

## **Esquema del Proyecto**

### **Datos de entrada**

Los datos de entrada son 4 archivos con mediciones biométricas tomadas de pacientes. Existen múltiples registros de datos por cada paciente.

### **Objetivos del proyecto**

El objetivo del proyecto es identificar a pacientes con potenciales problemas de salud.

### **Metodología**

#### **1.- Limpieza de datos**

Algunos datos son incorrectos. El alumno debe utilizar técnicas de limpieza de datos para eliminar outliers y garantizar que únicamente datos limpios entren a la etapa de modelado. Existen casos donde los datos no están dentro de los rangos plausibles y otros en donde el identificador del paciente no es usado propiamente.

#### **2.- Integración de datos**

El mismo identificador de usuario es utilizado en los diferentes archivos, es necesario integrar adecuadamente los datos en un solo dataframe.

#### **3.- Filtrado de datos**

Es necesario especificar el criterio de selección de datos, ya que existen múltiples medidas por cada alumno. La calidad del modelo resultante dependerá de esta selección.

#### **4.- Modelado de datos**

El alumno justificará el método de aprendizaje máquina a utilizar. Para este problema en particular se sugiere utilizar un método no supervisado seguido de un método supervisado; sin embargo, el alumno puede utilizar cualquier método proveyendo una justificación adecuada.

#### **5.- Presentación de datos**

Se sugiere utilizar Shiny para presentar el modelo final en una página web. El usuario podrá introducir una serie de valores y el resultado será diagnosticar si el paciente se encuentra en riesgo de enfermedad.

### **Evaluación**

Se evaluarán los cinco aspectos de la metodología. Cada aspecto contribuirá con un 20% a la calificación.