**"Plataforma Interactiva para la Estratificación de Riesgo y Fenotipado de Pacientes con Alzheimer"**.

La idea no es solo predecir si alguien tiene Alzheimer (un simple 0 o 1), sino ofrecer un panorama completo que podría ser útil para un profesional médico.

**## Fase 1: Análisis Exploratorio de Datos (EDA) y Preprocesamiento**

Esta es la base, pero con un enfoque claro.

1. **Limpieza de Datos:** Elimina la columna DoctorInCharge y PatientID (para el modelo). Verifica si hay valores nulos o atípicos y decide cómo manejarlos.
2. **Análisis de Correlación:** Crea un mapa de calor (heatmap) para visualizar las correlaciones entre las variables y la variable objetivo (Diagnosis). Esto te dará pistas sobre qué factores son más influyentes.
3. **Visualización:**
   * Compara las distribuciones de variables clave (como MMSE, Age, BMI) entre pacientes diagnosticados y no diagnosticados usando gráficos de violín o boxplots.
   * Usa gráficos de barras para comparar la prevalencia de factores como FamilyHistoryAlzheimers o Smoking en ambos grupos.

**## Fase 2: Modelo Predictivo y Creación de un "Índice de Riesgo"**

Aquí es donde empezamos a innovar sobre la clasificación estándar.

1. **Construcción del Modelo:** Entrena varios modelos de clasificación (regresión logística, Random Forest, XGBoost) para predecir la variable Diagnosis.
2. **Evaluación Enfocada:** No te quedes solo con la exactitud (accuracy). Mide la **sensibilidad (recall)** y la **precisión (precision)**. En un contexto médico, es crucial no tener falsos negativos (alta sensibilidad). El área bajo la curva ROC (AUC) también es una métrica clave.
3. **La Innovación: El Índice de Riesgo:** En lugar de usar el modelo para predecir un 0 o 1 (diagnóstico sí/no), utiliza la probabilidad que genera el modelo (con predict\_proba). Esta probabilidad (un valor de 0.0 a 1.0) se convierte en tu **"Índice de Riesgo de Alzheimer"**. Es una medida mucho más útil y matizada que una simple clasificación binaria.

**## Fase 3: Identificación de Perfiles de Pacientes (Clustering)**

Esta es la parte más exploratoria e interesante. El objetivo es descubrir si existen "tipos" o "fenotipos" de pacientes con Alzheimer.

1. **Filtrado:** Selecciona solo a los pacientes que han sido diagnosticados con Alzheimer (Diagnosis == 1).
2. **Clustering:** Aplica un algoritmo de clustering como **K-Means** sobre este subconjunto de datos. Puedes usar características clave de estilo de vida, clínicas y cognitivas. El objetivo es agrupar a los pacientes en, por ejemplo, 3 o 4 clusters distintos.
3. **Análisis de Clusters:** Una vez formados los clusters, analiza las características promedio de cada uno. Podrías descubrir perfiles como:
   * **Perfil Metabólico:** Pacientes con alto BMI, diabetes, hipertensión y problemas de colesterol.
   * **Perfil de "Desgaste Cognitivo":** Pacientes con puntuaciones muy bajas en MMSE y FunctionalAssessment pero sin tantos problemas metabólicos.
   * **Perfil de Estilo de Vida:** Pacientes con historial de tabaquismo, consumo de alcohol y baja actividad física.
4. **Visualización:** Utiliza técnicas de reducción de dimensionalidad como **PCA** o **t-SNE** para visualizar estos clusters en un gráfico 2D. ¡Esto queda genial en un portafolio! 📊

**## Fase 4: Dashboard Interactivo con IA Explicable (XAI)**

Esta es la culminación del proyecto, donde unes todo en una herramienta funcional. Puedes usar **Streamlit** o **Dash** en Python para crearlo fácilmente.

El dashboard tendría estas funciones:

1. **Entrada de Datos del Paciente:** El usuario (simulando ser un médico) puede introducir los datos de un nuevo paciente a través de sliders, campos de texto y botones.
2. **Cálculo del Índice de Riesgo:** Al presionar "Calcular", tu modelo predictivo (el mejor de la Fase 2) se ejecuta en segundo plano y muestra el **Índice de Riesgo** (ej: "Riesgo: 78%").
3. **IA Explicable (XAI):** ¡La parte más innovadora! 🚀 Usa una librería como **SHAP** para mostrar *por qué* el modelo dio ese resultado. Genera un gráfico (como un "force plot" o "waterfall plot") que muestre qué factores del paciente aumentaron su riesgo (ej: FamilyHistory = Yes, MMSE = 15) y cuáles lo disminuyeron (ej: PhysicalActivity = 5 horas/semana).
4. **Asignación de Perfil:** Si el riesgo es alto, tu herramienta podría asignar al paciente al "perfil" o cluster más probable de la Fase 3, añadiendo una descripción (ej: "Este paciente corresponde a un Perfil Metabólico").

**## ¿Por qué este proyecto es innovador y bueno para tu portafolio?**

* **Va más allá de la clasificación:** Muestra que puedes pensar en problemas de negocio (o en este caso, clínicos) de manera más profunda.
* **Demuestra un set de habilidades completo:** Cubres EDA, preprocesamiento, modelado supervisado (clasificación), modelado no supervisado (clustering) y despliegue (dashboard).
* **Incorpora técnicas modernas:** El uso de IA Explicable (XAI) con SHAP es una habilidad muy demandada y demuestra que estás al día con las mejores prácticas.
* **Es tangible y visual:** Un dashboard interactivo es mucho más impresionante y fácil de mostrar en una entrevista que un simple Jupyter Notebook.