Homework #1

Participating in a Kaggle Competition

Kaggle

- You must register to the Kaggle (www.kaggle.com)
  - An online community of data scientists and machine learning practitioners
  - Allows users to
    - Find and publish data sets
    - Explore and build models in a web-based data-science environment
    - Work with other data scientists and machine learning engineers
    - Enter competitions to solve data science challenges

- You must participate the following competition on Kaggle
  - "고객이 호텔 예약을 취소할 것인지를 예측하는 프로젝트"
    - URL: https://www.kaggle.com/competitions/sm-datamining-hw1/
    - Invitation: https://www.kaggle.com/t/3adfa56b127146949593421a0e5fc28e

To Do (1/2)

- Implement a customer predictor that performs as follows:

  ① Read the training data (`train.csv`)

  ② Preprocess the training data

  ③ Build any of the following classifiers you want
    - Decision tree, k-NN, naïve Bayes, logistic regression, artificial neural network

  ④ Read the test data (`test.csv`) and make predictions

  ⑤ Output a prediction file for the test data (e.g., `submission.csv`)

To Do (2/2)

- You ***must*** use ***scikit-learn*** to build a classifier
  - https://scikit-learn.org/stable/supervised_learning.html

- You ***must*** use the following classifiers provided by scikit-learn
  - Decision trees
  - k-Nearest neighbors
  - Naïve Bayes
  - Logistic regression
  - Artificial neural networks

Submission

① Submit your prediction file to the leaderboard on Kaggle

② Submit the following files to Snowboard
   – Source file (`.py`)
   – Report (`.docs`, `.hwpx`, `.pdf`, …)
   – SnowBoard → 데이터마이닝및분석 → 9주차 → Homework #1

▪ Due: **2023.5.18 (Thu) 23:59**
   – 1-day delay: 80% credit
   – 2-day delay or more: 0% credit

Homework Support

- **[Off-line]** If you need help from me, I recommend you to request an off-line meeting with me
  - The number of students ≥ 1
  - Schedule a meeting time with me

- **[On-line]** Slack or emails
  - Of course, you can use the Slack for Q&A, discussion, etc.

- **[Teaching Assistant]** You can get help from the T.A.
  - Eunjo Jang (wkddmswh99@sookmyung.ac.kr)
  - Seongyeon Yang (yeon55@sookmyung.ac.kr)