

Informe EPRIEX

Laura Romero Jaque

Durante la realización del contrato EPRIEX he llevado a cabo las siguientes actividades:

- **Curado de bases de datos:**

Se realizaron labores de curado de las 3 bases de datos proporcionadas para el proyecto sepsis, para ello fue necesario el uso de programas de manejo de datos como R.

Primero se realizó una descripción de las variables que componían cada base de datos, registrando su nomenclatura, nombre, tipo de variable (numérica/categórica) así como sus unidades.

A continuación se estudiaron las variables comunes entre cada base de datos, buscando un estándar común para el correcto fusión de las bases de datos sin pérdida de información y manteniendo la calidad de los datos.

Se estudiaron las variables comunes entre cada base de datos, buscando un estándar común para el correcto fusión de las bases de datos sin pérdida de información, para ello fue clave homogenizar categorías, recodificando ciertas variables categóricas como sexo, estado, exitus, etc., por ejemplo en el caso de que tuvieran solo dos valores, estos se cambiarían a 0 y 1, con el fin de que todas las variables sigan una estructura común.

- **Data quality.**

Con el fin de realizar un correcto análisis y seguimiento de la calidad de los datos así como de la integridad de los mismos, se realizaron dos funciones mediante R para poder medir y observar la evolución de dichas variables en el tiempo.

- **Complejidad (*Completeness*):** Esta dimensión hace referencia al grado en que se rellenan los valores de datos necesarios y no se dejan en blanco.
- **Consistencia (*Consistency*):** Esta dimensión comprueba si los valores de los datos almacenados para el mismo registro en fuentes dispares no presentan contradicciones y son exactamente iguales, tanto en términos de significado como de estructura y formato.
- **Semántica:** Esta dimensión hace referencia así existe un estándar consensuado del significado de todos los atributos que intervienen en el conjunto de datos, para que puedan ser verificados.
- **Precisión:** Esta dimensión refiere a en qué medida los valores de los datos representan la realidad.

- **Página web**

Con este estudio realizado se procedió a generar el modelado conceptual de un modelo de bases de datos relacional que permitiera establecer un estándar teniendo en cuenta los datos proporcionados y bibliográficos. Y teniendo este como base se creó una aplicación web mediante php,html,css y javascript, así como la base de datos en mysql y un servidor remoto Linux y web Apache, fue posible crear la página web Hospital Management of Sepsis, epidisease. Para un mayor detalle del funcionamiento de la página web están disponibles tanto el manual de usuario como el código.

