

Data Preparation Overview

After this video you will be able to..

- Articulate the importance of data preparation
- Define the objectives of data preparation
- List some activities in preparing data

Preparing Data

Goal: Create data for analysis



Clean



Format

- Select features to use
- Transform data

Data Cleaning



- **Data quality issues**
 - Missing values
 - Duplicate data
 - Inconsistent data
 - Noise
 - Outliers

Addressing Data Quality Issues

- Some techniques:
 - Remove data with missing values
 - Merge duplicate records
 - Generate best estimate for invalid values

Cleaning Data

Data Cleaning



Data Cleansing

Getting Data in Shape

Data Munging

Data Wrangling

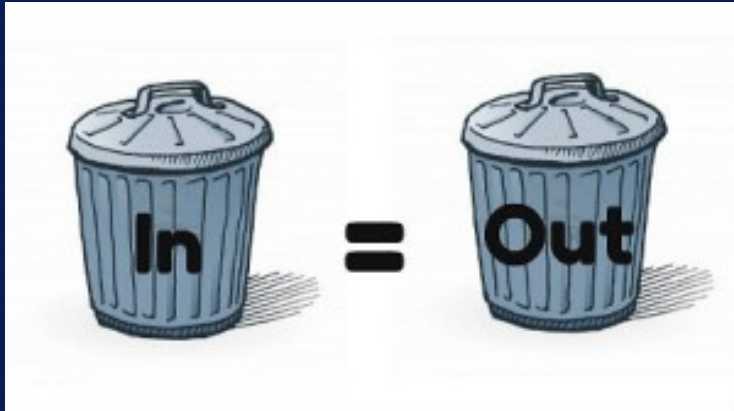


Data Preprocessing

Data Wrangling

- Feature selection
 - Combing features
 - Adding/Removing features
- Feature transformation
 - Scaling
 - Dimensionality reduction

Always Remember!



Data preparation is very important for meaningful analysis.

Garbage in = Garbage out