Fase 2: entendimiento de los datos

En el proyecto he utilizado datos provenientes de la información nutrimental de cada alimento que he comido desde hace varias semanas proporcionadas por la aplicación Samsung Healt la cual contiene datos sobre varios alimentos y se ha usado como punto de referencia sobre cada alimento que he comido pues me he visto con la obligación de considerar estos datos no tan certeros ya que es una información genérica sobre lo que contiene cada alimento en cierta cantidad de masa o que la forma en como se prepararon los alimentos puede variar y con eso cambiar los valores de cierta manera de las variables a considerar de cada alimento para este trabajo. Por lo que se ha proseguido de una manera en considerarlos no tan certeros, pero sí aproximados a lo real. En este trabajo los datos manejados representan la información nutrimental cómo las calorías, carbohidratos, los lípidos, las proteínas y el sodio de cada uno de los alimentos consumidos cada día de la semana durante los últimos meses, por lo cual se ha considerado actualizar la tabla cada semana de la información de cada alimento consumido en cada día el domingo de cada semana por lo cual se van a seguir registrando alimentos semanalmente hasta la fecha de conclusión de registro de datos de este proyecto de investigación.

Las columnas de la base de datos que se lleva registrando parecen más prometedoras las de carbohidratos, lípido y proteína por la relación que tienen y que se ha presentado durante el análisis de los datos que son los que tienen una mayor relación con la variable dependiente que es calorías y la variable que al parecer podemos descartar de nuestro modelo es el sodio.

En todo estudio estadístico es siempre mejor tener la mayor cantidad de datos posibles para poder realizar un modelo más preciso que permita explicar el comportamiento de algún fenómeno por lo que con una mayor cantidad de datos más exacto es el modelo y en este caso con la cantidad de datos que se tienen son una buenos para realizar conclusiones y poder explicar cuan posibles es que algo ocurra, pero siempre con más datos es mejor.

En cuanto a los atributos que hay para realizar un modelo en este caso se ha reflejado que con los que se manejan han sido muy buenos para realizar el trabajo ya que han mostrado una gran relación en el análisis y entendimiento de las calorías que hay en cada alimento por lo que ha sido más que suficiente para este modelo en lo que se ha visto reflejado.

Todos los datos obtenidos provienen de la misma fuente que es Samsung Healt y son actualizados cada semana, todos los datos vienen completos. Hasta el momento la cantidad de datos a los que se puede acceder son 190 y algunos son con valores aproximados a lo que realmente se ha estado consumiendo.

Una de las hipótesis puede ser que si falta una de las variables ya escogidas disminuye la calidad en la predicción de las calorías.