Base de datos que quiere medir el índice de felicidad en un país en base a el índice económico, etc

Data de dulces y caramelos . Es un ranking de caramelos.

Dataset que nos permite evaluar cuales son los factores de risfo de salud cardiaca

¿Como afectan nuestros features a los modelos de machine learning?

¿más features siempre es mejor?

NO

Si tenemos features que no son relevantes para el entrenamiento de nuestro modelo, generará ruido

* Features Irrelevantes
* Valores Faltantes
* Introducción de ruido
* Costo computacional

¿qué podemos hacer?

Técnicas de feature Selection y feature Extraction (PCA)

Es la más importante y la más usada.

Consiste en tomar nuestros features y convertirlas en un espacio menor en donde solo conservaremos la info más importantw

Regularización

Usado en varios modelos lineales

Penalizar aquellos features que no aportan michos

Balanceo: OverSampling y UnderSampling

Lo qe hace es tratar de balancear los datos de cada features.