Introduction to

Machine Learning and Data Mining

Capstone Project examples

Khoat Than

School of Information and Communication Technology

Hanoi University of Science and Technology

2020

Prediction of apps' rating

- **Problem**: study to build a system that can make accurate prediction about the average rating for an app, using some descriptions about the app.
- **Input**: some descriptions about the app
- **Output**: average rating from users for a given app
- **Method** to be used: Ridge regression or neural network
- **Dataset**: a set of apps and their descriptions in terms of text, each app has a rating collected from App Store.

Prediction of hotels' rating

- **Problem**: study to build a system that can make accurate prediction about the rating for a hotel when it has just been launched, using some descriptions about that hotel. The rating belongs to {1*, 2*, 3*, 4*, 5*}.
- **Input**: some descriptions about the hotel
- **Output**: rating for that hotel
- **Method** to be used: Random Forest
- **Dataset**: a set of hotels and their descriptions. The data will be collected from Agoda.com.

Users' preference in music

- **Problem**: analyze the preference/interest of online users about music, over demographic/time/sex, …
- **Input**: set of songs/MV, and a set of users and their interactions with the songs/MV
- **Output**: preference, new conclusion/finding, visualization, …
- **Method** to be used: clustering by K-means, classification with Random forest, …
- **Dataset**: set of songs/MV, and a set of users and their interactions with the songs/MV. The data will be collected from youtube.com.

Comparison of different methods

- **Problem**: do an extensive evaluation about the performance of different ML&DM methods for solving a real-life problem
- **Dataset**: a dataset from that real-life problem
- **Output**: new conclusion/finding, recommendation, …
- **How to do?**
  - Select at least 3 methods/models to be evaluated.
  - Implement or use some existing codes of those methods.
  - Do extensive experiments to compare those methods, using different measures (e.g., accuracy, time, memory, …) and a good evaluation strategy. The comparison might also be in different scenarios. Use tables, figures, … to summarize the results.
  - Analyze the results, compare the performance, make conclusions.