Preprocessing data

Presented to you by

Dipl. Inform.(FH) Jony Sugianto, M. Comp. Sc.
WA:0812-130-86659
Github: https://github.com/jonysugianto

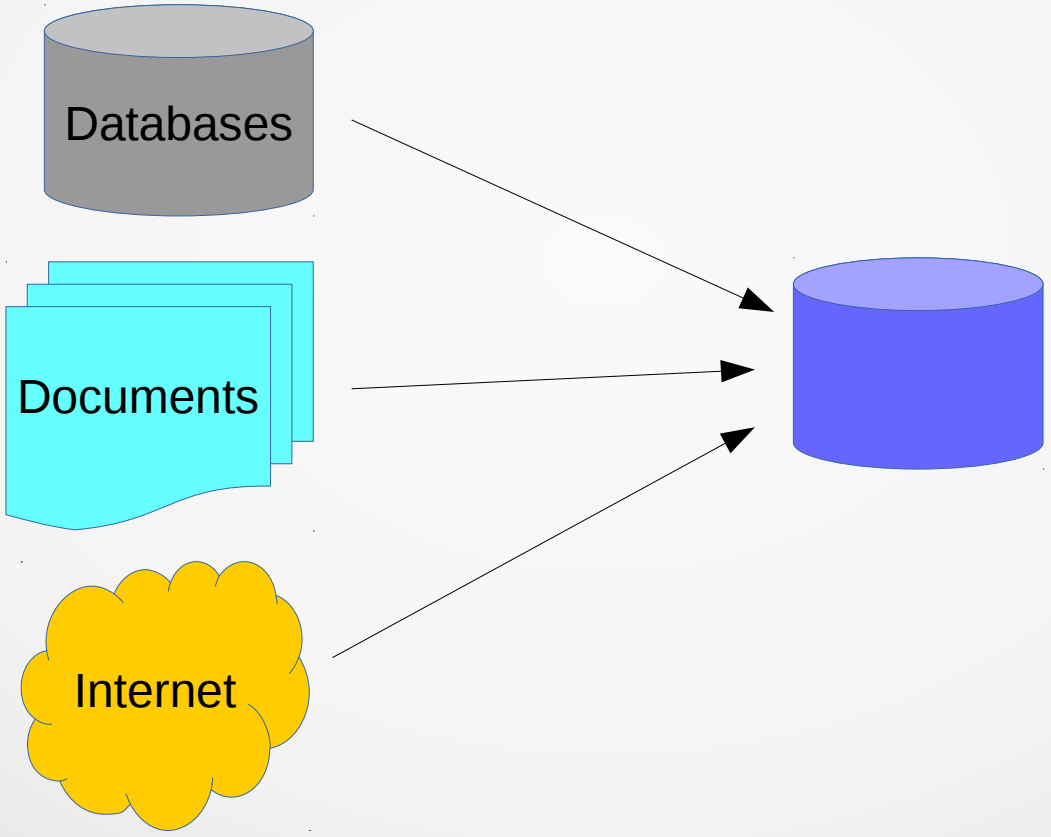
Why Preprocessing ?

The real world data is gathered from various sources and influenced by negative factors such the presence of noise, missing values, inconsistent and superfluous data and huge sizes in both dimensions, examples and features.

Preprocessing to improve quality data

Data Integration

Merging of data from multiple data stores.



Preprocessing to improve quality data

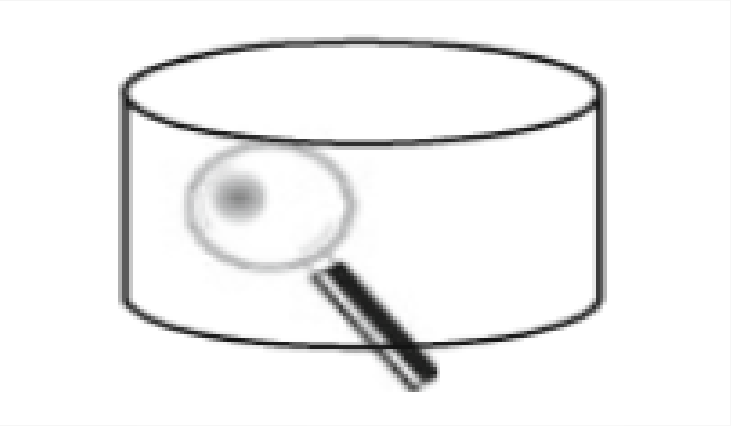
Data Cleaning

Correct bad data, filter some incorrect data out of the data set and reduce the unnecessary detail of data.

