Improving Matches for a Dating Site

Learning Outcomes

- Upon successful completion of this lab, you will have demonstrated the abilities to:
  - Build and compare different models for a particular problem
  - Make decisions about different aspects of the data mining pipeline

Note: In this lab, you have the freedom of choosing different models (**at least 3 models**), evaluation methodologies (e.g., cross-validation), performance metrics, and you can perform model selection, before evaluation the model on the test data set. Try your best!

Instructions:

You are working as a data mining scientist, and now you have a case from a dating site. The site has been complained for its terrible recommendations, which simply match the locations and ages. Now, it intends to improve its recommendation system, using data mining techniques. After a survey on the recommendation, the site realizes that a specific customer often categorizes other customers into three types:

1. People he/she didn't like;
2. People he/she liked in small doses;
3. People he/she liked in large doses.

It also realizes that the following features are highly related to the choices of customers:

1. Number of frequent flyer miles earned per year;
2. Percentage of time spent playing video games;
3. Liters of ice cream consumed per week.

Finally, you accept the case and the site also passes you the data from its survey.

NOTE: In your report, you should analyze the cases and provide at least three solutions, and validate your solutions.

The data (datingData_training/test.txt) contains four columns. Each row contains the information of a specific customer. The first column denotes the number of frequent flyer miles earned per year; the second column indicates the percentage of time spent playing video games; the third column is the liters of ice cream consumed per week. The fourth column indicates the types of people, which is labeled by some other customers.

Train/Test Data Sets:

- **Train (https://learn.ontariotechu.ca/courses/19275/files/2375396?wrap=1)** ⬇ (https://learn.ontariotechu.ca/courses/19275/files/2375396/download?download_frd=1)
- **Test (https://learn.ontariotechu.ca/courses/19275/files/2375397?wrap=1)** ⬇ (https://learn.ontariotechu.ca/courses/19275/files/2375397/download?download_frd=1)

Report:

1. Your report should have a cover letter including the group member names
2. Organize all your *diagrams* and *interpretations* in your lab report *(PDF format)*
3. Include your code and report in a folder (you can zip the folder) and submit it