

## 1. Comprensión del Negocio (Business Understanding) - Fase de CRISP-DM

Esta es la primera fase de CRISP-DM y es crucial porque se enfoca en entender los objetivos del proyecto desde una perspectiva de negocio (o en tu caso, de aplicación clínica/médica). Luego, traduce esos objetivos en una definición clara del problema de minería de datos, identificando los criterios de éxito y los posibles riesgos.

### Tareas principales de la fase de Comprensión del Negocio y su aplicación en nuestro proyecto:

#### 1. Determinación de los Objetivos del Negocio:

- **¿Qué se debería hacer?** Entender qué necesita la organización y cómo la minería de datos puede ayudar a lograr esos objetivos. Esto implica definir el problema de manera precisa y sus metas.
- **En nuestro proyecto (lo que hicimos):** El objetivo principal se definió claramente desde el inicio: "desarrollar un modelo de Deep Learning capaz de clasificar imágenes de radiografías de tórax para identificar la presencia de neumonía." Esto se enmarca en un contexto de apoyo al diagnóstico, buscando "mejorar la eficiencia y precisión en la detección de esta condición." Esto establece el propósito clínico y el valor esperado del proyecto.
- **Lo que se podría haber hecho (ejemplo de mayor detalle):** Si se hubiera trabajado para un hospital específico, se podrían haber definido objetivos más detallados, como "reducir en un 20% el tiempo de diagnóstico de neumonía en urgencias" o "lograr una sensibilidad del 95% para neumonía bacteriana".

#### 2. Evaluación de la Situación Actual:

- **¿Qué se debería hacer?** Analizar el estado actual del problema, los recursos disponibles, las limitaciones, los riesgos y los beneficios esperados.
- **En nuestro proyecto (lo que hicimos):** Implícitamente, se entendió que el diagnóstico de neumonía por radiografía es una tarea que requiere experiencia y tiempo. El uso de "Deep Learning" como solución ya apunta a los recursos

tecnológicos y la capacidad del equipo. El objetivo de "apoyo al diagnóstico" reconoce que la IA asistirá a un proceso existente.

- **Lo que se podría haber hecho (ejemplo de mayor detalle):** Realizar un análisis más profundo sobre la tasa actual de errores en el diagnóstico manual, el tiempo promedio para obtener un resultado, o los costos asociados a diagnósticos tardíos o incorrectos. También, identificar si hay problemas de acceso a radiólogos en ciertas áreas.

### 3. Determinación de los Objetivos de la Minería de Datos:

- **¿Qué se debería hacer?** Traducir los objetivos del negocio en objetivos técnicos específicos de minería de datos.
- **En nuestro proyecto (lo que hicimos):** El objetivo de minería de datos se convirtió en un problema de **clasificación multiclase de imágenes**. Se definió que el modelo debe poder distinguir entre "NORMAL, VIRAL\_PNEUMONIA y BACTERIAL\_PNEUMONIA".
- **Criterios de Éxito:** La elección de métricas como **AUC**, **Precision**, **Recall** y **Accuracy** desde el principio indica que ya se estaba pensando en cómo se mediría el éxito del modelo desde una perspectiva técnica y, en el caso de **Precision** y **Recall**, también médica (ej., la importancia de minimizar falsos negativos en medicina).

### 4. Planificación del Proyecto:

- **¿Qué se debería hacer?** Desarrollar un plan detallado para lograr los objetivos de minería de datos, incluyendo los pasos a seguir, los recursos necesarios, los plazos y los roles.
  - **En nuestro proyecto (lo que hicimos):** Se estructuró el proyecto en fases claras (Importación de librerías, Carga de Datos, Aumento de Datos, Definición del Modelo, Compilación, Callbacks, Entrenamiento, Evaluación, Visualización, Matriz de Confusión) y, más importante, se planteó un enfoque iterativo a través de las "Ejecuciones", lo cual es una forma de planificar el desarrollo y la optimización progresiva.
-

## 2. Comprensión de los Datos (Data Understanding) - Fase de CRISP-DM

En esta fase, el objetivo es familiarizarse con los datos, explorarlos a fondo, identificar posibles problemas de calidad y descubrir los primeros conocimientos o patrones que podrían ser relevantes para el problema de minería de datos.