Preprocessing to improve quality data

Noise Identification

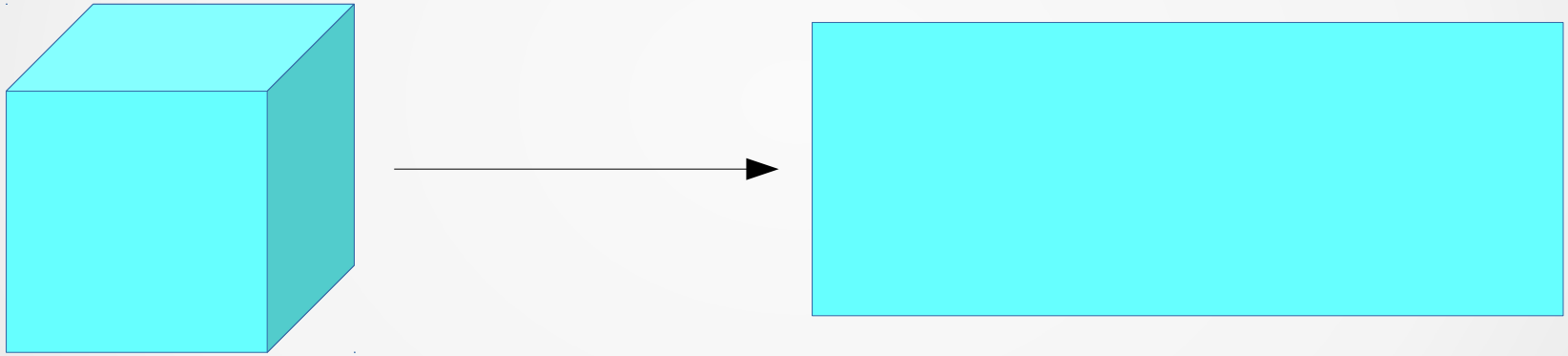
To detect random errors or outliers in a measured variable.



Preprocessing to improve quality data

Data Transformation

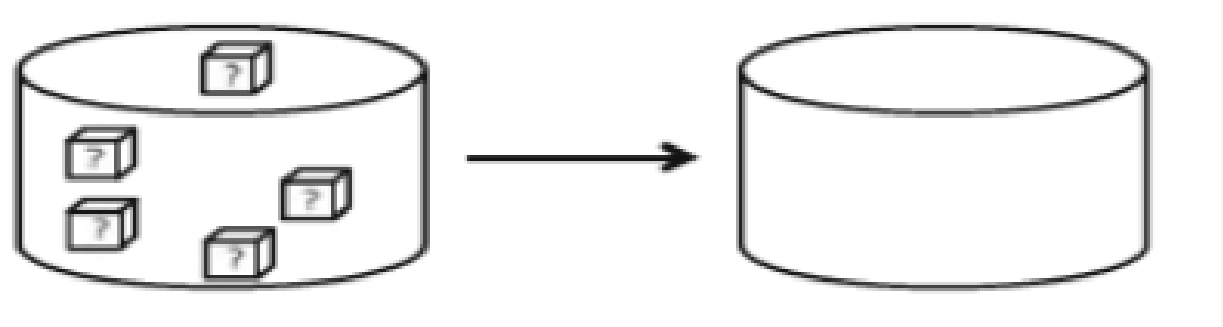
Used to get data into a form that makes it more suitable for a chosen learning method.



Preprocessing to improve quality data

Missing values Imputation

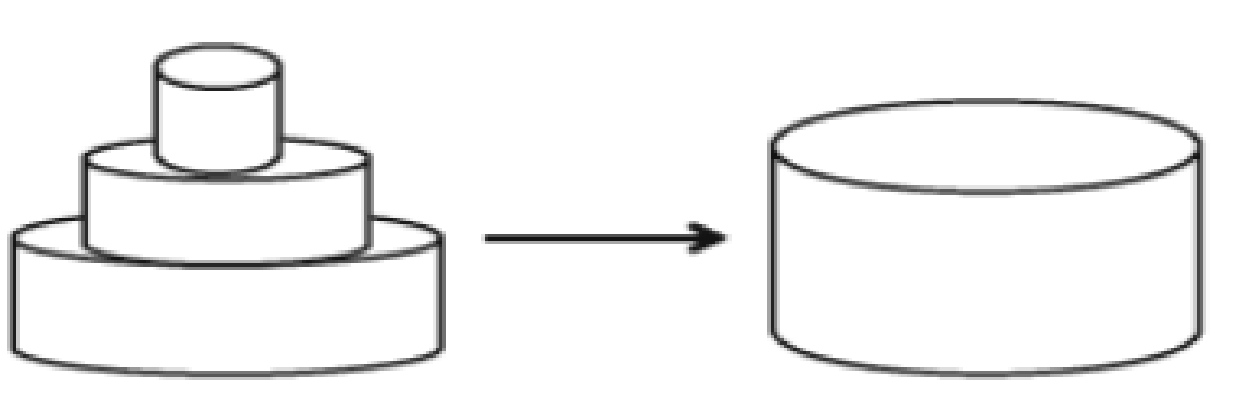
To fill the variables that contain missing values with some intuitive data.



