de escalar servicios de forma independiente, de acuerdo con la carga o demanda funcional.

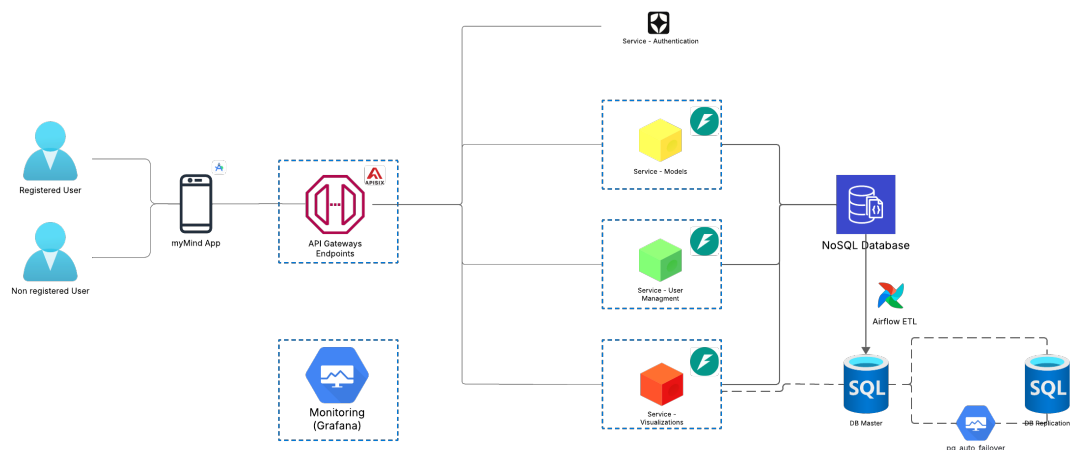


Figura 1: Arquitectura a alto nivel del sistema *myMind*

Los usuarios acceden a la solución mediante una aplicación móvil, que funciona como punto de interacción principal. Las solicitudes generadas desde la aplicación son dirigidas hacia un gateway de APIs, el cual actúa como punto de entrada unificado al sistema y se encarga de gestionar el enrutamiento del tráfico hacia los servicios internos, aplicar políticas de autenticación y autorización a través de la integración con Auth0, y controlar el acceso según el tipo de solicitud.

La arquitectura está segmentada en tres servicios principales, desplegados como contenedores independientes. El primero corresponde al servicio de usuarios, el cual administra los procesos de registro, inicio de sesión y gestión del perfil de los usuarios. El segundo es el servicio de modelos, responsable del procesamiento de las transcripciones de audio, análisis de emociones y sentimientos, operando de manera aislada para permitir una mejor gestión de recursos computacionales. El tercero es el servicio de reportes, que se encarga de consolidar los resultados generados por el servicio de modelos y ponerlos a disposición de los usuarios mediante informes personalizados.

Todos los servicios se comunican con una base de datos NoSQL, utilizada para almacenar la información operativa del sistema, incluyendo transcripciones, resultados de análisis y configuraciones. Esta base de datos está integrada con un sistema de extracción, transformación y carga (ETL), el cual transfiere los datos hacia una base de datos relacional en la nube. Esta última actúa como respaldo y como almacén de datos estructurado, facilitando la generación de reportes y el análisis histórico de la información. Además, se incorpora un mecanismo de replicación y conmutación por error automático que garantiza la disponibilidad y continuidad del servicio en caso de fallos.

Esta arquitectura modular, desacoplada y controlada permite no solo una evolución independiente de cada componente, sino también un monitoreo efectivo de su desempeño, integrando herramientas para el análisis de métricas en tiempo real.

4 Herramientas y tecnologías

Para la implementación de la solución propuesta, se propuso utilizar diversas herramientas y tecnologías que permiten asegurar el funcionamiento previsto del sistema:

- **Kotlin:** Lenguaje de programación utilizado para desarrollar la aplicación móvil de *myMind*, aprovechando su interoperabilidad con Java y su compatibilidad con Android Studio.
- **Python:** Lenguaje de programación principal para el backend, empleado en la construcción de los servicios REST, procesamiento de datos y modelos de análisis emocional mediante NLP.
- **Docker:** Para la contenerización y despliegue de cada uno de los servicios, asegurando entornos aislados y consistentes entre ambientes de desarrollo, prueba y producción.
- **Android Studio:** Utilizado en el desarrollo de la aplicación móvil de *myMind*, que actúa como interfaz principal de interacción para los usuarios finales.
- **FastAPI:** Framework principal del backend que permite definir servicios REST de forma eficiente y con validación automática de datos.
- **MongoDB:** Base de datos NoSQL utilizada como almacén principal de información operativa, especialmente útil para documentos como transcripciones y datos analíticos intermedios.
- **MySQL:** Base de datos relacional que actúa como respaldo en la nube y como almacén centralizado para análisis posteriores.
- **Apache APISIX:** Servidor de API Gateway que permite enrutar peticiones hacia los servicios internos y manejar autenticación, balanceo de carga y control de tráfico.
- **Apache Airflow:** Utilizado para la orquestación del flujo ETL entre las bases de datos, asegurando la ejecución programada y monitoreada de estos procesos.
- **Grafana y Prometheus:** Solución de observabilidad empleada para recolectar métricas del sistema, visualizar el comportamiento de los servicios y configurar alertas ante comportamientos anómalos.
- **Auth0:** Plataforma de autenticación delegada que gestiona los procesos de login, verificación de identidad y autorización de acceso de forma segura.
- **Failover y tolerancia a fallos:** Mecanismos integrados en la arquitectura para garantizar disponibilidad del sistema ante fallos parciales o caídas de componentes individuales.

VI DESARROLLO DE LA SOLUCIÓN

1 Desarrollo del Back-end

1.1 Módulo de autenticación de usuario

El sistema de autenticación y autorización del ecosistema *myMind* está gestionado a través del proveedor externo *Auth0*, lo cual permite separar las credenciales sensibles del backend propio y delegar la gestión de seguridad a una plataforma especializada.

Para habilitar la comunicación segura entre servicios, se configuró una aplicación tipo **Machine-to-Machine** en la plataforma *Auth0*, la cual permite generar tokens de acceso mediante el flujo `client_credentials`. Estos tokens son utilizados exclusivamente por los componentes internos del sistema (en este caso el Api Gateway) para autenticarse ante otros servicios protegidos por *Auth0*, permitiendo así un control estricto sobre el acceso entre módulos del ecosistema.

Adicionalmente, los usuarios que interactúan con el sistema a través de la interfaz obtienen su propio token de acceso. Dichos tokens son validados para autorizar el acceso a funcionalidades específicas, asegurando que cada solicitud esté autenticada y provenga de un usuario legítimo.

En ambos casos, los tokens son obtenidos desde un endpoint de *Auth0* que sólo puede ser accedido por los actores autorizados del sistema (ya sean aplicaciones cliente o servicios internos), quienes poseen las credenciales necesarias (como **Client ID** y **Client Secret**) para solicitar dichos tokens.

Al manejar por medio de esta plataforma las credenciales de los usuarios, se establece que:

- Las contraseñas nunca son almacenadas directamente en la base de datos del sistema *myMind*.
- *Auth0* se encarga del cifrado y almacenamiento seguro de las credenciales.

De igual manera, esto responde directamente con restricciones dadas por los atributos de calidad como:

- **Seguridad:** Se delega el manejo de credenciales y autenticación a una plataforma robusta y auditada.
- **Escalabilidad:** Permite integrar múltiples flujos de autenticación, incluyendo redes sociales, multifactor y autenticación empresarial.
- **Simplicidad:** Reduce la complejidad del backend, evitando la necesidad de implementar lógica personalizada para gestión de sesiones o almacenamiento seguro de contraseñas.
- **Experiencia del usuario:** El usuario no debe registrar una cuenta ajena a las que ya tiene establecidas. Puede usar sus cuentas de Google, microsoft, entre otras.

De esta manera, el tráfico entre servicios y las interacciones con los usuarios están protegidos mediante tokens válidos y firmados, validados sin necesidad de compartir información sensible.

