

UNIVERSIDAD PRIVADA FRANZ TAMAYO
FACULTAD DE INGENIERÍA EN SISTEMAS

CARRERA DE INGENIERIA EN SISTEMAS



“Evaluación Procesual”

AUTORES:

Carlos Andrew Lubi Aguilar

TUTOR:

Enrique Alejandro Laurel Cossio

LAPAZ - BOLIVIA

2025

Introducción

El presente proyecto busca desarrollar un software de logopedia enfocado en mejorar la pronunciación y fonética de niños de 3 a 8 años con dificultades del habla. Mediante el uso de tecnologías de Big Data e Inteligencia Artificial, el sistema analizará grabaciones de voz para identificar errores de pronunciación y generar recomendaciones personalizadas.

La etapa de recolección de datos es esencial, ya que permitirá construir una base de audios reales sobre la que se entrenarán los modelos de análisis fonético. Cada grabación será realizada con consentimiento de los padres o tutores y bajo condiciones controladas, garantizando calidad, privacidad y precisión en los resultados.

Este proyecto combina el aprendizaje tecnológico con una aplicación social importante: apoyar el desarrollo comunicativo infantil a través de una herramienta educativa accesible y moderna.

Planteamiento del Problema

En Bolivia y en muchos países de Latinoamérica, miles de personas especialmente niños con trastornos orofaciales, adultos en rehabilitación neurológica y pacientes con dificultades del habla carecen de acceso a terapias de logopedia de calidad. Las terapias tradicionales suelen ser presenciales, costosas y limitadas en su alcance, lo que genera brechas de acceso para personas en áreas rurales o con bajos recursos. Además, los métodos convencionales no aprovechan las grandes cantidades de datos generados durante las prácticas de rehabilitación (grabaciones de voz, métricas de fonética, progresos por sesión). Esto provoca que no exista un análisis profundo ni una personalización adecuada de los tratamientos, reduciendo la eficacia en la rehabilitación del habla.

Formulación del problema

¿Cómo puede se puede identificar de manera efectiva los errores de pronunciación en niños de 3 a 8 años, para ayudar en su fonética y dicción, en el contexto socioeducativo de Bolivia?

Objetivos

Objetivo general

Desarrollar una plataforma inteligente de logopedia que, mediante técnicas de Big Data e inteligencia artificial, mejore la fonética y la dicción de personas con problemas orofaciales, ofreciendo evaluaciones automáticas, seguimiento de progreso y recomendaciones personalizadas para terapeutas y pacientes.

Objetivos específicos

- Diseñar el flujo de captura, preprocesado y almacenamiento de grabaciones de voz y métricas acústicas.
- Desarrollar modelos de Machine Learning que evalúen la calidad fonética y estimen el nivel de falencia del paciente.
- Construir un dashboard web para visualización de métricas, informes y recomendaciones.
- Validar la plataforma mediante pruebas con casos controlados y métricas de mejora.

Alcance

El proyecto incluirá:

- Módulo de captura (web) para grabar ejercicios y enviar las grabaciones al servidor.
- Pipeline de preprocesamiento: normalización, extracción de features (F0, jitter, shimmer, HNR, MFCCs).
- Almacenamiento centralizado en la nube (bases de datos orientadas a big data y almacenamiento de blobs para audios).
- Modelos de clasificación/regresión y dashboard para terapeutas.
- Biblioteca de ejercicios y motor de recomendaciones personalizadas.

No incluye:

- Integración completa con historia clínica de terceros.
- Certificaciones médicas oficiales.
- Hardware propietario (se usará micrófono estándar del dispositivo).

Justificaciones

Justificación social

El proyecto responde a una necesidad de salud pública y educación en contextos con acceso limitado a servicios de logopedia. En Bolivia y en otras regiones de Latinoamérica, muchos niños con trastornos del habla no reciben atención adecuada por carencia de especialistas, costes elevados o ubicación geográfica remota. La plataforma propuesta democratiza el acceso a herramientas de diagnóstico y seguimiento, ofreciendo evaluaciones objetivas y materiales terapéuticos adaptativos que favorecen la inclusión educativa, la comunicación efectiva y el desarrollo socioemocional de la infancia. Además, al documentar y sistematizar progresos, se posibilita la intervención temprana, reduciendo impacto académico y social a largo plazo.

