**Feature Selection**

1. **Removing redundant features**
2. **Selecting features using text vectors**
3. **Dimensionality reduction**

What is feature selection?

Selecting features to be used for modeling

Doesn't create new features

Improve model's performance

**1- Removing redundant features**

Redundant features

* Remove noisy features
* Remove correlated features
* Remove duplicated features
* Missing Values
* Low Variance:

Let’s think of a scenario where we have a constant variable (all observations have same value, 5) in our data set. Do you think, it can improve the power of model? Ofcourse NOT, because it has zero variance. In case of high number of dimensions, we should drop variables having low variance compared to others because these variables will not explain the variation in target variables.

Correlated features

* Statistically correlated: features move together directionally
* Linear models assume feature independence
* Pearson correlation coefficient

df.corr()

# bu komut veya digger gorsel metotlarla korelasyon katsayilarina ulasabiliriz ama soru su: katsayi kac olursa feature’I cikarmam gerekiyor. Sadece birini cikaracaksam hangisi olacagina nasil karar vericem?

**2- Selecting features using text vectors**