Mini Project: predicting beer rating

Background: The rating of beer is impacted by various factors, some of which come from beer (e.g., ABV), some of which come from user (e.g., age, gender). In this project, we want to build an ML model that can predict beer rating based on relevant features. This is important for various applications such as personalized beer recommendation.

Method: Fortunately, when speaking about prediction, we have a powerful tool called machine learning. Using machine learning, we can train a predictive model. You can use the Scikit learn library for doing this. Note that Sk-learn library contains many useful internal functions such as accuracy_score(), train_test_split(), etc.

Training data and labels: For every machine learning problem, we need to identify three things: what is input feature, what is output label, and what should be the ML method?
Input feature: A numerical vector, e.g., if you want to use [ABV, age], then the vector can be like [7.7, 1307459047]. Feel free to choose your own features!
Output label: Beer rating, e.g., [3.5].
ML method: Your choice! While we only studied limited ML methods now, I would not limit your choices here. You are free to use any methods. (And change ML method is easy, just change one line of code; of course, don't forget to include the proper package).
Beer rating dataset is here: www.ece.villanova.edu/~xjiao/course/ECE5400/dataset/beer_50000.json.

A glance at the training data:
Please use 80% of the data in the dataset as the training data, and 20% data as the testing dataset. An easiest way to do this is to use train_test_split() function. This is very important! If you have overlap in the train and test dataset, this is considered "cheating". Since the split is random, you can repeat your experiment multiple times and you may get different results which is fine. You can report the highest you get.

Evaluation:
Please use MAPE as the evaluation metric for this.

Your delivery:

1. Your source code: **Put a link** of your Google Colab Python code (remember the **sharing** setting so a 3-party can view) in the report so the TA can easily click and run your code.
2. A **1-page report**, including at least how you design your classifier (feature, label, algorithm), what is your final MAPE. Please clearly show the **name** of your team members.

Grading (50 points):

1. Team: Form your own team of 2-3 people. Each one gets same grade.
2. Your program quality (executable/readable): 30 points.
3. Your report quality (clear/ necessary components): 20 points.
4. Send your report (with your code link) to xun.jiao@villanova.edu and dma2@villanova.edu, with **subject title** ECE 5400 HW3 Mini Project. (**cc all of your teammates**).

Deadline:

1. By Mar.7, 3:00pm.