1.2 Módulo de enrutamiento API Gateway

Para centralizar y controlar el flujo de tráfico entre los distintos servicios del ecosistema *myMind*, se implementó un módulo de enrutamiento basado en **Apache APISIX**, el cual actúa como *API Gateway*. Este componente permite definir, modificar y proteger rutas HTTP de forma flexible y extensible, proporcionando un punto de entrada único para los consumidores del sistema.

El entorno se compone de los siguientes elementos principales:

- **Apache APISIX:** Es el componente principal del gateway, encargado de recibir las solicitudes externas e internas y enrutar el tráfico hacia los servicios correspondientes, aplicando filtros, autenticación, validaciones o transformaciones si es necesario. **De este modo, se integra con el módulo de autenticación descrito anteriormente, utilizando plugins como openid-connect para proteger rutas sensibles.**
- **etcd:** Es la base de datos distribuida utilizada por APISIX para almacenar toda la configuración del gateway (rutas, plugins, upstreams, certificados, etc.), garantizando consistencia y disponibilidad del estado del sistema.
- **APISIX Dashboard:** Es una interfaz gráfica de administración conectada directamente a **etcd**, que permite visualizar, gestionar y editar las rutas del gateway de forma accesible para los desarrolladores. Esta herramienta facilita el monitoreo del comportamiento del gateway y la gestión dinámica de nuevas rutas, sin necesidad de modificar archivos de configuración o reiniciar servicios.

De este modo, es posible incorporar nuevas rutas, modificar las existentes o aplicar políticas de autenticación y control de acceso (como el plugin **openid-connect** de Auth0) de forma flexible, sin necesidad de redeplegar todo el sistema. Además, la separación entre el archivo de configuración base y la carga dinámica permite escalar y mantener el gateway fácilmente.

1.3 Servicio de usuarios

El servicio de usuarios tiene como objetivo principal manejar todo lo relacionado con la información personal del usuario, de manera que esta pueda obtenerse, editarse o eliminarse desde la base de datos de MongoDB. El servicio fue desarrollado utilizando *FastAPI*, aprovechando su rendimiento, asincronía nativa y facilidad de integración. Este servicio permite gestionar perfiles de usuario autenticados a través de *Auth0* e incluye múltiples endpoints para registrar, consultar y actualizar información como nombre, correo electrónico, preferencias de notificación y foto de perfil.

La conexión con la base de datos *MongoDB* se realiza de forma asíncrona mediante el cliente **motor**, lo que permite operaciones rápidas y no bloqueantes. La base de datos utilizada se llama **mymind_users**, y cada documento representa un usuario autenticado, identificado por el campo `_id`, que corresponde al identificador `sub` generado por Auth0. Esta integración directa entre autenticación y persistencia garantiza la unicidad y coherencia de los datos del usuario.

Control de acceso

Una de las prioridades en el diseño fue garantizar que las solicitudes al backend provengan exclusivamente del gateway *APISIX*, tanto en entornos locales como en producción. Para ello, se implementó un middleware personalizado (**verify_request_from_apisix**) que verifica la dirección IP del cliente. En el entorno local, esta verificación se realiza comparando contra la IP del contenedor **apisix**, resuelta mediante DNS dentro de la red definida para los contenedores.

Este middleware permite controlar estrictamente el acceso al backend, asegurando que únicamente las solicitudes autorizadas puedan llegar a los endpoints, lo cual refuerza la seguridad y el aislamiento del servicio.

Endpoints implementados

El servicio expone múltiples funcionalidades clave:

- **Registro de usuario (/register):** Crea un documento con los datos básicos del usuario y una lista vacía de transcripciones.
- **Consulta de perfil:** Permite obtener información general del usuario mediante `/profile`, así como información específica como nombre (`/name`), correo electrónico (`/email`), foto de perfil (`/profile-pic`) y preferencias de notificación (`/notifications`), entre otras funcionalidades.
- **Actualización selectiva:** A través de las rutas como `/update-name`, `/update-email`, `/update-notifications` y `/update-profile-pic`, el usuario puede modificar atributos individuales de su perfil.
- **Manejo de transcripciones:** Permite la obtención, actualización y eliminación de transcripciones generadas por el usuario.

Todas las rutas están protegidas mediante autenticación, utilizando dependencias que validan el token JWT proporcionado por Auth0, asegurando así que solo usuarios autenticados puedan interactuar con el sistema.

1.4 Servicio de modelos

El servicio de modelos de *myMind* está encargado de procesar las transcripciones (tanto de audio como de texto) y realizar sobre estas el análisis de sentimientos y emociones. Este servicio está construido con *FastAPI* y se conecta a una base de datos MongoDB.

Control de acceso

El servicio verifica el origen de las solicitudes para permitir únicamente aquellas provenientes de **APISIX**.

Además, todas las rutas están protegidas mediante autenticación usando *Auth0*, lo que garantiza que sólo usuarios autenticados puedan acceder a las funcionalidades del servicio, igual que el servicio de usuarios.

Modelos Utilizados

Los modelos utilizados dentro del sistema son los descritos en la sección de Análisis de Modelos, los cuales corresponden a:

- **Reconocimiento automático de voz (ASR):** Se utiliza el modelo *NVIDIA NeMo* para convertir el audio en texto.
- **Análisis de sentimiento:** Se emplea la librería *Pysentimiento*.
- **Análisis de emociones:** Se utiliza un modelo *ajustado manualmente (fine-tuned)* sobre la base de RoBERTuito para identificar emociones ajustado con etiquetas del modelo de Plutchik para gestionar emociones.

Procesamiento de peticiones

Dado que la transcripción de audio y el posterior análisis con modelos de lenguaje puede demorar algunos segundos, se adoptó una arquitectura basada en *procesamiento asíncrono con colas* usando *Celery* y *Redis* como broker y backend.

Esto permite que las solicitudes de los usuarios no bloqueen el servicio mientras se procesa una transcripción. En lugar de esperar una respuesta inmediata, cada solicitud es encolada como una tarea que será procesada por un *worker* en segundo plano. Posteriormente, el usuario puede consultar el estado y resultado de su tarea mediante un canal específico gestionado con Redis (pub/sub).

Flujo de procesamiento de peticiones

El flujo completo del servicio sigue los siguientes pasos:

1. El usuario envía una solicitud a la ruta `/transcription/add-transcription/audio`.
2. La solicitud es validada y se encola como tarea de Celery en la cola `transcriptions`.
3. Un *worker* de Celery toma la tarea y realiza:
 - La transcripción del audio.
 - El análisis de sentimiento usando `pysentimiento`.
 - El análisis de emociones con el modelo ajustado.
4. Al terminar el procesamiento, se publica el resultado en un canal Redis único asociado al `task_id` correspondiente.
5. El cliente puede consultar el resultado en tiempo real mediante la ruta `/transcription/stream/{task_id}`, la cual escucha el canal Redis en espera de la respuesta.

1.5 Servicio de visualizaciones

En el marco del proyecto, se desarrolló el servicio de visualización de transcripciones de audio, con el objetivo de facilitar a los usuarios el análisis de sus emociones y sentimientos detectados en el procesamiento automático del habla. Este servicio se encarga de traer todos los datos relevantes del usuario autenticado desde la base de datos mongo para posteriormente ser traídos en formato json hasta el front, permitiendo identificar tendencias emocionales y consultar registros históricos recientes por medio de gráficas.