Justificación tecnológica

El uso de arquitecturas Big Data y técnicas de inteligencia artificial permite procesar y analizar grandes volúmenes de señales de voz, extraer características acústicas relevantes y adaptar modelos de predicción que mejoren con el tiempo. Tecnologías como pipelines de ingestión, procesamiento distribuido, almacenamiento de blobs y modelos de Deep Learning (CNN/LSTM/Transformers de audio) garantizan escalabilidad, rendimiento y capacidad de respuesta en tiempo real. La adopción de estas tecnologías moderniza la práctica clínica y crea una base técnica reproducible que puede integrarse con otras soluciones de salud digital.

Justificación científica

Desde el punto de vista científico, la plataforma genera un corpus de datos fonéticos infantiles anotado clínicamente que contribuye al avance del conocimiento en fonética, adquisición del lenguaje y rehabilitación del habla. Los análisis cuantitativos de métricas acústicas (F0, jitter, shimmer, HNR, MFCCs) y su correlación con variables clínicas permiten validar hipótesis sobre patrones de error articulatorio y eficacia de ejercicios. Además, la experimentación controlada

y los modelos predictivos ofrecerán evidencia empírica sobre la efectividad de intervenciones automatizadas o asistidas.

Justificación económica

El sistema reduce costos a largo plazo al disminuir la necesidad de sesiones presenciales frecuentes y al optimizar el tiempo del terapeuta mediante herramientas de análisis automático. Beneficios económicos directos incluyen: menor gasto familiar en desplazamientos y sesiones, aumento de la capacidad de atención por terapeuta (al poder gestionar más pacientes con apoyo tecnológico) y posibilidad de escalado comercial (licenciamiento a centros educativos o de salud). Los costos principales corresponden a desarrollo, infraestructura en la nube y validación clínica; sin embargo, se espera un retorno social y económico positivo al expandir el acceso y mejorar la tasa de éxito terapéutico.

Variables del proyecto

Variables independientes

- Tipo de ejercicio.
- Técnica de análisis utilizada
- Frecuencia de sesiones.
- Personalización del plan.
- Condiciones de grabación.

Variables dependientes

- Score_fonetica
- Calidad_Vocal
- HNR_dB
- F0_Hz
- Índice_Perturbacion
- Precisión del modelo en detección de errores.
- Mejora temporal.