Preprocessing to improve quality data

Data Normalization

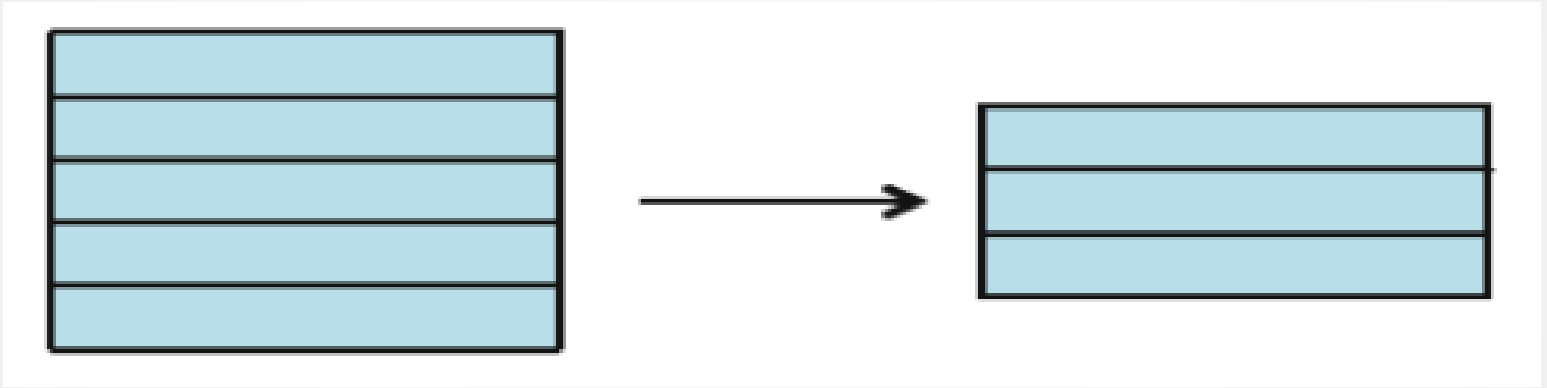
To express data in the same measurements units, scale or range.



Preprocessing to reduce the size of data

Instance Selection

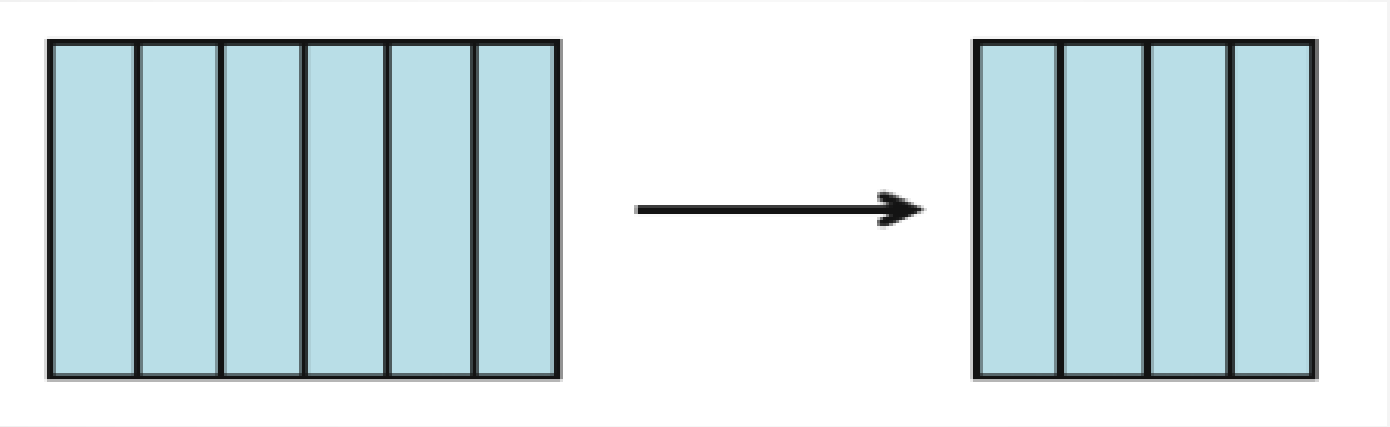
Sample the data records and only choose the representative subset from the data.



Preprocessing to reduce the size of data

Feature Selection

the reduction of the data set by removing irrelevant or redundant features.



Preprocessing to reduce the size of data

Discretization

Transforms quantitative data into qualitative data, that is, numerical attributes into nominal attributes with a finite number of intervals.