**Tareas principales de la fase de Comprensión de los Datos y su aplicación en nuestro proyecto:**

### 1. Recopilación de Datos Inicial:

- **¿Qué se debería hacer?** Identificar y recolectar todas las fuentes de datos necesarias.
- **En nuestro proyecto (lo que hicimos):** Los datos (imágenes de radiografías de tórax) ya estaban disponibles y estructurados en carpetas (presumiblemente de un repositorio como Kaggle). La acción de "Se cargan las imágenes con `image_dataset_from_directory`" es el paso de recopilación efectiva en este contexto.

### 2. Descripción de los Datos:

- **¿Qué se debería hacer?** Obtener una primera visión de los datos, incluyendo su formato, volumen, tipos de datos y el significado de los atributos.
- **En nuestro proyecto (lo que hicimos):**
  - **Formato y Tipo:** Se identificó que se trabajaría con imágenes digitales (radiografías).
  - **Volumen:** Se establecieron parámetros como el `BATCH_SIZE = 32`, lo que indirectamente habla del volumen de datos a procesar en cada iteración.
  - **Dimensiones:** Se definió el `IMG_SIZE = (224, 224)`, que describe las dimensiones de las imágenes de entrada.

- **Clases:** Se identificaron las tres clases a clasificar: NORMAL, VIRAL\_PNEUMONIA y BACTERIAL\_PNEUMONIA, y se supo que `label_mode='categorical'` sería apropiado para su codificación.

### 3. Exploración de Datos:

- **¿Qué se debería hacer?** Realizar un análisis exploratorio de datos (EDA) para entender las distribuciones, identificar patrones, correlaciones y anomalías. Esto puede incluir visualizaciones.
- **En nuestro proyecto (lo que hicimos):**
  - **Identificación del Desbalance de Clases:** Aunque la solución (balanceo manual) se implementó en la Ejecución 2, la **detección de que la categoría de Neumonía Bacteriana poseía "aproximadamente 1500 datos extra por sobre las otras 2 categorías"** es un hallazgo clave de la exploración de datos en esta fase. Este desbalance es una característica importante de los datos que se identificó tempranamente y que requería una acción.
  - **Identificación de la necesidad de preprocesamiento:** El hecho de que se definiera la normalización (`Rescaling`) y el aumento de datos (`flip`, `rotation`, `zoom`, `contrast`) en las fases iniciales del proyecto demuestra que se comprendió la naturaleza de los datos de imagen y los requisitos para su procesamiento en una CNN.
- **Lo que se podría haber hecho (ejemplo de mayor detalle):** Se podrían haber visualizado muestras de imágenes de cada clase para entender las diferencias visuales, o graficar la distribución de las clases para confirmar el desbalance. Inspeccionar metadatos si los hubiera (ej., histogramas de edades de pacientes).

### 4. Verificación de la Calidad de los Datos:

- **¿Qué se debería hacer?** Evaluar la calidad de los datos, incluyendo la completitud, consistencia, validez y precisión. Identificar y documentar problemas de calidad.

- **En nuestro proyecto (lo que hicimos):** Al provenir de un dataset de Kaggle, la calidad de las imágenes en términos de integridad (archivos no corruptos) y consistencia en el formato suele ser alta. La identificación del desbalance de clases (mencionado anteriormente) es un aspecto de "calidad" en el sentido de que afecta la representatividad y el aprendizaje del modelo.
- **Lo que se podría haber hecho (si hubiese sido necesario):** Realizar una auditoría más profunda para detectar imágenes duplicadas, errores en las etiquetas (etiquetado incorrecto de una radiografía), o problemas de resolución/calidad inconsistente entre imágenes.