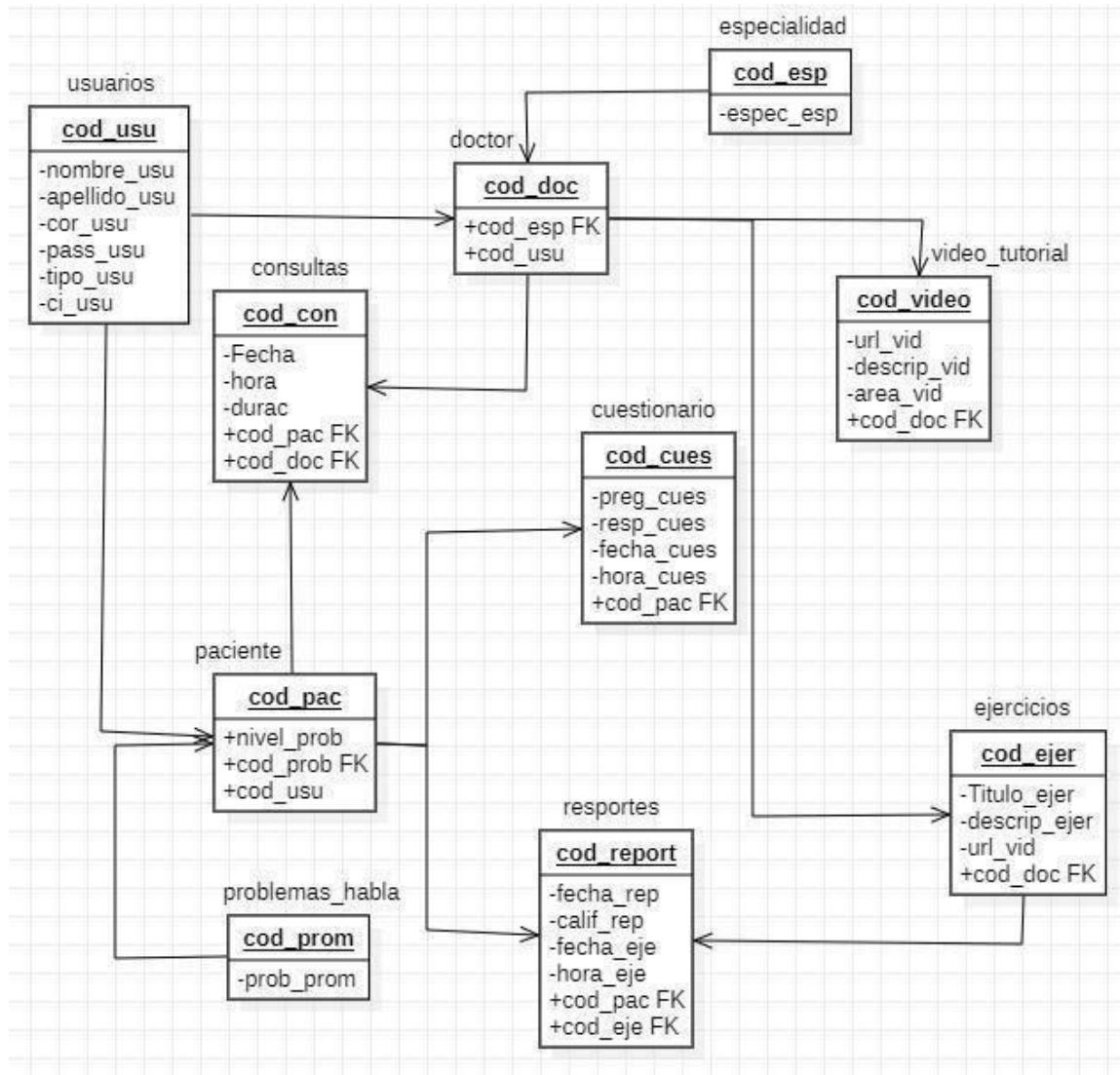
Variables de control (para mantener constantes o registrar)

- Edad del niño
- Género.
- Nivel de ruido ambiente.
- Duración de la grabación.

Variables intervinientes (posibles confusores a monitorear)

- Motivación del niño
- Competencia del terapeuta que etiqueta los audios.
- Calidad de supervisión parental durante ejercicios en casa.

Diagrama de clases de bajo nivel



Esquema de tablas

- pacientes (id_paciente PK, nombre, fecha_nac, genero, telefono, email, observaciones)
- terapeutas (id_terapeuta PK, nombre, credenciales, email)

- sesiones (id_sesion PK, id_paciente FK, id_terapeuta FK, fecha, nota)
- grabaciones (id_grabacion PK, id_sesion FK, id_ejercicio FK, ruta_blob, duracion, formato, timestamp)
- analisis (id_analisis PK, id_grabacion FK, f0, jitter, shimmer, hnr, mfcc_json, score_fonética, created_at)
- ejercicios (id_ejercicio PK, nombre, descripcion, tipo, dificultad, media_url)
- planes (id_plan PK, id_paciente FK, id_terapeuta FK, fecha_inicio, fecha_fin, objetivos)

Plan de recolección de datos (3–8 años)

Objetivo:

Obtener grabaciones de voz de niños de 3 a 8 años para analizar su pronunciación y entrenar modelos que detecten errores de dicción.

Población y muestra:

30 niños (15 con dificultades de habla y 15 sin ellas), reclutados en colegios y centros de terapia, con consentimiento de los padres.

