Attribute Selection

When you remove an attribute?

1. Relevance or Importance
The columns is not relevant to business objective, or not effect on target.
Example: name of student in grades and GPA dataset
Example: Employee ID in dataset of salaries

2. Redundancy
The columns consists of very low probabilities, that cause high redundancy in values; so it has low-to-zero effect on target.
Example: age category in datasets of teenagers only

3. Empty
The column contains a lot of empty values. It will cause a problem in model accuracy.

4. Distinct
The column contains very distinct values. So, it consists of very high probabilities; so it cause high difficultly in pattern extraction.