CSCI - 4146 - 6409 - The Process of Data Science - Fall 2022

Assignment 2

The submission must be done through Brightspace. Due date and time as shown on Brightspace under Assignments.

- To prepare your assignment solution, use the assignment template notebook available on Brightspace.
- The detailed requirements for your writing and code can be found in the evaluation rubric document on Brightspace.
- Questions will be marked individually with a letter grade. Their weights are shown in parentheses after the question.
- Assignments can be done by a pair of students or individually. If the submission is by a pair of students, only one of the students should submit the assignment on Brightspace.
- We will use plagiarism tools to detect any type of cheating and copying (your code and PDF).
- Your submission is a single Jupyter notebook and a PDF (With the compiled results generated by your Jupyter notebook). The PDF is the printout of the final notebook with all the code and results. File names should be:
  - **A2-<your_name1>-<your_name2>.ipynb**
  - **A2-<your_name1>-<your_name2>.pdf**
- **Forgetting to submit both files results in 0 markings for both students.**

Note:

In this assignment, we will continue to use the same dataset from Assignment 1. As you already have the data exploration ready, you will apply the predictive algorithm to address the business problem you formulated in Assignment 1. Link to the dataset: https://www.kaggle.com/stefanoleone992/tripadvisor-european-restaurants

Build a model to forecasts the average rating of a restaurant

1. [0.1] Explain what is the task you're solving (e.g., supervised / unsupervised, classification / regression / clustering or similarity matching, etc)
2. [0.1] Select the evaluation metric. Justify your choice.
3. [0.2] Use a feature selection method to select the features to build a model.
4. [0.6] Build a model
   1. Motivate the choice of the model. Explain how your data satisfies the model's requirements.
   2. Perform hyperparameter tuning if applicable.
   3. Train and evaluate your model
   4. Plot learning curve and analyze it
5. [0.1] Analyze the model's performance according to the evaluation metric. Can the model be used to solve the business problem?