Datos a recolectar:

- Audios WAV
- Edad, sexo, tipo de ejercicio, fecha
- Observaciones del terapeuta

Procedimiento:

1. Consentimiento del tutor
2. Grabación de vocales, sílabas y palabras simples
3. Subida del audio al sistema con metadatos
4. Revisión y etiquetado por terapeuta

Criterios de calidad:

Audio limpio (SNR \geq 20 dB), duración adecuada, datos anónimos y cifrados.

Resultado esperado:

Dataset inicial de 70 audios infantiles listos para análisis fonético

Importancia del Proyecto

Este proyecto busca impacto social y tecnológico, respondiendo a necesidades reales:

- Contribuir a la inclusión educativa y social de personas con problemas de comunicación.
- Aprovechar el Big Data y la inteligencia artificial para generar modelos predictivos y adaptativos en terapias del habla.
- Reducir costos y ampliar el acceso a terapias mediante una plataforma web y móvil conectada a un sistema centralizado de datos.
- Brindar a profesionales de la salud un dashboard avanzado con métricas, patrones de progreso y predicciones sobre la evolución del paciente.

Possible Solución

Se plantea el desarrollo de una plataforma inteligente de logopedia asistida por Big Data, que:

1. Capture la voz del usuario a través de un micrófono de PC o dispositivo móvil.
2. Procése la señal con algoritmos de reconocimiento de patrones y análisis acústico (frecuencia fundamental, jitter, shimmer, HNR).
3. Genere un dataset estructurado y lo almacene en la nube para análisis masivo.

4. Aplique Big Data Analytics y Machine Learning para:
 - Identificar patrones de mejora o retroceso en la fonética.
 - Predecir tiempos estimados de rehabilitación.
 - Recomendar ejercicios personalizados.
5. Muestre resultados en un panel interactivo para logopedas, familias y pacientes.

Cómo se enfrentará técnicamente la solución

- Recolección de datos: Grabaciones de voz, evaluaciones clínicas y métricas en tiempo real.
- Preprocesamiento: Normalización de audio, extracción de características acústicas y limpieza de datos.
- Almacenamiento: Base de datos en la nube.
- Análisis: Algoritmos de Machine Learning (clasificación, regresión y clustering).
- Visualización: Dashboard web con gráficos dinámicos para médicos y terapeutas (Power BI, Tableau o librerías como D3.js).
- Interacción: Aplicación móvil/web para los usuarios, con ejercicios y feedback en tiempo real.

Marco Teórico

Logopedia

La logopedia es la disciplina que aborda los trastornos del habla, lenguaje y comunicación. Tradicionalmente, el proceso terapéutico se basa en la observación directa del logopeda y la repetición de ejercicios por parte del paciente. Sin embargo, los avances tecnológicos han permitido la aparición de la logopedia digital, un enfoque que utiliza sistemas computacionales para registrar, analizar y retroalimentar el desempeño vocal del usuario.

De acuerdo con Ladefoged (2001), el estudio de la fonética acústica se centra en las propiedades físicas del sonido, incluyendo parámetros como la frecuencia fundamental (F0), jitter (variación de frecuencia), shimmer (variación de amplitud) y la relación armónico-ruido (HNR). Estos indicadores cuantitativos permiten evaluar de manera objetiva la calidad de la voz y su estabilidad.

Fonética Acústica

La fonética acústica estudia las propiedades físicas del sonido del habla, analizando su frecuencia, intensidad y duración. Según Ladefoged (2001), estos parámetros permiten caracterizar la voz humana y evaluar su estabilidad. En el proyecto, los sonidos emitidos por los usuarios son transformados en señales digitales que se analizan mediante algoritmos especializados. Este análisis proporciona información cuantitativa sobre la calidad vocal y permite identificar patrones de pronunciación anormales, esenciales para la corrección terapéutica.

Frecuencia Fundamental (F0)

La frecuencia fundamental (F0) representa el número de vibraciones por segundo de las cuerdas vocales y determina el tono de la voz. Es uno de los parámetros más relevantes en la evaluación fonética, ya que revela el control respiratorio y muscular del paciente. Alteraciones significativas en la F0 pueden indicar tensiones, problemas de articulación o variaciones emocionales. Su análisis se realiza mediante transformadas de Fourier y algoritmos de detección de picos en Python, lo que permite obtener valores precisos del tono vocal en cada grabación.

