Guion Detallado de la Presentación

Sección 1: Introducción y Análisis Exploratorio (Alejandro)

(Tiempo estimado: 6-7 minutos)

Diapositiva 1: Título

QUÉ DECIR:
"Buenos días/tardes a todos. Mi nombre es Alejandro Pérez, y junto con mis compañeros Jair Gutierrez y Yusmany Rejopachi, hoy les presentaremos nuestro proyecto: Análisis de Emociones en la Voz con Inteligencia Artificial. A lo largo de esta presentación, exploraremos el fascinante desafío técnico de enseñar a una máquina a interpretar los patrones acústicos del habla para clasificar emociones."
PUNTO CLAVE:

- Establecer un tono profesional y centrado en el aspecto técnico desde el inicio.

Diapositiva 2: Justificación Técnica

QUÉ DECIR:
"Para comenzar, ¿por qué enfocarnos en el audio? La voz humana es mucho más que palabras; es una señal acústica increíblemente rica. Contiene características prosódicas, como el tono y el ritmo, y patrones espectrales, como la distribución de energía en las frecuencias, que cambian sistemáticamente con nuestro estado emocional. El reto es que esta información no es estructurada. Aquí es donde entra la Inteligencia Artificial, que nos da la capacidad de procesar esta señal de alta dimensionalidad, extraer patrones complejos de manera objetiva y, en última instancia, modelar la emoción contenida en la voz."
PUNTO CLAVE:

- Enfatizar que el audio es una fuente de datos rica pero compleja, y la IA es la herramienta ideal para analizarla.

Diapositiva 3: Descripción del Problema

QUÉ DECIR:
"El problema técnico que abordamos es, por lo tanto, la clasificación de emociones a partir de una señal de audio. Esto es un desafío porque la voz varía enormemente entre personas, idiomas y contextos. Nuestro objetivo es construir un modelo que pueda manejar esta variabilidad y diferenciar de manera confiable entre siete emociones clave: alegría, tristeza, enojo, miedo, sorpresa, disgusto y neutralidad. Para lograrlo, debemos crear un pipeline de datos robusto, que es una secuencia de pasos que va desde la limpieza del audio hasta la clasificación final."
PUNTO CLAVE:

- Definir claramente el reto: clasificar emociones a pesar de la alta variabilidad del

habla. Mencionar las 7 emociones.

Diapositiva 4: Objetivo General

QUÉ DECIR:
"Con base en lo anterior, nuestro objetivo general es: Desarrollar y evaluar un modelo de inteligencia artificial para la clasificación de emociones humanas a partir del análisis de características acústicas y espectrales del habla, estableciendo un pipeline completo desde el preprocesamiento de la señal hasta la predicción del modelo."
PUNTO CLAVE:

- Leer el objetivo de manera clara y concisa. Es la tesis de su proyecto.

Diapositiva 5: Objetivos Específicos

QUÉ DECIR:
"Para alcanzar nuestro objetivo general, lo desglosamos en cinco objetivos específicos: Primero, integrar y preprocesar diversos conjuntos de datos. Segundo, extraer y analizar las características acústicas más relevantes. Tercero, diseñar y entrenar nuestros modelos de machine learning. Cuarto, evaluar su rendimiento con métricas estándar. Y finalmente, implementar técnicas de reducción de dimensionalidad para visualizar y comprender mejor nuestros datos."
PUNTO CLAVE:

- Presentar los objetivos como los pasos lógicos que siguieron en el proyecto.

Diapositiva 6: Metodología Iterativa

QUÉ DECIR:
"Para llevar a cabo este proyecto, no seguimos un camino lineal, sino una metodología iterativa. Como muestra el diagrama, nuestro proceso es un ciclo. Comenzamos con la adquisición de datos, su análisis y preprocesamiento, y luego pasamos a la extracción de características y el entrenamiento del modelo. La fase clave es la evaluación. Si los resultados no son los esperados, este modelo nos permite regresar a pasos anteriores, como refinar el preprocesamiento o la selección de características, para mejorar el rendimiento. Es un ciclo de mejora continua."
PUNTO CLAVE:

- Explicar que el desarrollo en IA no es lineal, sino un ciclo de prueba, error y mejora.

Diapositiva 7: Adquisición de Conjuntos de Datos

QUÉ DECIR:
"La base de cualquier proyecto de IA son los datos. Utilizamos tres datasets principales para asegurar la diversidad y robustez de nuestro modelo: MESD, que nos da el contexto del español mexicano; RAVDESS, que ofrece audios de alta calidad profesional para establecer un benchmark; y TESS de la Universidad de Toronto, que aporta datos muy consistentes y

controlados. Es importante mencionar que tuvimos que reemplazar un dataset que ya no estaba disponible, lo cual es un desafío común en la ciencia de datos."
PUNTO CLAVE:

- Justificar por qué se usó una combinación de datasets: para obtener variedad cultural, de calidad y de consistencia.