- **Promedio de transcripciones del usuario autenticado en los últimos 7 días**
 - **Ruta:** GET `/transcriptions/user/ultimos-7-dias`
 - **Respuesta:** Objeto que incluye el promedio de probabilidades de emociones y sentimientos de las transcripciones realizadas por el usuario en los últimos 7 días.
 - **Uso:** Permite obtener un análisis promedio de las emociones y sentimientos de las transcripciones realizadas por el usuario en el último período de 7 días.
- **Todas las transcripciones del usuario autenticado**

- **Ruta:** GET /transcriptions/user
- **Respuesta:** Lista de objetos con el ID de la transcripción, fecha y hora de todas las transcripciones realizadas por el usuario autenticado.
- **Uso:** Permite obtener un listado completo de las transcripciones del usuario en orden descendente según la fecha.
- **Última transcripción registrada de ese usuario**
 - **Ruta:** GET /transcriptions/user/latest
 - **Respuesta:** Objeto que incluye el ID, fecha, hora, emoción, sentimiento, y probabilidades asociadas a la última transcripción registrada por el usuario.
 - **Uso:** Facilita la consulta del análisis emocional más reciente del usuario.
- **Top de emociones y sentimientos de los últimos 7 días del usuario**
 - **Ruta:** GET /transcriptions/user/ultimos-7-dias/top-emotions-sentiments
 - **Respuesta:** Top 3 emociones y el sentimiento más común en las transcripciones realizadas por el usuario en los últimos 7 días.
 - **Uso:** Permite obtener un resumen de las 3 emociones más comunes y el sentimiento predominante en las transcripciones de los últimos 7 días.
- **Transcripción específica del usuario por audio ID**
 - **Ruta:** GET /transcriptions/user/audio/{audio_id}
 - **Respuesta:** Objeto que incluye el ID, fecha, hora, emoción, sentimiento y probabilidades asociadas a la transcripción especificada por ID del usuario autenticado.
 - **Uso:** Permite consultar el análisis emocional detallado de una transcripción específica del usuario autenticado.

Control de acceso

Todos los endpoints mencionados requieren autenticación mediante un token JWT, el cual se procesa utilizando la función `get_current_user`. Esta función extrae de manera segura el identificador del usuario directamente del token, evitando que se envíe el `user_id` manualmente como parámetro. De esta manera, se garantiza que las consultas a la base de datos se realicen exclusivamente sobre las transcripciones pertenecientes al usuario autenticado.

Consultas a la base de datos

Las consultas fueron diseñadas utilizando el ORM de SQLAlchemy. Se aplicaron filtros basados en el identificador del usuario autenticado (obtenido desde el token) y, cuando fue necesario, por rangos de fechas, asegurando que los resultados correspondieran únicamente al período o criterios solicitados.

Control de errores

Se implementaron mecanismos para manejar los siguientes escenarios:

- Retorno de un error HTTP 404 en caso de que el usuario no tenga transcripciones registradas.
- Manejo de errores de autenticación en caso de no recibir un token válido o en caso de que el token haya expirado.

2 Aplicación Móvil

2.1 Logo de la aplicación

El logo de la aplicación representa visualmente el propósito emocional y cognitivo de *myMind*. Se diseñó como un personaje en color morado que abraza un cerebro rosado, integrando la dimensión afectiva (el abrazo) y el nombre de *myMind* que se traduce al español como *miMente* representado por el cerebro. Esta imagen busca generar empatía, calidez y cercanía, simbolizando el acompañamiento emocional que ofrece la aplicación.

2.2 Psicología del color

Para la definición de la paleta de colores de la aplicación, se tomó como base la psicología del color, la cual sostiene que cada tonalidad provoca reacciones espontáneas en las personas y posee un significado simbólico claro y definido [40]. En este sentido, se optó por utilizar tonos pastel de azul y morado, buscando generar una experiencia visual armónica y emocionalmente positiva.

El logo de la aplicación refuerza esta elección cromática: un personaje morado abraza un cerebro rosado, lo que convierte al morado en un vínculo visual entre el azul de la interfaz y el rosado del cerebro. Según *Psicología del color*, el azul es considerado el color más profundo e inmaterial, asociado tradicionalmente con la calma y la tranquilidad [40]. La Escuela Británica de Artes Creativas y Tecnología indica que el azul claro evoca sensaciones de salud, comprensión y alivio [41]. Por su parte, el Colegio de Psicólogos destaca que el morado simboliza espiritualidad, equilibrio y serenidad [42].

2.3 Mockups iniciales

Durante las primeras fases del diseño, se crearon cuatro propuestas visuales en Figma, incluyendo estructura de navegación, paleta de colores y estilo gráfico. Estas propuestas fueron compartidas entre los integrantes del equipo para realizar una votación interna, priorizando criterios como usabilidad, estética y coherencia emocional. El diseño más votado fue ajustado según sugerencias y se convirtió en la base del prototipo funcional.

Posteriormente, se desarrolló un prototipo interactivo en Figma, que permitía simular la experiencia de uso de la aplicación antes de iniciar el desarrollo. Este prototipo incluía la navegación entre pantallas, transiciones animadas y simulaciones de interacción con botones y formularios, lo cual facilitó la validación temprana de la experiencia de usuario (UX) y la toma de decisiones de diseño.

2.4 Elección de Android Studio

Desde el inicio del proyecto se definió que la aplicación móvil se desarrollaría en Android, principalmente por razones de familiaridad tecnológica, disponibilidad de dispositivos de prueba, y la necesidad

de concentrar esfuerzos en el modelo de procesamiento del lenguaje natural. Se utilizó Android Studio como entorno de desarrollo, aprovechando sus herramientas de diseño visual, gestión de dependencias y compatibilidad con Kotlin.

2.5 Proceso de desarrollo

El desarrollo de la aplicación se organizó siguiendo principios de arquitectura modular y separación de responsabilidades. Se adoptó el patrón de diseño **MVVM** (Modelo-Vista-ViewModel), lo que permitió una mayor mantenibilidad, escalabilidad y separación lógica entre los componentes visuales, los modelos de datos y la lógica de negocio.

Las vistas se implementaron mediante actividades y fragmentos, utilizando *ViewBinding* para acceder a los elementos de la interfaz de forma segura y eficiente. Se definieron adaptadores personalizados para la visualización de listas, y se mantuvo una estandarización en los nombres de clases, métodos y estilos visuales, garantizando coherencia y legibilidad en todo el proyecto.

Cada funcionalidad de red fue encapsulada en un servicio con su respectiva interfaz e implementación, siguiendo buenas prácticas de programación orientada a interfaces. Además, se aplicaron algunos patrones de diseño para fortalecer la estructura del código; por ejemplo, el patrón **Singleton** fue utilizado para centralizar el acceso al cliente Retrofit, asegurando que todas las solicitudes compartieran la misma instancia de configuración.

La comunicación con el backend se realiza mediante solicitudes HTTP utilizando Retrofit, incluyendo autenticación y autorización mediante la plataforma Auth0. También se implementó el manejo de permisos sensibles como el de grabación de audio, almacenamiento y acceso a internet, con mensajes adecuados de interacción para el usuario.

El backend está estructurado en tres servicios independientes, lo cual permitió organizar la lógica de red de manera clara:

- **Servicio de usuarios:** gestiona el registro, inicio de sesión, perfil y preferencias del usuario.
- **Servicio de audios:** permite subir grabaciones, consultar transcripciones y gestionar el historial.
- **Servicio de visualizaciones:** recupera los análisis de emociones y sentimientos y genera los datos necesarios para los gráficos.

El proceso de desarrollo fue iterativo y estuvo guiado por la metodología ágil **Scrum**. A través del tablero Kanban se documentaron y organizaron las tareas semanales, permitiendo entregas incrementales que facilitaban la revisión continua del producto. Cada semana se presentaban avances funcionales que eran evaluados, ajustados y corregidos en función del feedback recibido por parte del equipo, garantizando así una evolución constante del prototipo hacia su versión final. Este enfoque también permitió detectar errores tempranos y realizar mejoras continuas en la experiencia de usuario y en la estabilidad técnica de la aplicación.

2.6 Resultado final

A continuación, se presentan las pantallas finales de la aplicación, resultado del proceso iterativo de diseño, validación y desarrollo. Cada una cumple una función específica dentro del flujo del usuario y está diseñada para ofrecer una experiencia intuitiva, segura y emocionalmente significativa. Las figuras al final de esta sección muestran una selección de las pantallas clave.

