



Data Mining

資料探勘

Project 1

Hung-Yu Kao, Fall 2019

Association Analysis

2

- Dataset1: Select from kaggle.com / UCI
- Dataset2: Use IBM Quest Synthetic Data Generator
 - <https://sourceforge.net/projects/ibmquestdatagen/>
 - Generate different datasets
- Implement Apriori Algorithm and apply on these datasets
 - Hash? Tree? (optional)
 - FP-growth
- Use association analysis tools (e.g., WEKA) to generate association rules from the datasets you generate
- Testing data release soon
- Compare your results

- Due: 10/22 9am

Find and answer

3

- What are rules with
 - ▣ High support, high confidence ?
 - ▣ High support, low confidence ?
 - ▣ Low support, low confidence ?
 - ▣ Low support, high confidence ?