

Taller de Ciencia de Datos: Competencia Kaggle sobre Predicción de Complicaciones en Cáncer de Mama

1. Introducción

Este taller se desarrollará como una competencia en Kaggle, donde los estudiantes trabajarán en equipos de **dos personas** para construir un modelo de clasificación que prediga si un paciente con cáncer de mama en una EPS tendrá una complicación en los próximos 6 meses. La evaluación se basará en la métrica F1-score sobre un conjunto de test privado.

2. Lineamientos Generales

1. Creación de Cuenta en Kaggle

- Si aún no tienen cuenta deben registrarse en [Kaggle](#).

2. Inscripción en la Competencia

- La competencia estará alojada en Kaggle con el nombre "**FA I 2025-I: modelos tradicionales- CA MAMA**".
- Los estudiantes deben inscribirse en equipos de **2 personas**.
- Cada equipo debe ingresar a la competencia y aceptar los términos.

3. Lectura del Overview y Diccionario de Datos

- Revisar el **Overview** de la competencia para entender el objetivo, los datos disponibles y la forma de evaluación.
- Analizar el **Diccionario de Datos**, donde se explican las variables disponibles en el dataset, su significado y posibles valores.

4. Revisión Inicial del Código de Entendimiento de los Datos

- Se proporcionará un **notebook base** con análisis exploratorio y visualización de datos.
- Los estudiantes deben ejecutar y analizar el notebook para familiarizarse con el dataset y detectar patrones iniciales.

3. Evaluación de la Competencia

- **Métrica de evaluación:** El desempeño del modelo será medido mediante el **F1-score** sobre un **test set privado**.

- **Asignación de Puntos:** De acuerdo a los resultados finales de la competencia habrá un valor adicional entre 0 y 1 a la nota promedio de los quices
 - Promedio de quices + Remuneración Adicional por Desempeño en la competencia
- **Remuneración Adicional por Desempeño:**
 - Se asignará un **puntaje adicional** dentro del 20% de la nota de quices, basado en la siguiente función de normalización del F1-score:

$$P_i = \left(\frac{F1score_i}{F1score_{max}} \right)^3$$

Donde:

- $F1score_i$ es el F1-score del equipo i.
- $F1score_{max}$ es el mayor F1-score obtenido en la competencia.
- Esto significa que el equipo con mejor desempeño recibirá un puntaje adicional de **1**, y los demás serán escalados proporcionalmente.

4. Entregables y Fechas Clave

- **Cierre de la competencia en Kaggle:** 25-05-2025

5. Reglas y Consideraciones

- **Máximo 5 envíos por día** en Kaggle.
- **No se permite compartir código entre equipos.**
- Se recomienda documentar el proceso de modelado y decisiones tomadas.