Jitter

El jitter mide las variaciones de frecuencia entre ciclos consecutivos de vibración de las cuerdas vocales. Un valor elevado de jitter indica inestabilidad en la frecuencia, lo que puede reflejar un control insuficiente del tono o irregularidades

en la fonación (Boersma & Weenink, 2023). En este proyecto, el análisis del jitter es esencial para detectar fluctuaciones anormales que afectan la claridad de la pronunciación. Su registro continuo permite observar mejoras progresivas en la estabilidad vocal conforme el paciente avanza en la terapia.

Shimmer

El shimmer cuantifica las variaciones en la amplitud de la señal vocal entre ciclos sucesivos. Es un indicador de irregularidad en la intensidad o fuerza de la voz. Una voz saludable presenta valores bajos de shimmer, mientras que niveles altos pueden asociarse a fatiga vocal, disfonía o falta de control respiratorio. La plataforma medirá el shimmer en cada sesión para identificar alteraciones y ofrecer retroalimentación sobre la uniformidad de la emisión sonora del paciente.

Relación Armónico-Ruido (HNR)

La relación armónico-ruido (HNR) evalúa la proporción entre los componentes armónicos y el ruido presente en una señal vocal. Un valor bajo de HNR indica la presencia de turbulencias o distorsiones, lo que afecta la claridad del sonido. Este parámetro resulta fundamental en la detección de problemas laringeos y en la valoración de la calidad general de la voz. El sistema calculará la HNR junto a otras métricas acústicas para determinar un índice global de calidad fonética.

Big Data

El término Big Data hace referencia al manejo y análisis de grandes volúmenes de información que superan la capacidad de las herramientas tradicionales de gestión de datos. Según Gartner (2012), Big Data “comprende volúmenes de información de gran tamaño, velocidad y variedad que demandan tecnologías innovadoras para capturar, gestionar y analizar datos con el fin de generar valor”. En el ámbito del presente proyecto, Big Data permite almacenar y procesar múltiples grabaciones de voz obtenidas de diferentes pacientes, sesiones y ejercicios, generando una base de conocimiento sobre patrones fonéticos comunes.

El uso de Big Data en logopedia permite crear modelos predictivos que identifican errores sistemáticos en la pronunciación y evalúan la evolución del paciente a lo largo del tiempo. Esto posibilita pasar de terapias reactivas a terapias predictivas y adaptativas, basadas en evidencia cuantitativa. En este sentido, el proyecto se alinea con la tendencia mundial de la salud digital, donde

los datos masivos se convierten en una fuente de mejora continua para la atención personalizada.

Procesamiento de Señales de Voz

El análisis acústico de la voz se fundamenta en la fonética acústica, que estudia las propiedades físicas del sonido del habla. Según Ladeffoged (2001), los parámetros acústicos como F0, jitter, shimmer y HNR permiten caracterizar cuantitativamente la calidad vocal y detectar anomalías en la producción del habla.

Ingeniería de Características

La selección y transformación de variables es crucial en proyectos de machine learning. Wickham (2014) enfatiza la importancia de los datos tidy (datos ordenados) donde cada variable forma una columna, cada observación forma una fila, y cada tipo de unidad observacional forma una tabla.

Limpieza de Datos en Big Data

En el contexto de las 5V de Big Data, la veracidad se garantiza mediante procesos rigurosos de limpieza y validación. Gartner (2012) destaca que la calidad de los datos es fundamental para generar valor en proyectos de análisis masivo.

Arquitectura del Sistema

Usuarios y roles

- Administrador del sistema: gestión de cuentas, permisos, mantenimiento de la plataforma.
- Terapeuta / Logopeda: crea planes, revisa grabaciones, etiqueta audios, monitoriza progreso.
- Paciente / Tutor: realiza ejercicios, sube grabaciones y visualiza retroalimentación simple.

Flujos de datos (descripción paso a paso)

1. **Captura:** Tutor/paciente graba ejercicio en la app → archivo WAV + metadatos.

