Università
della
Svizzera
italiana

**Facoltà
di scienze
informatiche**

Data Analytics

Academic Year 2024-25

Course Assignment N.16: Dating Profile

Prof. Fabio Crestani & Dr Ana-Maria Bucur

TAs: Lili Lu & Enrico Verdolotti

For this assignment you will work in pairs to carry out simple tasks of data analysis given a specific dataset. The goal of this assignment is to use Python and complementary libraries on a given dataset in order to *explore* and *analyze* the given data and *draw conclusions*.

Description

This dataset includes user ratings for different dating profiles taken from a dating website as well as some information for the users. The goal is to recommend dating profiles to some of the users.

Your tasks are to:

- Explore and describe the data (*i.e.*, standard descriptive statistics, visualize the variables with different graphs, draw distributions and histograms of variables, are there outliers? Any interesting observation? Any correlations? Etc.)
- Pre-process the data (*i.e.*, handle and fill unknowns if there are any, etc.)
- Build two recommendation systems for profile rating for the users in the test sample (consider rating as unknown– remove it first) ○ The real value of the variable rating is given to be able to evaluate your result
- Evaluate and compare the performance of the models

Submission procedure and evaluation

You should produce a report of your work and its evaluation along with the source code. It will be a concise explanation of how you tackled the different tasks, the reasons of your choices, successive conclusions, plots you produced, results of the decisions and their accuracy, *etc*.

Use Jupyter Notebook to produce results of the commands in a single .ipynb file. For more information check: https://jupyter.org/documentation.

The report (max 5 pages) and the code of the project need to be submitted via iCorsi.

Please, upload all the required items in a single file and name it following the structure: **noProject_FirstnameLastname.[zip|tar.gz|7z]**.

The dataset regarding this project can be downloaded from: Dataset.