CAMT Data Mining: A Case Study

Manawin Songkroh
College of Arts, Media and Technology

manawin@live.com

Outline

- Data Mining Definition
- CAMT's Profile
- Literature Review
- Purpose of the study
- Data used
- Proposed Tool: Rapid Miner

Why Data Mining?

- is good for ***information technology*** era
- saves ***time and cost*** (Fayyad et al., 1996)
- has been accepted by ***organizations in many fields (NASA, US Treasury Network, Banking Industry, Retailer, Medical, Bioinformatics....)***

Data Mining in the real world

- Marketing: market-basket analysis
- Investment: Managing portfolio (LBS Capital Management) http://www.lbs.com/lbs_tech.htm
- Fraud Detection: PRISM System for Credit Card Fraud, FAIS System for detecting money laundering activities.

DM & KDD

- “KDD refers to the overall process of discovering useful knowledge from data and data mining refers to a particular step in this process.” (Fayyad et. al., 1996, p.39)
- The additional steps in KDD process are data preparation, data selection, data cleaning and etc.

Literature Review

- Hsieh (2004) uses an integrated data mining and behavioral scoring model to manage existing credit card customer in a bank.

CAMT Profile

- over 1000 students, founded in 2004
- 125 staffs (75 teaching and 50 supporting)
- multidisciplinary college: MMIT, Animation, Software Engineering, KM (PHD)

Current Problems in CRM

- Low number of applicants in Software Engineering
- High dropout and expel rate in MMIT

Purpose

- to cluster students for better CRM plan
- to build the predictive model for tentative drop-out students

Stats

Personnel/Students	amount
Lecturer	75
Supporting Staff	25
Temporary STaff	20
Undergraduate	700
Master	60
PHD	100

RapidMiner

- http://rapid-i.com/content/view/26/84/
- Window, and other systems with Java
- RapidMiner 4.6
- Open-Source from German Firm

Data Used

- CAMT Student Records from Registration Office of Chiang Mai University.

Data File-

- .dbf form

Project (Study) Management)

- Data Acquisition
- Data Preparation & Understanding
- Data Experimentation
- Data Validation
- Writing Paper

Next Presentation

- Detailed steps in accomplishing the paper
- Results from Data Preparation and Understanding & Model Selection
- Q & A