2. **Subida:** App envía audio a API Gateway → audio colocado en Blob Storage y se envía mensaje en broker indicando nuevo archivo.
3. **Preprocesamiento:** Worker toma mensaje → descarga audio → aplica limpieza de ruido, normalización y segmentación → calcula features → sube features a DB NoSQL y marca estado de análisis.
4. **Inferencia:** Servicio de inferencia toma features → ejecuta modelos de clasificación/regresión → genera score_fonética, etiquetas de errores y recomendaciones de ejercicios → resultados almacenados en DB relacional/noSQL.
5. **Presentación:** Dashboard/usuario solicita resultados → API devuelve reportes visuales, gráficas de evolución y recomendaciones.
6. **Feedback humano:** Terapeuta revisa y etiqueta manualmente subset de audios → etiquetas vuelven al data lake para reentrenamiento.
7. **Retraining:** Periodicamente, pipeline de ML extrae datos anotados → reentrena modelos distribuidos → despliegue de nuevos weights en model serving.
8. **Monitoreo:** Sistemas de logging y métricas vigilan latencia, accuracy en producción y drift; alertas si desempeño cae por debajo de umbrales.

Elección de Variables

Variables Seleccionadas para el Análisis

Variables Acústicas Principales:

- F0_Hz: Frecuencia fundamental (tono de voz)
- Jitter_porcentaje: Variación de frecuencia entre ciclos vocales
- Shimmer_porcentaje: Variación de amplitud entre ciclos vocales
- HNR_dB: Relación armónico-ruido (calidad vocal)

Justificación de la Selección:

Estas variables fueron seleccionadas porque:

- Son métricas estandarizadas en análisis vocal (Boersma & Weenink, 2023)
- Capturan diferentes dimensiones de la calidad vocal
- Tienen correlación clínica con trastornos del habla

Variable Objetivo

Variable Principal: Calidad_Vocal (derivada de HNR_dB)

- Excelente: $HNR > 20 \text{ dB}$
- Buena: $15 < HNR \leq 20 \text{ dB}$
- Regular: $10 < HNR \leq 15 \text{ dB}$
- Mala: $HNR \leq 10 \text{ dB}$

Variables Secundarias de Evaluación:

- Clasificacion_Voz: Basada en rangos de F0
- Indice_Perturbacion: Combinación de jitter y shimmer

Metodología de Limpieza de Datos

Pipeline de Limpieza Implementado

Pipeline de limpieza secuencial:

1. Carga y exploración inicial
2. Detección de valores nulos y outliers
3. Limpieza estratificada:
 - Eliminación de registros completamente nulos
 - Tratamiento de outliers por método IQR
 - Imputación por mediana para valores nulos restantes
4. Enriquecimiento con nuevas variables
5. Validación y almacenamiento

Estrategias Específicas de Limpieza

Valores Nulos:

- Eliminación de filas donde TODAS las variables numéricas son nulas
- Imputación con mediana para valores nulos individuales
- Justificación: La mediana es robusta a outliers

Criterios de Calidad Aplicados:

- Conservar $\geq 85\%$ de los datos originales
- SNR implícito mediante filtrado de outliers
- Distribuciones post-limpieza dentro de rangos fisiológicos

Enriquecimiento de Datos

Nuevas Variables Creadas:

1. Clasificacion_Voz: Categorización por rangos de F0
2. Calidad_Vocal: Clasificación basada en HNR
3. Indice_Perturbacion: Métrica combinada de estabilidad vocal
4. Variables Z-Score: Normalización para modelos ML

Validación de la Calidad Post-Limpieza

Métricas de Evaluación

Conservación de Datos:

- Tasa de conservación objetivo: > 85%
- Filas eliminadas: < 15% del total original
- Variables numéricas sin valores nulos

Calidad Distribucional:

- Distribuciones más Gaussianas post-limpieza
- Reducción de sesgos por outliers
- Mantenimiento de relaciones entre variables

Verificación de Consistencia

Análisis Pre/Post Limpieza:

- Comparación visual de distribuciones
- Verificación de estadísticos descriptivos
- Validación de correlaciones entre variables