- **MainActivity**: pantalla contenedora inicial que gestiona la navegación hacia otras vistas.
- **WelcomeActivity**: pantalla inicial que da la bienvenida y dirige al usuario hacia el inicio de sesión o registro.
- **RegisterActivity**: formulario de registro donde se capturan los datos básicos del usuario y se valida su contraseña.
- **SignInActivity**: permite iniciar sesión de forma segura mediante integración con Auth0.
- **PrivacyAgreeActivity**: muestra los términos y condiciones de privacidad, solicitando aceptación del usuario antes de continuar.
- **HomeActivity**: actividad principal que integra la barra de navegación inferior y coordina los distintos fragmentos.
- **InstructionsFragment**: guía al usuario nuevo en el uso de la aplicación, explicando paso a paso sus funcionalidades.
- **RecordingFragment**: interfaz de grabación de audio con botones para pausar, reanudar, eliminar o confirmar el audio.
- **ConfirmationFragment**: muestra la grabación realizada, permitiendo escucharla antes de enviarla al backend.
- **LoadingFragment**: indica que el audio está siendo procesado por el backend para generar la transcripción y análisis emocional.
- **ReportFragment**: presenta los resultados del análisis emocional a través del tiempo en gráficos interactivos y explicativos.
- **SpecificReportFragment**: muestra un reporte emocional asociado a un audio específico seleccionado desde el historial o el último enviado.
- **AudioHistoryFragment**: lista todos los audios grabados por el usuario, con opciones para visualizar o eliminar cada uno.
- **ProfileFragment**: pantalla donde se visualiza el perfil del usuario, incluyendo nombre e imagen.
- **EditProfileFragment**: permite editar los datos del perfil, actualizar la imagen, el nombre de usuario y cambiar la contraseña.
- **SettingsFragment**: pantalla de configuración donde se puede personalizar el horario de notificaciones.
- **PrivacyFragment**: presenta las políticas de privacidad de la aplicación, accesibles en cualquier momento desde el menú.
- **AboutFragment**: ofrece información institucional sobre la aplicación, su propósito y sus creadores.
- **SuccessActivity**: pantalla de confirmación tras un proceso exitoso, como la finalización del registro o una actualización exitosa.

El flujo general de navegación puede verse en la Figura 2.

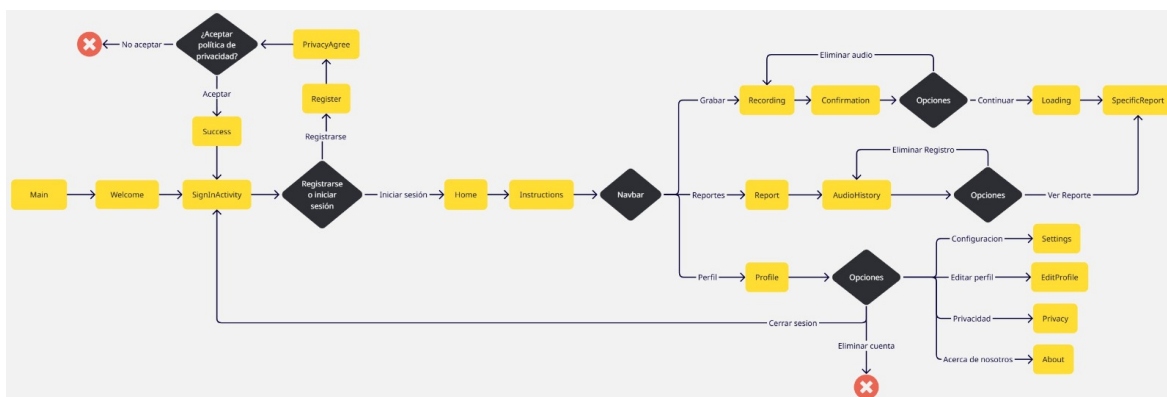
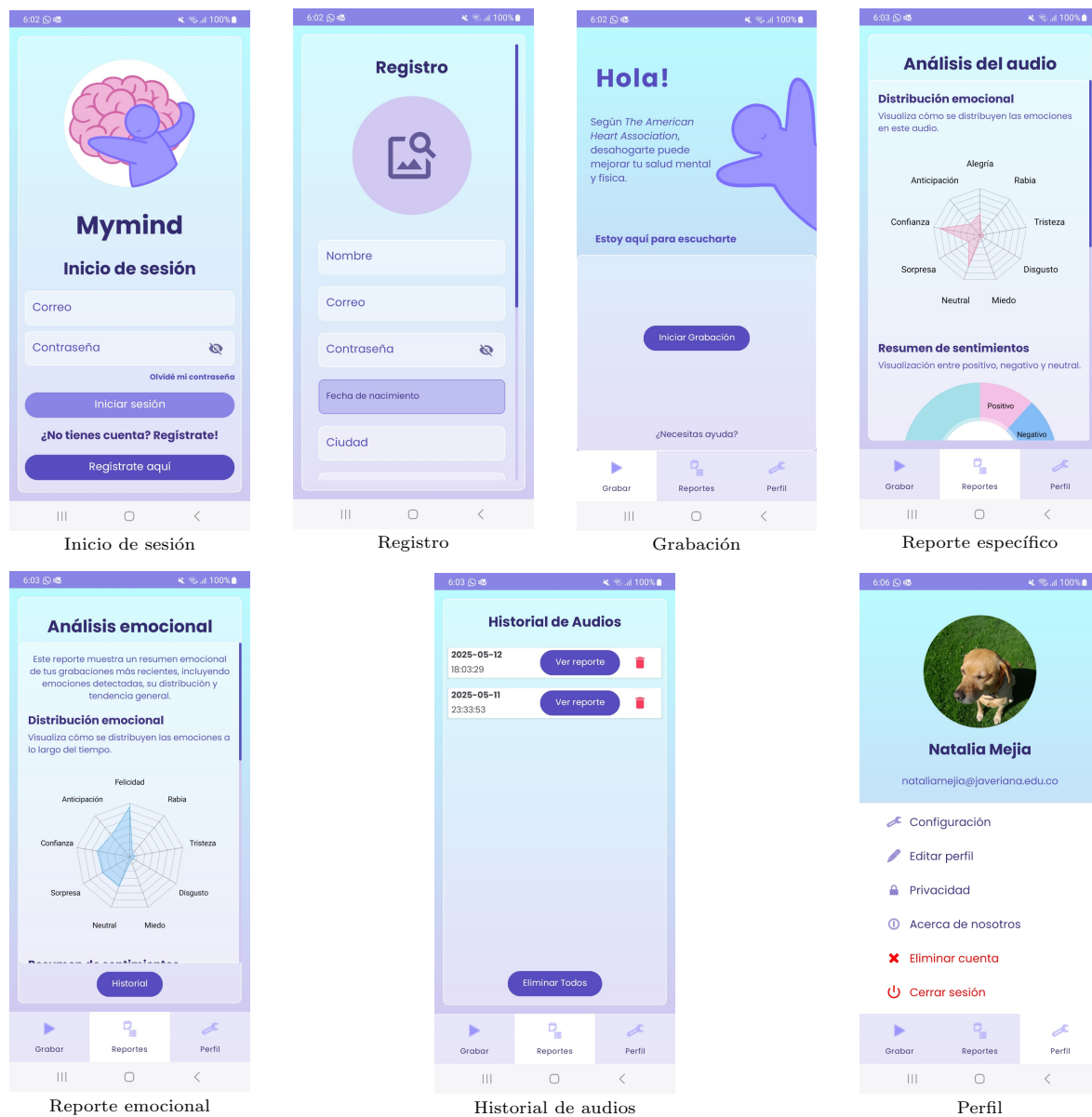


Figura 2: Diagrama de flujo de la aplicación

La aplicación móvil myMind permite a los usuarios registrar y analizar su estado emocional a través de notas de voz. El proceso inicia con el registro o inicio de sesión, seguido de una guía de instrucciones. Desde la pantalla principal, el usuario puede grabar su voz y enviarla al backend, donde es transcrita y analizada emocionalmente mediante técnicas de procesamiento de lenguaje natural. Los resultados se presentan en gráficos interactivos que muestran tanto el análisis del último audio como un historial emocional a lo largo del tiempo. Además, el usuario puede consultar reportes específicos de cada grabación, gestionar su perfil, acceder al historial de audios, y configurar preferencias como notificaciones y privacidad. Toda la experiencia está diseñada para ser sencilla, intuitiva y emocionalmente empática.