### 3. Preparación de Datos (Fase de CRISP-DM)

La fase de Preparación de Datos se encarga de transformar los datos brutos en un formato limpio y adecuado para el modelado. Se centra en construir el conjunto de datos final, asegurando que sean de alta calidad, consistentes y apropiados para el algoritmo.

Aquí se detallan las tareas principales y cómo se abordaron en nuestro proyecto:

#### 1. Selección de Datos:

- ¿Qué hace? Decide qué datos (registros o atributos/columnas) son relevantes para el proyecto y cuáles deben ser excluidos. Puede ser que ciertos datos no aporten valor, estén incompletos o sean redundantes.
- En nuestro proyecto:
  - **Lo que hicimos:** Se seleccionaron exclusivamente las imágenes de radiografías de tórax, cargándolas con `image_dataset_from_directory`. Estas imágenes ya venían pre-clasificadas en carpetas por tipo de neumonía (Normal, Viral, Bacteriana), lo que facilitó su selección y organización inicial.

Además, se decidió trabajar únicamente con las imágenes, sin incorporar metadatos adicionales del paciente (edad, sexo, etc.), ya que el objetivo principal era la clasificación basada puramente en la imagen.

- Lo que no fue necesario (ejemplo de qué se podría haber hecho): No fue necesario filtrar otros tipos de radiografías (ej., de rodilla, mano) porque el dataset ya estaba pre-filtrado para radiografías de tórax. Si el dataset hubiera incluido otro tipo de datos no relevantes, esta sería la etapa para eliminarlos.

## 2. Limpieza de Datos:

- ¿Qué hace? Aborda problemas de calidad de datos, como valores atípicos (outliers), datos faltantes, inconsistencias o errores. Es fundamental para asegurar la precisión del modelo.
- En nuestro proyecto:
  - Lo que hicimos: No se realizó una limpieza explícita de datos faltantes o valores atípicos severos en las imágenes porque el dataset de Kaggle (fuente común para este tipo de proyectos) tiende a estar relativamente "limpio" en cuanto a la integridad de los archivos y la calidad visual de las imágenes. Se asumió que las imágenes proporcionadas eran completas y válidas.
  - Lo que se podría haber hecho (si hubiese sido necesario):
    - Manejo de Datos Faltantes: Si hubiera habido imágenes corruptas o carpetas vacías, se habrían eliminado o se habría implementado un mecanismo para manejarlas.
    - Detección de Valores Atípicos/Anómalos: Si algunas radiografías tuvieran artefactos graves (ej., sobreexposición extrema, objetos extraños que no son del tórax) que no representaran un caso real de neumonía, se podrían haber identificado y eliminado para evitar confundir al modelo.
    - Consistencia de Etiquetas: Se verificó implícitamente la consistencia al usar `label_mode='categorical'`, asegurando que las etiquetas de las carpetas se tradujeran uniformemente a las clases del modelo.

### 3. Construcción de Datos (Feature Engineering):

- ¿Qué hace? Crea nuevos atributos (características o "features") a partir de los datos existentes que podrían ser más útiles para el modelo. A veces, la información en los datos brutos no es directamente utilizable o es más efectiva si se transforma.
- En nuestro proyecto:
  - Lo que hicimos: Para este proyecto de Deep Learning, la construcción de características (Feature Engineering) a partir de las imágenes no se realizó manualmente. Las Redes Convolucionales (CNN) están diseñadas para extraer automáticamente las características relevantes de las imágenes a través de sus capas convolucionales. Nuestro modelo (CNN secuencial con bloques convolucionales) se encarga de aprender y construir estas "características" a medida que se entrena.
  - Lo que se podría haber hecho (ejemplo de cuándo aplicaría): Si hubiéramos integrado datos clínicos de los pacientes (edad, historial médico, resultados de laboratorio), podríamos haber creado nuevas características combinando esos datos (ej., un "índice de riesgo" o "puntuación de gravedad"), que luego se podrían haber usado junto con las características visuales de la imagen para mejorar la clasificación.