Integración con el Proyecto de Logopedia

Aplicación en el Contexto Clínico

Para Terapeutas:

- Datos limpios para evaluación objetiva
- Métricas estandarizadas para seguimiento
- Visualizaciones claras del progreso

Para Modelos ML:

- Dataset preparado para entrenamiento
- Variables normalizadas para algoritmos
- Estructura consistente para análisis predictivo

Escalabilidad Big Data

Preparado para:

- Procesamiento de lotes de grabaciones
- Integración con pipeline de análisis acústico
- Alimentación de dashboard en tiempo real

Resultados y Próximos Pasos

Entregables Generados

- resultados_voz_limpio.csv: Dataset listo para análisis
- Visualizaciones exploratorias completas
- Documentación del proceso de limpieza

Preparación para Fase de Modelado

El dataset limpio está optimizado para:

- Análisis de clustering de patrones vocales
- Modelos de clasificación de trastornos
- Regresión para predicción de progreso
- Sistema de recomendaciones personalizadas

Implementación Técnica Detallada

Visualizaciones Generadas

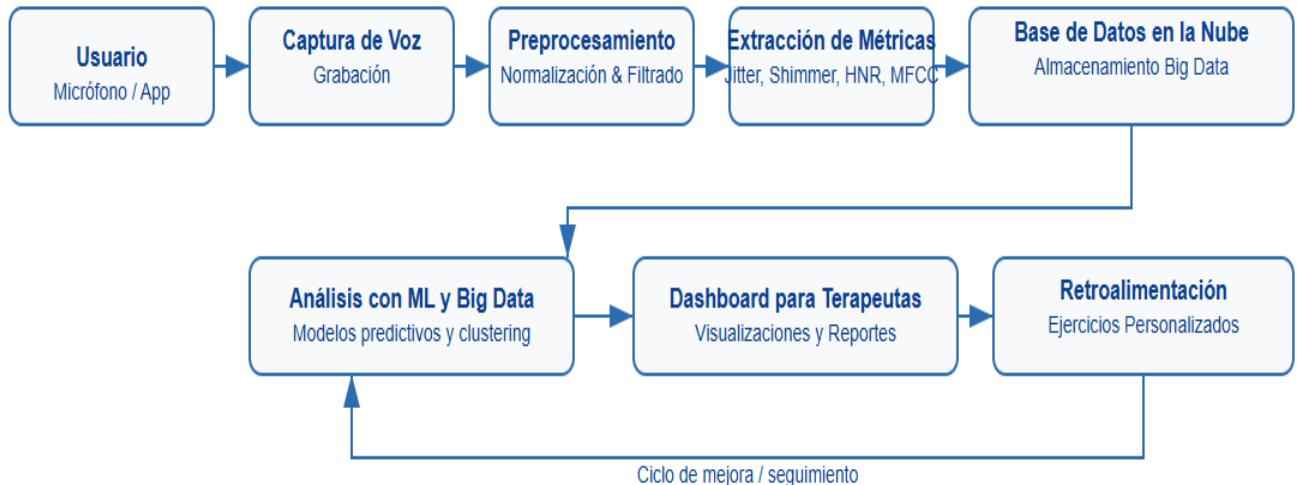
1. Mapa de calor de valores nulos

2. Distribuciones pre/post limpieza
3. Detección de outliers por boxplots
4. Matriz de correlación entre variables
5. Análisis de clasificación vocal
6. Relaciones entre métricas acústicas

Logros Alcanzados

- Dataset limpio y estandarizado para análisis ML
- Variables enriquecidas con significado clínico
- Proceso reproducible para futuras iteraciones
- Documentación completa del pipeline de limpieza

Diagrama de Explicación del Proyecto



Bibliografía

- Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer program]. University of Amsterdam.
- Gartner IT Glossary. (2012). *Definition of Big Data*. Gartner Inc.
- Ladefoged, P. (2001). *A Course in Phonetics* (4th ed.). Harcourt College Publishers.
- (DTW) Techniques. *Journal of Computing*, 2(3), 138–143.
- Wickham, H. (2014). *Tidy Data*. *Journal of Statistical Software*, 59(10), 1–23.