Este prototipo es completamente funcional, implementa todas las funcionalidades descritas en el planteamiento del proyecto. Permite grabar notas de voz, enviarlas al servidor para su transcripción y análisis emocional, y visualizar los resultados en reportes personalizados. La aplicación cumple con los objetivos establecidos: recolecta datos mediante audio, realiza transcripciones automáticas, aplica modelos de procesamiento de lenguaje natural, y presenta al usuario un seguimiento gráfico de sus emociones. Todo esto se lleva a cabo en un entorno seguro, privado y amigable para el usuario, lo que permite ofrecer una experiencia tecnológica coherente con la meta de promover el bienestar emocional de los estudiantes.

A continuación, se presentan capturas de las pantallas principales:

Figura 3: Capturas principales de la aplicación móvil *myMind*

3 Despliegue

3.1 Implementación del ETL (Extract, Transform, Load)

Para la integración de datos entre la base de datos operativa MongoDB y el almacén de datos analítico MySQL, se desarrolló un proceso ETL (Extracción, Transformación y Carga) automatizado

utilizando Apache Airflow. Este proceso es fundamental para permitir el análisis posterior de los datos recopilados por la aplicación MyMind.

Arquitectura del Proceso ETL El proceso ETL implementado sigue una arquitectura de tres componentes principales:

- **Apache Airflow:** Orquestador de flujos de trabajo que permite programar y monitorear la ejecución del proceso ETL.
- **Docker/Docker Compose:** Provee la contenedorización del entorno, asegurando la portabilidad y la reproducibilidad del proceso.
- **Python:** Lenguaje de programación utilizado para desarrollar las tareas de extracción, transformación y carga.

La arquitectura general del proceso se puede visualizar como un flujo desde MongoDB hacia MySQL, pasando por las tareas gestionadas por Airflow dentro de contenedores Docker.

Estructura de Datos El proceso ETL maneja dos estructuras de datos distintas que requieren transformación:

- **MongoDB (Origen):** Contiene una colección de usuarios con documentos anidados, que incluyen datos de perfil y transcripciones como subdocumentos.
- **MySQL (Destino):** Implementa un modelo relacional con dos tablas principales:
 - **users:** Tabla normalizada que contiene la información de perfil de los usuarios.
 - **transcriptions:** Tabla que almacena las transcripciones, relacionada con la tabla de usuarios mediante claves foráneas.

Desarrollo del DAG de Airflow El flujo de trabajo se implementó como un DAG (Directed Acyclic Graph) en Airflow, compuesto por tres tareas principales:

1. **extract_mongo_data:** Extrae los documentos de MongoDB y maneja la conversión de tipos BSON.
2. **transform_data:** Procesa los documentos separando información de usuarios y transcripciones, convierte formatos de fecha, y estructura los datos para inserción.
3. **sync_and_load_mysql:** Sincroniza y carga los datos en MySQL, implementando actualizaciones e inserciones, además de eliminar registros que ya no existen en la fuente.

3.2 Implementación del Sistema de Failover para MySQL

Para garantizar la alta disponibilidad del almacén de datos analítico, se implementó un sistema de replicación y failover para MySQL. Este sistema asegura la continuidad operativa en caso de fallas en el servidor principal de base de datos.

Arquitectura de Replicación La arquitectura de replicación MySQL implementada consta de dos componentes principales:

- **MySQL Master:** Servidor principal que gestiona todas las operaciones de escritura y lectura durante la operación normal.
- **MySQL Slave:** Servidor de réplica que mantiene una copia sincronizada de los datos, preparado para asumir el rol de master en caso de fallo.

Ambos servidores se implementaron como contenedores Docker dentro del mismo entorno que aloja Airflow, facilitando su gestión y monitoreo integrado.

Configuración de la Replicación La replicación MySQL se configuró utilizando el método de replicación basada en binlogs, considerado el más robusto y eficiente:

1. **Configuración del Master:** El servidor master se configuró habilitando el registro binario con las siguientes directivas en su archivo de configuración:

```
log-bin=mysql-bin
server-id=1
binlog_format=ROW
binlog_expire_logs_seconds=2592000
sync_binlog=1
innodb_flush_log_at_trx_commit=1
```

2. **Configuración del Slave:** El servidor slave se configuró con los parámetros necesarios para recibir y aplicar los cambios del master:

```
server-id=2
relay-log=relay-bin
log_slave_updates=ON
read_only=ON
```

3. **Usuario de Replicación:** Se creó un usuario dedicado con privilegios de replicación para establecer la conexión segura entre master y slave:

```
CREATE USER 'replication_user'@'%' IDENTIFIED WITH
mysql_native_password BY 'replication_password';
GRANT REPLICATION SLAVE ON *.* TO 'replication_user'@'%';
```

Sistema de Detección y Failover Se implementó un sistema de monitoreo y failover automático que realiza las siguientes funciones:

- **Monitoreo Continuo:** Un script de monitoreo verifica cada 10 segundos la disponibilidad del servidor master ejecutando comandos de ping MySQL.

- **Detección de Fallos:** Cuando se detecta que el master no responde, se inicia automáticamente el proceso de failover.
- **Promoción del Slave:** El script de failover ejecuta los comandos necesarios para promocionar el slave a master:

```
STOP SLAVE;  
RESET SLAVE ALL;  
SET GLOBAL read_only=OFF;
```

- **Redirección de Conexiones:** La conexión configurada en Airflow se actualiza automáticamente para apuntar al nuevo servidor master, asegurando la continuidad del ETL.

Recuperación y Restauración Para facilitar la recuperación después de un failover, se implementó un script de restauración que permite:

1. Reiniciar el servidor que falló.
2. Configurarlos como slave del nuevo master.
3. Sincronizar los datos utilizando la posición actual de los binlogs.
4. Restablecer la replicación en la dirección inversa.

Beneficios del Sistema de Failover La implementación del sistema de failover proporciona múltiples beneficios para la infraestructura del proyecto MyMind:

- **Alta Disponibilidad:** Garantiza que el almacén de datos analítico esté siempre disponible incluso durante fallos de hardware o software.
- **Continuidad Operativa:** El proceso ETL sigue funcionando sin interrupciones significativas gracias a la redirección automática de conexiones.
- **Integridad de Datos:** La replicación sincrónica asegura que no se pierdan datos durante el proceso de failover.
- **Mantenimiento Sin Tiempo de Inactividad:** Permite realizar tareas de mantenimiento en un servidor mientras el otro continúa dando servicio.

Esta arquitectura de replicación y failover complementa el proceso ETL, asegurando que los datos analíticos estén siempre disponibles para la aplicación MyMind y sus funcionalidades de análisis emocional y generación de reportes.

VII RESULTADOS

En esta sección se presentan los resultados del proceso de pruebas realizado sobre la solución desarrollada, abarcando tanto la aplicación móvil como los servicios backend. El objetivo principal fue validar el correcto funcionamiento, seguridad, estabilidad, rendimiento y usabilidad de la plataforma *myMind*, así como su conformidad con los requerimientos funcionales y no funcionales establecidos en las etapas de análisis y diseño.