### 4. Integración de Datos:

- ¿Qué hace? Combina datos de múltiples fuentes o bases de datos en un solo conjunto coherente.
- En nuestro proyecto:
  - Lo que hicimos: La integración de datos no fue una tarea compleja en este proyecto. Las imágenes ya estaban agrupadas en una estructura de directorios, y la librería `image_dataset_from_directory` las cargó de manera unificada.

- Lo que se podría haber hecho (si hubiese sido necesario): Si las imágenes de radiografías y sus etiquetas (Normal, Viral, Bacteriana) hubieran estado en bases de datos separadas, o si hubiéramos incorporado datos clínicos de otra fuente (ej., un sistema de gestión hospitalaria), esta sería la fase para unirlos en un único conjunto de datos coherente para cada radiografía.

## 5. Formateo de Datos:

- ¿Qué hace? Transforma los datos en el formato requerido por la herramienta de modelado. Esto incluye escalar, normalizar o codificar variables.
- En nuestro proyecto:
  - Lo que hicimos (y es crucial):
    - Redimensionamiento de Imágenes: Las imágenes se redimensionaron a `IMG_SIZE = (224, 224)`, un tamaño estándar compatible con las arquitecturas de CNN.
    - Normalización/Escalado de Píxeles: Se aplicó `Rescaling` para transformar los valores de los píxeles de las imágenes del rango 0-255 a un rango de 0-1. Esto es fundamental para el buen funcionamiento de las redes neuronales, ya que ayuda a que el proceso de optimización sea más estable y converge más rápido.
    - Codificación One-Hot de Etiquetas: Se utilizó `label_mode='categorical'` para convertir las etiquetas de texto de las clases ('NORMAL', 'VIRAL\_PNEUMONIA', 'BACTERIAL\_PNEUMONIA') a un formato numérico "one-hot" (ej., `[1, 0, 0]`, `[0, 1, 0]`, `[0, 0, 1]`). Este formato es indispensable para la función de pérdida `CategoricalCrossentropy` utilizada en el modelo.
    - División en Conjuntos: Los datos se dividieron en conjuntos de entrenamiento, validación y prueba (70%, 15%, 15% respectivamente desde la Ejecución 2) para asegurar una evaluación justa del rendimiento del modelo.



- Aumento de Datos (Data Augmentation): Se aplicaron técnicas de aumento de datos (**flip**, **rotation**, **zoom**, **contrast**) solo al conjunto de entrenamiento. Esto genera nuevas variantes de las imágenes existentes, aumentando la diversidad del dataset de entrenamiento sin necesidad de recolectar más datos. Es una forma efectiva de enriquecer el conjunto de datos y mejorar la capacidad del modelo para generalizar, reduciendo el sobreajuste. Como se vio en la Ejecución 3, la calibración precisa de estos parámetros fue clave para el rendimiento.
- Optimización del Pipeline de Datos: El uso de **AUTOTUNE**, **cache**, **shuffle** y **prefetch** en la carga de datos permitió acelerar el procesamiento y la alimentación eficiente del modelo, optimizando el flujo de datos.

## 6. Manejo del Desbalance de Clases:

- ¿Qué hace? Aborda situaciones donde algunas clases (categorías) tienen muchísimas más muestras que otras, lo que puede sesgar el aprendizaje del modelo hacia la clase mayoritaria.
- En nuestro proyecto:
  - Lo que hicimos: Identificamos que la categoría de "BACTERIAL\_PNEUMONIA" tenía un número significativamente mayor de imágenes (aproximadamente 1500 datos extra). En la Ejecución 2, se implementó un balanceo manual de los datos (por ejemplo, mediante técnicas como submuestreo de la clase mayoritaria o sobremuestreo de las minoritarias, aunque el texto indica "manual") para igualar la representación de cada clase. Esto fue crucial para evitar que el modelo se sesgara y mejorar su capacidad de aprender de las clases menos representadas, aunque inicialmente impactó el accuracy global.
  - Lo que se podría haber hecho (ejemplo de otras técnicas): Se podrían haber explorado técnicas automáticas de balanceo como SMOTE (Synthetic Minority Over-sampling Technique) para

sobremuestrear clases minoritarias, o usar ponderación de clases directamente en la función de pérdida del modelo.