Diapositiva 8: Análisis Exploratorio de Datos (EDA)

QUÉ DECIR:
"Antes de entrenar cualquier modelo, es crucial entender nuestros datos. Esta fase se conoce como Análisis Exploratorio de Datos o EDA. Nos planteamos preguntas como: ¿Existen diferencias claras en el espectro de la voz entre emociones? ¿Cómo cambia el tono? Para responderlas, nos apoyamos en tres tipos de visualizaciones que veremos a continuación: histogramas, mapas de calor y gráficos de dispersión."
PUNTO CLAVE:

- Presentar el EDA como una fase de investigación para entender la materia prima del proyecto.

Diapositiva 9: Visualización: Distribución del Pitch

QUÉ DECIR:
"Una de las características más intuitivas de la voz es el pitch o tono. Esta gráfica nos muestra cómo se distribuye el pitch para cada emoción. Como podemos ver, emociones de alta energía como la 'alegría' o la 'sorpresa' tienden a usar tonos más agudos (más a la derecha en el eje X), mientras que la 'tristeza' se concentra en tonos más graves. Esta clara diferencia nos indica que el pitch es una característica muy poderosa para nuestro modelo."
PUNTO CLAVE:

- Conectar la gráfica con la intuición humana: tonos agudos para alegría, graves para tristeza.

Diapositiva 10: Visualización: Correlación de MFCC

QUÉ DECIR:
"Aquí analizamos los MFCCs, que son una forma de representar el timbre de la voz. Este mapa de calor nos muestra qué tan relacionado está cada uno de los 13 coeficientes MFCC con las diferentes emociones. Por ejemplo, un rojo intenso significa una fuerte correlación positiva. Esta gráfica es fundamental porque nos permite identificar qué coeficientes son más informativos y cuáles son redundantes, lo que es clave para la etapa de selección de características."
PUNTO CLAVE:

- Explicar que esta gráfica no es solo para "ver bonito", sino una herramienta para tomar decisiones sobre qué datos usar.

"Con esto concluye mi parte. Ahora los dejo con mi compañero Jair, quien les hablará

sobre el preprocesamiento y la reducción de dimensionalidad."

Sección 2: Preprocesamiento y Reducción de Dimensionalidad (Jair)

(Tiempo estimado: 6-7 minutos)

Diapositiva 11: Preprocesamiento de Datos

QUÉ DECIR:
"Gracias, Alejandro. Buenos días, mi nombre es Jair Gutierrez. Una vez que entendimos nuestros datos, el siguiente paso fue prepararlos. El audio del mundo real es ruidoso y desordenado, por lo que aplicamos un pipeline de preprocesamiento que incluyó: normalizar todos los audios a un formato estándar, eliminar archivos corruptos, filtrar el ruido, segmentar los audios largos en ventanas más pequeñas con solapamiento, y finalmente, balancear las clases con una técnica llamada SMOTE para que el modelo no se incline a favor de las emociones con más ejemplos."
PUNTO CLAVE:

- Resumir el preprocesamiento como un proceso de "limpieza y orden" para asegurar la calidad de los datos de entrada.

Diapositiva 12: Preprocesamiento para Reducción Dimensional

QUÉ DECIR:
"Un paso crucial antes de aplicar técnicas como PCA y LDA es la estandarización de características. Nuestros datos, como el pitch y los MFCCs, tienen escalas muy diferentes. El pitch puede ir de 100 a 500 Hz, mientras que los MFCCs son valores mucho más pequeños. Sin estandarización, las características con valores más grandes dominarían el análisis. Por eso, escalamos todo para que ninguna característica tenga una ventaja injusta."
PUNTO CLAVE:

- Explicar el "porqué" del escalado: para asegurar un análisis justo y equitativo de todas las características.

Diapositiva 13: ¿Cómo Funciona StandardScaler?

QUÉ DECIR:
"Para lograr esta estandarización, usamos StandardScaler de la librería scikit-learn. Su funcionamiento es simple pero muy efectivo. Primero, calcula la media y la desviación estándar de cada característica. Luego, a cada dato le resta la media y lo divide por la desviación estándar. El resultado es que todas nuestras características terminan con una media de 0 y una varianza de 1. Es como poner a todos los corredores en la misma línea de salida antes de la carrera."
PUNTO CLAVE:

- Usar una analogía (como la línea de salida) para hacer el concepto técnico más

fácil de entender.