Estrategia y tipos de prueba

La estrategia de aseguramiento de calidad se estructuró en distintos niveles para verificar tanto los componentes individuales como la integración del sistema bajo condiciones reales de uso. Las pruebas se agruparon en cuatro categorías principales:

- **Pruebas unitarias:** Se validó el comportamiento de módulos individuales en el cliente (desarrollado en Kotlin para Android) y en el backend (desarrollado en Python sobre Ubuntu 24.04 LTS). Se utilizaron `pytest` y `pytest-cov` para los servicios backend, y **Android Studio + Logcat** para el cliente. Se alcanzó una cobertura de código superior al 85 % en los componentes críticos. Todos los casos fueron exitosos tras iteraciones de corrección.
- **Pruebas de integración:** Evaluaron flujos funcionales completos desde el registro del usuario hasta la generación de reportes emocionales. Se validó la interacción entre cliente, backend y API Gateway (**APISIX**), incluyendo autenticación con Auth0, procesamiento asíncrono, y almacenamiento en MongoDB y MySQL. Todos los flujos definidos fueron ejecutados con éxito dentro de los tiempos de respuesta requeridos.
- **Pruebas de carga:** Utilizando **Apache JMeter**, se simularon niveles de concurrencia (baja: 10 usuarios, media: 50, alta: 100, extrema: 200). Hasta 100 usuarios, el sistema mantuvo la disponibilidad ($\leq 90\%$) y tiempos de respuesta aceptables. Bajo carga extrema, se identificaron puntos de saturación previstos, sin pérdida de datos ni fallos críticos.
- **Pruebas UX/UI:** Realizadas manualmente en emuladores y dispositivos físicos. Se validaron formularios, botones, navegación, mensajes de retroalimentación y consistencia visual. Todas las interfaces funcionaron correctamente, sin errores gráficos ni rupturas de flujo.

Entornos y herramientas de prueba

Las pruebas se realizaron en entornos controlados, utilizando:

- **Android Studio** y dispositivos reales/emuladores: pruebas de la app móvil.
- **pytest, Docker, MongoDB, MySQL, Airflow 2.8.1:** pruebas de backend, sincronización y replicación.
- **Apache JMeter:** simulación de carga y pruebas de rendimiento.
- **Scripts de shell:** simulación de fallos y validación de failover automático.

Cobertura de requerimientos

El sistema cumple con la totalidad de los requerimientos funcionales y no funcionales establecidos. A continuación se resumen los principales hallazgos:

Requerimientos funcionales

- **RF-001 a RF-007:** Registro, inicio de sesión, validación de contraseñas y gestión de perfil; grabación, pausa, reanudación y eliminación de audios.
- **RF-008 a RF-012:** Transcripción automática, análisis de emociones y sentimientos, y extracción de palabras clave.
- **RF-013 a RF-017:** Almacenamiento estructurado, eliminación automática de audios, estandarización y replicación hacia bases de datos analíticas.

Requerimientos no funcionales

- **RNF-004:** Seguridad garantizada mediante tokens JWT, autenticación delegada (Auth0) y cifrado HTTPS.
- **RNF-006:** Consultas de historial completadas en ≤ 2 segundos (validado con JMeter).
- **RNF-007:** El 90 % de las transcripciones se completan en menos de 8 segundos.
- **RNF-008:** Los reportes se generan y visualizan en ≤ 25 segundos.
- **RNF-009:** Sistema de recuperación ante fallos probado con replicación MySQL y failover automático.
- **RNF-014:** Disponibilidad ≤ 96 % durante pruebas de carga y estrés.
- **RNF-015:** Cumplimiento estricto de la política de tratamiento de datos; su aceptación es requerida para continuar con el registro.

Restricciones y limitaciones

Generales:

- Usuarios no pueden modificar modelos ni lógica del backend.
- Acceso restringido a mayores de edad o con consentimiento.
- Toda la comunicación está cifrada.
- Capacidad de recuperación ante fallos con respaldo y replicación activa.
- Procesamiento de audios limitado a un máximo de 1 minuto.

De software y hardware:

- Claves y credenciales protegidas mediante variables externas.
- Lenguajes utilizados: Kotlin (app móvil) y Python (backend con FastAPI).

- Requisitos mínimos: Android 8+, infraestructura Ubuntu 24.04.
- Licencias open-source respetadas para librerías y modelos.
- Conectividad constante y capacidad de procesamiento simultáneo garantizada.

Historias de usuario Todas las funcionalidades descritas en las historias de usuario HU-01 a HU-16 fueron implementadas y superaron las pruebas correspondientes, incluyendo autenticación, grabación, procesamiento de emociones y visualización de resultados.

Análisis y conclusiones

La implementación de *myMind* permitió el desarrollo de un sistema técnicamente robusto y funcionalmente completo, capaz de cumplir con los requerimientos establecidos y adaptarse a condiciones reales de uso. Desde el punto de vista técnico, se alcanzaron altos estándares de calidad, medidos a través de métricas de rendimiento, cobertura de código, disponibilidad y estabilidad operativa.

Cada uno de los módulos fundamentales del sistema —autenticación, grabación, transcripción, análisis emocional, almacenamiento y visualización— fue validado de manera rigurosa, tanto de forma individual como integrada. La arquitectura basada en microservicios facilitó el cumplimiento de los requisitos funcionales y no funcionales, mejorando el monitoreo, la resiliencia ante errores y la capacidad de escalar el sistema. Los modelos de procesamiento de lenguaje natural empleados demostraron ser precisos y eficientes en el análisis de emociones y sentimientos en textos en español.

La aplicación móvil mostró un comportamiento estable y consistente en diversos dispositivos Android, garantizando una experiencia de usuario fluida, segura y empática. Además, la solución está preparada para su evolución futura, gracias a una separación clara de responsabilidades, el uso de tecnologías estándar y una base de pruebas exhaustiva y bien documentada.

En conjunto, los resultados obtenidos confirman que *myMind* ha alcanzado un alto nivel de madurez técnica y funcional, cumpliendo plenamente con los objetivos del proyecto y consolidándose como una plataforma confiable, ética y lista para su despliegue en entornos reales.

VIII CONCLUSIONES

4 Análisis de Impacto del Proyecto

El desarrollo del proyecto *myMind* partió de una problemática real y urgente. La dificultad de muchos estudiantes universitarios para identificar y gestionar sus emociones, sumada a la baja utilización de los canales institucionales de apoyo emocional, es una situación especialmente relevante en la vida universitaria. Esta realidad no solo reflejaba un problema de acceso, sino también una falta de conciencia emocional. Frente a ello, el sistema propuesto tenía como fin crear una herramienta accesible, técnicamente robusta y éticamente respetuosa, que ofreciera un primer paso hacia el autoconocimiento, sin pretender sustituir los servicios de salud mental existentes, sino complementarlos desde la tecnología.

Desde la ingeniería de sistemas, *myMind* representa una integración avanzada de múltiples dominios tecnológicos: desarrollo móvil en Kotlin, arquitectura desacoplada basada en microservicios, procesamiento de lenguaje natural especializado en español, almacenamiento distribuido y pipelines ETL automatizados. Esta complejidad se gestionó con principios de modularidad, seguridad, escalabilidad, mantenibilidad y, sobre todo, responsabilidad ética en el tratamiento de datos emocionales.

En el corto plazo, el impacto de *myMind* se refleja en la posibilidad de que estudiantes accedan a un espacio seguro para registrar sus emociones mediante voz, recibir reportes personalizados e iniciar una práctica de autorreflexión guiada por sus propios registros. Este proceso tiene beneficios comprobados para la salud mental: la acción de “desahogarse”, incluso sin mediación externa, puede disminuir niveles de ansiedad y favorecer la identificación emocional, facilitando el desarrollo de tolerancia a la frustración, mayor regulación emocional y claridad sobre el malestar sentido.

A mediano plazo, la herramienta se proyecta como un apoyo cotidiano a la salud emocional, generando conciencia de patrones anímicos y ofreciendo indicadores visuales sobre su evolución. Esto no reemplaza la atención profesional, pero sí la puede acompañar, fortaleciendo el proceso terapéutico o guiando la decisión de buscar ayuda.

En el largo plazo, *myMind* puede contribuir a transformar la relación entre jóvenes y su salud emocional, promoviendo una cultura más abierta al autocuidado afectivo. La aplicación abre un canal íntimo y no invasivo de expresión emocional, validando la experiencia subjetiva de los usuarios y colocándola en el centro del proceso.

