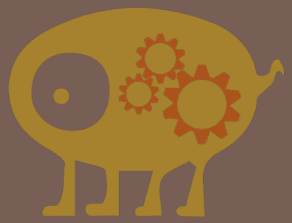

Data Mining
資料探勘

Project 2

Hung-Yu Kao, Fall 2023

Classification

- **Goal**
  - Understand what classification systems do and the difference between the real behavior of the classification model and observed data
- Description
  - Construct a classification model to observe the difference between real 'right' data and modeled data

- **Due:** 11/14 9am

Flow

- **Step 1**: Design a set of rules to classify data, e.g., classify students with good performance.
  - You should design **k** features/attributes for your problems first.
  - Use 'absolutely right' rules to generate your positive and negative data (the number of data = **M**)
- **Step 2**: Use the data generated in Step 1 to construct your classification model
  - Decision tree is basic requirement, you can add more classification models.
- **Step 3**: Compare the rules in the decision tree from Step 2 and the rules you used to generate your 'right' data
- **Step 4**: Discuss anything you can

Example: “How to select a good apple ?"

- Your “absolute right” rule (R)
  - Color: dark red
  - Knock voice: sharp
  - Head color: green
  - Weight: medium (hidden)
- Use (R) to generate your data
  - Add more attributes (e.g, price, #dark_spot, …, 20+ is better)
- Use classifiers to classify your data
  - Decision tree
  - Naïve Bayes
  - …