En resumen, la fase de preparación de datos es donde se hace el "trabajo sucio" pero fundamental. Es donde se asegura que los datos sean fiables, relevantes y estén en la forma correcta para que el modelo pueda aprender de ellos de la manera más efectiva posible. Sin una buena preparación de datos, incluso el mejor algoritmo de Machine Learning dará resultados pobres.

#### 4. Modelado (Modeling) - Fase de CRISP-DM

En esta fase, el enfoque principal es la construcción y el ajuste de los modelos de minería de datos. Se seleccionan las técnicas de modelado apropiadas y se configuran sus parámetros para lograr los objetivos definidos en la fase de Comprensión del Negocio.

**Tareas principales de la fase de Modelado y su aplicación en nuestro proyecto:**

##### 1. Selección de la Técnica de Modelado:

- **¿Qué se debería hacer?** Elegir uno o más algoritmos de minería de datos (clasificación, regresión, clustering, etc.) que sean adecuados para el tipo de problema y los datos disponibles. Se consideran las ventajas y desventajas de cada técnica.
- **En nuestro proyecto (lo que hicimos):** Para la clasificación de imágenes de radiografías, se eligió una **Red Neuronal Convolucional (CNN)**. Esta técnica es la más adecuada para tareas de visión por computadora debido a su capacidad intrínseca para extraer características jerárquicas de datos visuales. Tu proyecto menciona: "Se

implementa una red CNN secuencial con 3 bloques convolucionales que extraen características, seguidos por capas densas que realizan la clasificación."

## 2. Construcción del Modelo:

- **¿Qué se debería hacer?** Diseñar la arquitectura del modelo (si aplica, como en redes neuronales), definir sus componentes y configurar sus parámetros iniciales.
- **En nuestro proyecto (lo que hicimos):**
  - **Arquitectura:** Se definió una arquitectura CNN secuencial con "3 bloques convolucionales" que realizan la extracción de características y capas densas para la clasificación final.
  - **Capas:** Se especificaron las capas de la red, incluyendo capas convolucionales, de pooling, de regularización (**Dropout**) y la capa de salida.
  - **Capa de Salida:** "La capa final tiene 3 salidas **softmax**, correspondientes a las clases: NORMAL, VIRAL\_PNEUMONIA y BACTERIAL\_PNEUMONIA." La función **softmax** es ideal para clasificación multiclase, asignando una probabilidad a cada categoría.

## 3. Evaluación del Diseño del Modelo (interno):

- **¿Qué se debería hacer?** Evaluar cómo funciona el modelo *durante su construcción* o entrenamiento, ajustando los hiperparámetros y la configuración. Esta es una evaluación interna para optimizar el aprendizaje.
- **En nuestro proyecto (lo que hicimos):**
  - **Compilación del Modelo:** "Se usa el optimizador Adam con tasa de aprendizaje baja (0.0001) para entrenamiento estable. La función de pérdida es CategoricalCrossentropy y se evalúan métricas relevantes para medicina: precisión, recall y AUC además de la exactitud." La selección del optimizador (**Adam**), la tasa de aprendizaje y la función de pérdida son decisiones clave para el "cómo" el modelo aprende.