Además, el impacto global del proyecto se ubica dentro de un campo técnico en expansión: el procesamiento de lenguaje natural aplicado a emociones, un área que ha ganado enorme impulso en los últimos años. La selección, comparación y ajuste de modelos en español —incluyendo análisis de sentimiento, emociones y voz a texto— constituye una de las contribuciones más valiosas del proyecto, tanto desde el desarrollo aplicado como desde la documentación del proceso. En un contexto donde la mayoría de herramientas de PLN están entrenadas en inglés y carecen de enfoques culturales específicos, *myMind* representa una innovación relevante en el contexto hispanohablante y colombiano.

5 Conclusiones y Trabajo Futuro

El proyecto *myMind* logró con éxito todos los objetivos propuestos en su formulación:

- **Objetivo general:** Se desarrolló una aplicación móvil que permite a los estudiantes expresar-se mediante notas de voz en un entorno seguro y privado, brindando seguimiento emocional mediante reportes visuales generados a partir del análisis de lenguaje natural.

- **Diseñar una aplicación móvil para recolección y transcripción de notas de voz:** La app ofrece una interfaz clara, intuitiva y accesible que permite grabar audios de hasta un minuto, con controles de pausa y reanudación, y transcripción automática inmediata.
- **Aplicar modelos de PLN para extraer información emocional:** Se integraron modelos especializados para análisis de sentimientos y emociones, incluyendo modelos preentrenados y uno ajustado específicamente sobre corpus en español, priorizando la calidad del procesamiento afectivo en nuestro idioma.
- **Implementar el sistema de adquisición y procesamiento:** Se diseñó una arquitectura distribuida basada en microservicios que orquesta el flujo desde la grabación hasta la visualización final.
- **Desarrollar un sistema de generación de reportes personalizados:** El sistema muestra visualizaciones de los resultados emocionales mediante gráficas interactivas por emoción, fecha y tipo de sentimiento, facilitando así el seguimiento reflexivo y comprensible para el usuario.

A lo largo del desarrollo, se mantuvo como principio que **myMind no reemplaza a un psicólogo ni propone diagnósticos**, sino que invita al autoconocimiento, a la expresión emocional y a la complementación de los servicios ya existentes. Es una herramienta de acompañamiento y un espacio fomentado en el desahogo y la expresión libre.

De cara al futuro, el sistema podrá crecer hacia funcionalidades como:

- **Identificación de tópicos y entidades recurrentes en las transcripciones:** Incorporar técnicas de extracción de entidades nombradas (NER) y modelado temático para identificar los temas más mencionados por el usuario, como personas, lugares o eventos, permitiendo así una comprensión más profunda del contexto emocional de cada registro.
- **Alertas voluntarias ante variaciones abruptas en los reportes emocionales:** Implementar un sistema de notificaciones configurables que informe al usuario sobre cambios significativos en su estado emocional (por ejemplo, una caída súbita del sentimiento positivo), sin invadir su privacidad ni emitir juicios, solo como apoyo a la autoreflexión.
- **Integración con modelos de predicción de estado emocional:** Desarrollar modelos de aprendizaje automático que analicen las tendencias emocionales del usuario a lo largo del tiempo, con el fin de anticipar estados futuros y ofrecer herramientas de prevención emocional si el usuario lo autoriza.
- **Personalización avanzada de reportes y generación de informes exportables:** Permitir al usuario filtrar, comparar y analizar sus emociones por rango de fechas, eventos clave o tipos de sentimiento, además de generar informes en formatos descargables para uso personal o compartido con un profesional.
- **Escalamiento a nuevos idiomas, contextos institucionales y públicos vulnerables:** Adaptar la aplicación para su uso en distintos entornos lingüísticos y culturales, incluyendo poblaciones que puedan beneficiarse especialmente del acompañamiento emocional, como jóvenes en contextos de vulnerabilidad, migrantes o personas con discapacidad.
- **Soporte para Múltiples Idiomas:** Incluir opciones de idioma en la interfaz y el procesamiento de voz, con el objetivo de ampliar la accesibilidad y permitir que usuarios de distintas regiones puedan utilizar la aplicación en su lengua materna, mejorando la empatía y precisión del análisis.

- **Implementación de nuevos modelos de aprendizaje:** Ampliar el repertorio de modelos de PLN y aprendizaje automático utilizados, incorporando arquitecturas más recientes o específicas (como transformers multilingües o adaptaciones culturales) para capturar matices más sutiles del discurso emocional.
- **Sistema de Recomendaciones Personalizadas:** Desarrollar una funcionalidad que, a partir del análisis emocional previo, ofrezca recomendaciones contextualizadas —como ejercicios de respiración, audios de relajación, frases motivacionales o enlaces a recursos de salud mental— respetando siempre la autonomía y decisión del usuario.

En definitiva, **myMind** es el resultado de una propuesta que combina rigurosidad técnica, pertinencia social y sensibilidad humana. Ofrece una solución innovadora al permitir que la voz, el medio más natural de expresión emocional, sea escuchada por la tecnología y devuelta al usuario en forma de comprensión. Se enmarca en un campo en expansión como el PLN para emociones, aporta metodologías replicables en la selección y evaluación de modelos lingüísticos en español, y se alinea con los valores éticos de privacidad, cuidado y respeto. Es una aplicación que no solo funciona, sino que acompaña al usuario y le da este espacio tan valioso de comprensión a sí mismo.

IX REFERENCIAS

Referencias

- [1] P. U. Javeriana, *Institucional*, 2024. dirección: <https://www.javeriana.edu.co/institucional>.
- [2] P. U. Javeriana, *Salud Mental a Mi U: cuidado y bienestar emocional para estudiantes universitarios*, 2023. dirección: <https://www.javeriana.edu.co/-/noticia-salud-mental-a-mi-u-cuidado-y-bienestar-emocional-para-estudiantes-universitarios>.
- [3] P. U. Javeriana, *Boletín Estadístico 2022*, 2022. dirección: <https://app.powerbi.com/view?r=eyJrIjoibWQzNzIyZjYtMzEwNC00M2QwLWI1NTYtODlhODhiNjhhNGY0IiwidCI6ImRhZjc5OTBlLThhM2YtNDA5Yy05Yjc2LTJhNTQ3NTA5ODAwMCIscmMiOiR92>.
- [4] P. G. de la Nación, *Suicidio disparado en Colombia por cuenta de trastornos mentales: Procuraduría*, 2023. dirección: <https://www.procuraduria.gov.co/Pages/suicidio-disparado-colombia-cuenta-trastornos-mentales-procuraduria.aspx>.
- [5] U. Colombia, “De salud mental sí hablamos”: UNICEF Colombia propone abrir una conversación sobre la salud mental de niños, niñas y adolescentes, 2024. dirección: <https://www.minsalud.gov.co/Paginas/Salud-Mental.aspx>.
- [6] D. A. N. de Estadística, *¿Que pasa luego del grado? Un estudio comparativo de la situación de los egresados eafitenses al momento de terminar sus Estudios*, 2014. dirección: [https://www.eafit.edu.co/institucional/calidad-eafit/planeacion/Documents/Articulo%20comparativo%20M0%20y%20M1%20\(Compendio\).pdf](https://www.eafit.edu.co/institucional/calidad-eafit/planeacion/Documents/Articulo%20comparativo%20M0%20y%20M1%20(Compendio).pdf).
- [7] D. R. Zaraza Morales y J. R. Contreras Moreno, “El modelo de recuperación de la salud mental y su importancia para la enfermería colombiana,” es, *Revista Colombiana de Psiquiatría*, vol. 49, págs. 305 -310, dic. de 2020, ISSN: 0034-7450. dirección: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0034-74502020000400305&nrm=iso.
- [8] H. Zheng, K. Xu, H. Zhou, Y. Wang y G. Su, “Medication recommendation system based on natural language processing for patient emotion analysis,” *Academic Journal of Science and Technology*, vol. 10, n.º 1, págs. 62-68, 2024.
- [9] Y. Santoya, M. Garcés y M. Tezón, “Las emociones en la vida universitaria: análisis de la relación entre autoconocimiento emocional y autorregulación emocional en adolescentes y jóvenes universitarios,” *Psicogente*, vol. 21, n.º 40, 2018.
- [10] M. Sohal, P. Singh, B. S. Dhillon y H. S. Gill, “Efficacy of journaling in the management of mental illness: a systematic review and meta-analysis,” *Family medicine and community health*, vol. 10, n.º 1, e001154, 2022.
- [11] H. M. Wurtz, S. S. Willen y K. A. Mason, “Introduction: Journaling and mental health during COVID-19: Insights from the pandemic journaling project,” *SSM-Mental Health*, vol. 2, pág. 100141, 2022.
- [12] University of Rochester Medical Center, *Journaling for Emotional Wellness*, <https://www.urmc.rochester.edu/encyclopedia/content.aspx?ContentID=4552&ContentTypeID=1>, [Último acceso: 9 October 2024], 2024.