Diapositiva 14: Implementación Básica del Algoritmo

QUÉ DECIR:
"Para ilustrar el flujo de trabajo de Machine Learning, aquí mostramos una implementación básica. Primero, dividimos nuestros datos en un conjunto para entrenar y otro para probar. Luego, con el código de en medio, creamos nuestro StandardScaler y un modelo simple de Regresión Logística, y lo entrenamos usando el comando .fit(). Finalmente, usamos el modelo ya entrenado para hacer predicciones sobre los datos de prueba y evaluamos su precisión. Este es el ciclo fundamental que seguimos."
PUNTO CLAVE:

- Mostrar que el proceso de entrenamiento y prueba sigue una secuencia lógica y estandarizada.

Diapositiva 15: Reducción de Dimensionalidad: PCA vs. LDA

QUÉ DECIR:
"Ahora, hablemos de cómo visualizamos nuestros datos de alta dimensionalidad. Usamos dos técnicas: PCA y LDA. La diferencia clave es que PCA es no supervisado: simplemente busca la mayor dispersión en los datos, sin importar la emoción. En cambio, LDA es supervisado: utiliza las etiquetas de las emociones para encontrar la mejor proyección que separe las clases. PCA busca varianza, mientras que LDA busca separación."
PUNTO CLAVE:

- La diferencia fundamental: PCA ignora las etiquetas, LDA las usa.

Diapositiva 16: Visualización 3D: PCA

QUÉ DECIR:
"Aquí vemos el resultado de PCA en 3D. Los ejes que ven, PC1, PC2 y PC3, son los 'Componentes Principales', que son nuevas dimensiones creadas por el algoritmo para capturar la mayor cantidad de varianza o dispersión de los datos. Cada punto es un audio, coloreado por su emoción. Como pueden ver, aunque se forman algunas agrupaciones, hay un gran solapamiento. Esto es esperado, ya que PCA no está diseñado para separar clases, sino para representar la estructura general de los datos."
PUNTO CLAVE:

- Explicar qué son los ejes (Componentes Principales) y por qué el resultado se ve mezclado.

Diapositiva 17: Visualización 3D: LDA

QUÉ DECIR:
"Ahora, comparen con LDA. La diferencia es notable. Los ejes aquí se llaman 'Discriminantes Lineales', y son creados con el único propósito de maximizar la separación entre los colores (las emociones). El resultado es que los cúmulos de cada emoción son mucho más

compactos y están mejor definidos. Esta gráfica es una prueba visual de que las características que extrajimos sí contienen información muy útil para que un modelo pueda aprender a distinguir las emociones."
PUNTO CLAVE:

- Explicar que los ejes de LDA tienen un propósito diferente (separar) y que el éxito de la gráfica valida la calidad de las características extraídas.

Diapositiva 18: Comparación Final y Reflexión

QUÉ DECIR:
"Para resumir, esta tabla muestra las diferencias clave. Mientras que PCA es una gran herramienta para exploración, nuestra reflexión es que LDA es superior para este problema específico, porque su naturaleza supervisada se alinea perfectamente con nuestro objetivo de clasificación. La clara separación que logra nos da confianza para pasar a la siguiente etapa: el entrenamiento de un modelo más complejo."
PUNTO CLAVE:

- Declarar un "ganador" claro (LDA) para el propósito del proyecto y usarlo como transición a la siguiente sección.

"Ahora los dejo con mi compañero Yusmany, quien les explicará cómo usamos estos conceptos para entrenar nuestra Red Neuronal Convolucional."

Sección 3: Modelo de IA y Conclusiones (Yusmany)

(Tiempo estimado: 6-7 minutos)

Diapositiva 19: Fundamento Teórico: CNN

QUÉ DECIR:
"Gracias, Jair. Hola a todos, soy Yusmany Rejopachi. Con la confianza de que nuestros datos son separables, elegimos nuestro algoritmo principal: una Red Neuronal Convolucional o CNN. ¿Por qué una CNN? Porque son expertas en encontrar patrones en imágenes. Nosotros convertimos el audio en espectrogramas, que son esencialmente 'imágenes del sonido'. De esta forma, la CNN puede 'ver' las características de cada emoción. Aunque tienen desventajas como requerir muchos datos, su capacidad para extraer características automáticamente es una ventaja enorme. Para ilustrar, aquí tienen el audio de ejemplo que usaremos en las siguientes diapositivas."
PUNTO CLAVE:

- La idea central: Convertimos el audio en imágenes (espectrogramas) para que una CNN, que es experta en imágenes, pueda analizarlas.