- **Callbacks:** "Se utilizan tres callbacks: **EarlyStopping** (detiene el entrenamiento si no mejora la validación por 5 épocas), **ReduceLROnPlateau** (reduce el learning rate si no hay mejoras), **ModelCheckpoint** (guarda el mejor modelo basado en **val\_loss**).\" Los Callbacks son herramientas esenciales para controlar y optimizar el proceso de entrenamiento del modelo, previniendo el sobreajuste y guardando las mejores versiones.
- **Entrenamiento del Modelo:** "El modelo se entrena utilizando los datasets de entrenamiento y validación durante un máximo de 50 épocas.\" Aquí es donde el modelo ajusta sus pesos internos basándose en los datos.
- **Ajuste Iterativo de Hiperparámetros:** A lo largo de las "Ejecuciones\" (especialmente la Ejecución 3 y 4), se realizaron ajustes iterativos a hiperparámetros como el **BATCH\_SIZE** (tamaño de lote), los parámetros de aumento de datos (rotación, zoom, contraste) y la cantidad de filtros en las capas convolucionales. Estos ajustes buscan encontrar la configuración óptima para el modelo.

---

## 5. Evaluación (Evaluation) - Fase de CRISP-DM

Aquí es donde aclaramos tu duda. La fase de Evaluación en CRISP-DM no es *solo* si el modelo es relevante para el negocio, sino que es una **evaluación exhaustiva de la calidad y el rendimiento del modelo desde una perspectiva tanto técnica como de "negocio" (o aplicación práctica en tu caso)**. Se compara el modelo con los objetivos iniciales y se decide si el proyecto puede proceder a la fase de despliegue.

**Tareas principales de la fase de Evaluación y su aplicación en nuestro proyecto:**

### 1. Evaluación de los Resultados del Modelo:

- **¿Qué se debería hacer?** Medir el rendimiento del modelo utilizando métricas relevantes para el problema y el tipo de modelo.
- **En nuestro proyecto (lo que hicimos):** "Después del entrenamiento, el modelo se evalúa con el conjunto de prueba. Las métricas obtenidas incluyen: **Accuracy**, **Precision**, **Recall**, **AUC**." Se realizó esta evaluación para cada una de las Ejecuciones. Estas métricas son fundamentales para entender qué tan bien el modelo clasifica las imágenes.
- **Visualización de Curvas de Entrenamiento:** "Se grafican las curvas de pérdida y precisión para los conjuntos de entrenamiento y validación. Estas curvas permiten verificar si el modelo está aprendiendo correctamente y si hay signos de sobreajuste." Estas son evaluaciones visuales del proceso de aprendizaje.
- **Matriz de Confusión y Reporte de Clasificación:** "La matriz de confusión permite visualizar los aciertos y errores del modelo en cada clase. Se acompaña de un informe de clasificación que detalla precisión, recall y F1-score por clase." Estas herramientas proporcionan una evaluación detallada del rendimiento del modelo para cada clase (Normal, Neumonía Viral, Neumonía Bacteriana), permitiendo identificar dónde el modelo acierta y dónde falla.

## 2. Revisión del Proceso de Minería de Datos:

- **¿Qué se debería hacer?** Revisar el proceso completo (desde la comprensión del negocio hasta el modelado) para asegurar que todos los pasos se siguieron correctamente y no hay errores u omisiones que pudieran haber afectado los resultados.
- **En nuestro proyecto (lo que hicimos):** Tu análisis comparativo de las "Ejecuciones 1, 2, 3 y 4" es una forma de revisión del proceso. Al identificar que el desbalance de datos (Ejecución 2) o la complejidad excesiva (Ejecución 4) afectaron el rendimiento, y que el ajuste de parámetros (Ejecución 3) lo mejoró, estás evaluando el impacto de las decisiones tomadas en fases anteriores.