- [13] J. M. Smyth, J. A. Johnson, B. J. Auer, E. Lehman, G. Talamo y C. N. Sciamanna, "Online positive affect journaling in the improvement of mental distress and well-being in general medical patients with elevated anxiety symptoms: A preliminary randomized controlled trial," *JMIR mental health*, vol. 5, n.º 4, e11290, 2018.
- [14] H. Khanpour y C. Caragea, "Fine-grained emotion detection in health-related online posts," en *Proceedings of the 2018 conference on empirical methods in natural language processing*, 2018, págs. 1160-1166.
- [15] C. Beltrán y C. Rodríguez, "Procesamiento del lenguaje natural (PLN)-GPT-3, y su aplicación en la Ingeniería de Software," *TIA Tecnología, investigación y academia*, vol. 8, n.º 1, 2020.
- [16] C. Stryker y J. Holdsworth, *¿Qué es el PLN (procesamiento del lenguaje natural)?* <https://www.ibm.com/es-es/topics/natural-language-processing>, [Último acceso: 21 October 2024], 2024.
- [17] A. AWS, *¿Qué es la conversión de voz a texto?* [Último acceso: 21 October 2024], Amazon AWS, 2024. dirección: <https://aws.amazon.com/es/what-is/speech-to-text/>.
- [18] OMS, *Salud mental*, [Último acceso: 21 October 2024], 2024. dirección: https://www.who.int/es/health-topics/mental-health#tab=tab_1.
- [19] Microsoft, *¿Qué es un modelo de aprendizaje automático?* [Último acceso: 21 October 2024], 2024. dirección: <https://learn.microsoft.com/es-es/windows/ai/windows-ml/what-is-a-machine-learning-model>.
- [20] Z. Zhang, Y. Sun y R. Yan, "An Intelligent Mobile Application for Depression Relief using Artificial Intelligence and Natural Language Processing," en *CS & IT Conference Proceedings*, CS & IT Conference Proceedings, vol. 12, 2022.
- [21] R. V. Kankaria, A. Agrawal, H. Barot y A. Godbole, "RAAH. ai: An interactive chatbot for stress relief using Deep Learning and Natural Language Processing," en *2021 12th International conference on computing communication and networking technologies (ICCCNT)*, IEEE, 2021, págs. 1-7.
- [22] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del Arco et al., "Overview of mental-risks at iberlef 2024: Early detection of mental disorders risk in spanish," *Procesamiento del lenguaje natural*, vol. 73, págs. 435-448, 2024.
- [23] M. Taboada, "Sentiment Analysis: An Overview from Linguistics," *Annual Review of Linguistics*, vol. 2, págs. 325-347, 2016. DOI: 10.1146/annurev-linguistics-011415-040518.
- [24] K. R. Scherer, "Appraisal Theory," en *Handbook of Cognition and Emotion*. John Wiley & Sons, Ltd, 1999, cap. 30, págs. 637-663. DOI: 10.1002/0470013494.ch30.
- [25] M. Wankhade, A. C. S. Rao y C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, págs. 5731-5780, 2022. DOI: 10.1007/s10462-022-10144-1.
- [26] F. A. Acheampong, C. Wenyu y H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, e12189, 2020. DOI: 10.1002/eng2.12189.
- [27] M. Bain, D. Varis, N. Reimers, V. Raunak y A. Sordoni, "WhisperX: Time-Accurate Speech Transcription of Long-Form," *arXiv preprint*, 2023. arXiv: 2303.00747. dirección: <https://arxiv.org/pdf/2303.00747>.

- [28] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey e I. Sutskever, "Robust speech recognition via large-scale weak supervision," *arXiv preprint*, 2022. arXiv: 2212.04356. dirección: <https://arxiv.org/abs/2212.04356>.
- [29] M. Malik, M. Malik, K. Mehmood e I. Makhdoom, "Automatic speech recognition: a survey," *Multimedia Tools and Applications*, vol. 80, págs. 1-47, mar. de 2021. DOI: 10.1007/s11042-020-10073-7.
- [30] F. Palma Gomez, R. Sanabria, Y.-h. Sung, D. Cer, S. Dalmia y G. Hernandez Abrego, "Transforming LLMs into Cross-modal and Cross-lingual Retrieval Systems," *arXiv preprint*, 2024. arXiv: 2404.01616. dirección: <https://arxiv.org/abs/2404.01616>.
- [31] Y. Peng, X. Wang y J. Lee, "VoiceTextBlender: Augmenting Large Language Models with Speech Capabilities via Single-Stage Joint Speech-Text Supervised Fine-Tuning," *arXiv preprint*, 2025. arXiv: 2410.17485. dirección: <https://arxiv.org/pdf/2410.17485>.
- [32] Avila y J. M. Crego, "Leveraging Large Pre-trained Multilingual Models for High-Quality Speech-to-Text Translation on Industry Scenarios," en *Proceedings of the 2025 Conference on Computational Linguistics (COLING)*, 2025. dirección: <https://aclanthology.org/2025.coling-main.509.pdf>.
- [33] M. C. Stoian, S. Bansal y S. Goldwater, "Analyzing ASR Pretraining for Low-Resource Speech-to-Text Translation," en *Proceedings of ICASSP 2020 - IEEE International Conference on Acoustics, Speech and Signal Processing*, 2020, págs. 7909-7913. DOI: 10.1109/ICASSP40776.2020.9053847.
- [34] H. Ahlawat, N. Aggarwal y D. Gupta, "Automatic Speech Recognition: A survey of deep learning techniques and approaches," *International Journal of Cognitive Computing in Engineering*, vol. 6, págs. 201-237, 2025. DOI: 10.1016/j.ijcce.2024.12.007.
- [35] K. Bu, Y. Liu y X. Ju, "Efficient utilization of pre-trained models: A review of sentiment analysis via prompt learning," *Expert Systems with Applications*, vol. 213, pág. 118987, 2023. DOI: 10.1016/j.eswa.2023.118987.
- [36] P. Pereira, H. Moniz y J. P. Carvalho, "Deep emotion recognition in textual conversations: A survey," *Artificial Intelligence Review*, vol. 58, pág. 10, 2024. DOI: 10.1007/s10462-024-11010-y.
- [37] A. Ngo y J. Kocoń, "Integrating personalized and contextual information in fine-grained emotion recognition in text: A multi-source fusion approach with explainability," *Information Fusion*, vol. 118, pág. 102966, 2025. DOI: 10.1016/j.inffus.2025.102966.
- [38] S. M. Mohammad, F. Bravo-Marquez, M. Salameh y S. Kiritchenko, "SemEval-2018 Task 1: Affect in Tweets," en *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA, 2018.
- [39] M. Richards y N. Ford, *Fundamentals of Software Architecture: An Engineering Approach*. Sebastopol, CA: O'Reilly Media, Inc., 2023, ISBN: 978-1-098-12492-4.
- [40] E. Heller, *Psicología del color: Cómo actúan los colores sobre los sentimientos y la razón*. GG, 2012.
- [41] C. de Psicólogos, *¿Qué significa el color azul?* <https://ebac.mx/blog/que-significa-el-color-azul>, Consultado el 12 de mayo de 2025.

- [42] E. británica de artes creativas y tecnología, *Psicología del color: Morado en marketing*, <https://colegiodepsicologossj.com.ar/psicologia-del-color-morado-en-marketing/>, Consultado el 12 de mayo de 2025.

X ANEXOS