Diapositiva 20: Análisis Espectral (1/4): Forma de Onda

QUÉ DECIR:

"El primer paso es la forma de onda. Esta es la representación más básica del sonido. El eje Y es la amplitud, que percibimos como volumen, y el eje X es el tiempo. Nos permite ver la dinámica general, como las pausas o los picos de intensidad, pero no nos dice nada sobre las frecuencias."

PUNTO CLAVE:

- Es la vista más simple, muestra "cuánto" sonido hay, pero no "de qué tipo".

Diapositiva 21: Análisis Espectral (2/4): Espectrograma Lineal

QUÉ DECIR:
"Para ver las frecuencias, generamos un Espectrograma Lineal. Aquí, el eje Y representa las frecuencias en Hertz, desde las más graves abajo hasta las más agudas arriba. El color nos indica la intensidad de cada frecuencia en cada momento. Los colores amarillos significan alta energía. Ahora ya no solo vemos el volumen, sino también el 'color' del sonido."

PUNTO CLAVE:

- Esta gráfica descompone el sonido en sus frecuencias constituyentes.

Diapositiva 22: Análisis Espectral (3/4): Espectrograma Mel

QUÉ DECIR:
"El oído humano no percibe todas las frecuencias por igual; somos mucho más sensibles a los cambios en las frecuencias bajas. El Espectrograma Mel ajusta el eje Y a la escala Mel, que imita esta percepción humana. Como pueden ver, se le da más espacio y detalle a las frecuencias bajas. Esta es la representación más importante, ya que es la que realmente alimentará a nuestra Red Neuronal."

PUNTO CLAVE:

- Esta es la "imagen" optimizada para la percepción humana y, por lo tanto, para nuestro modelo de IA.

Diapositiva 23: Análisis Espectral (4/4): Cromograma

QUÉ DECIR:
"Finalmente, el Cromograma. Esta visualización es diferente: en lugar de frecuencias, nos muestra la energía de las 12 notas musicales (Do, Re, Mi, etc.). Esto es muy útil para capturar la melodía o entonación del habla. Por ejemplo, una pregunta suele terminar con una entonación ascendente, y un cromograma puede capturar ese patrón."

PUNTO CLAVE:

- El cromograma se enfoca en la "musicalidad" del habla.

Diapositiva 24: Arquitectura de la Red Neuronal

QUÉ DECIR:
"Con los espectrogramas Mel listos, los pasamos por nuestra arquitectura de CNN. La Capa Convolucional usa filtros para detectar patrones básicos como bordes o texturas en la imagen. La Capa de Pooling reduce el tamaño de la imagen para hacerla más manejable y

robusta. Finalmente, las Capas Densas toman todos estos patrones detectados y toman la decisión final, clasificando la emoción. Usamos técnicas como ReLU y Dropout para asegurar que el modelo aprenda de manera eficiente y no solo memorice."
PUNTO CLAVE:

- Describir la CNN como un embudo: empieza detectando patrones simples y pequeños, y termina combinándolos para tomar una decisión compleja.

Diapositiva 25: Recursos

QUÉ DECIR:
"Para la realización de este proyecto, utilizamos herramientas estándar en la industria de la ciencia de datos. Programamos en Python, y nos apoyamos en librerías como TensorFlow y Keras para construir la red neuronal, Librosa para el procesamiento de audio, y Scikit-learn para el preprocesamiento y la evaluación. Todo el entrenamiento intensivo se realizó en Google Colab, aprovechando sus GPUs."
PUNTO CLAVE:

- Mostrar que se utilizaron herramientas robustas y reconocidas.

Diapositiva 26: Alcance del Proyecto

QUÉ DECIR:
"Es importante definir claramente el alcance de nuestro trabajo. Lo que sí hicimos fue desarrollar un modelo completo para clasificar 7 emociones, con todo su pipeline de datos y una evaluación de su rendimiento. Lo que no está incluido es una aplicación para usuario final, el procesamiento en tiempo real o cualquier tipo de diagnóstico. Nuestro enfoque fue puramente el desarrollo y la evaluación del modelo de IA."
PUNTO CLAVE:

- Ser honestos y claros sobre lo que el proyecto es y lo que no es.

Diapositiva 27: ¡Gracias!

QUÉ DECIR:
"En conclusión, nuestro proyecto demuestra que es completamente viable utilizar técnicas de Inteligencia Artificial, como las Redes Neuronales Convolucionales, para analizar la compleja señal de la voz y extraer de ella patrones que corresponden a estados emocionales. Esperamos que esta presentación haya sido informativa y clara. Muchas gracias por su atención. Ahora, con gusto responderemos cualquier pregunta que puedan tener. También, pueden encontrarnos en nuestras redes."
PUNTO CLAVE:

- Terminar con una conclusión fuerte y abrir el espacio para preguntas.