## 3. Determinación de la Relevancia para los Objetivos del Negocio/Clínicos:

- **¿Qué se debería hacer?** Comparar los resultados del modelo con los criterios de éxito definidos en la fase de Comprensión del Negocio. ¿El modelo es suficientemente bueno para ser útil en el contexto clínico? ¿Cumple con las expectativas de eficiencia y precisión?
  - **En nuestro proyecto (lo que hicimos):** Tu "Conclusión General de Nuestro Proyecto" es el reflejo de esta tarea. Allí mencionas que la **Ejecución 3** condujo "al modelo con el mejor rendimiento general y el balance más deseable entre precisión y recall para todas las clases", lo cual lo hace "prometedor" como "herramienta de apoyo al diagnóstico" (tu objetivo inicial). También señalas por qué la Ejecución 4 no fue la mejor a pesar de un cambio ambicioso.
  - Las "Consideraciones Clínicas" también son parte de esta evaluación, ya que discuten la validación rigurosa por parte de expertos médicos y la calibración del umbral de decisión, que son requisitos de "negocio" o aplicación práctica.
- 

## 6. Despliegue (Deployment) - Fase de CRISP-DM

En esta fase final, los modelos que se consideran "útiles" se implementan en el entorno operativo. Esto puede variar desde un informe simple y estático hasta un sistema automatizado interactivo.

**Tareas principales de la fase de Despliegue y su aplicación en nuestro proyecto:**

### 1. Planificación del Despliegue:

- **¿Qué se debería hacer?** Decidir cómo se integrará el modelo en el entorno de producción. ¿Será una aplicación, un informe, una API? Se planifican los recursos, el monitoreo y el mantenimiento.
- **En nuestro proyecto (lo que hicimos):** Aunque no se describe un despliegue completo, el **objetivo inicial de crear una "herramienta de apoyo al diagnóstico"** implica una planificación futura de despliegue. No se detalla un plan técnico, pero la intención de uso está clara.

## 2. Planificación del Monitoreo y Mantenimiento:

- **¿Qué se debería hacer?** Establecer cómo se monitoreará el rendimiento del modelo una vez desplegado y cómo se mantendrá (reentrenamiento, actualizaciones).
- **En nuestro proyecto (lo que hicimos):** No se especifica un plan de monitoreo o mantenimiento en tu descripción. Esto sería algo a detallar en un proyecto ya en fase de producción.
- **Lo que se podría haber hecho:** Indicar que, una vez desplegado, el modelo necesitaría monitoreo constante de su rendimiento en datos nuevos (drift) y reentrenamiento periódico con nueva información para mantener su precisión.

## 3. Generación del Informe Final/Presentación:

- **¿Qué se debería hacer?** Presentar los resultados de la minería de datos de una manera que sea fácilmente comprensible para las partes interesadas (stakeholders).
- **En nuestro proyecto (lo que hicimos):** Toda tu documentación, tus "Ejecuciones" detalladas, los gráficos, matrices de confusión, reportes de clasificación, y especialmente tu "Conclusión General", son los componentes clave de este informe final. Este es el resultado tangible del proyecto que se presenta.

## 4. Revisión del Proyecto (Post-Mortem):

- **¿Qué se debería hacer?** Una revisión formal de lo que se aprendió del proyecto, lo que funcionó y lo que no, para mejorar futuros proyectos.
- **En nuestro proyecto (lo que hicimos):** Tu "Conclusión General de los Cambios" donde analizas el éxito y los desafíos de cada ejecución es un ejemplo perfecto de esta revisión. Extraes lecciones aprendidas sobre la optimización de hiperparámetros, el balanceo de datos y la complejidad del modelo.

## 5. Consideraciones Legales y Éticas:

- **¿Qué se debería hacer?** Abordar cualquier cuestión legal, ética o de privacidad relacionada con el uso de los datos y el despliegue del modelo.
- **En nuestro proyecto (lo que hicimos):** Tus "Consideraciones Clave para el Uso Clínico y Ético" (sesgo, transparencia, responsabilidad) son discusiones esenciales que deben ocurrir en esta fase de despliegue para asegurar que la herramienta sea utilizada de forma responsable y alineada con las normativas.