[Glassdoor](https://en.wikipedia.org/wiki/Glassdoor) introduced a new diversity rating for each review in 2020. I want you to predict this diversity rating out-of-sample using machine learning based other review attributes, such as overall rating and textual decriptions, among others.

* **Input:** this folder includes two .db file, one containing the training data and another containing the testing data. For the testing data, I hide the actual diversity rating for performance evaluation later.
* **Deliverable**: pleas**e write a Python code in the Code folder to produce the diversity rating prediction for reviews in the testing data. The output should be named ‘rating\_diversity\_OOS.csv” and contain two columns: id (the review id) and rating\_diversity\_OOS (the predicted diversity rating).**
  + Note: make sure your code print your model’s in-sample accuracy.

**Please make sure that I can simply copy paste your code into my own Code folder in my computer (like below) and run it to get the output file myself.**

Graphical user interface

Description automatically generated with medium confidence

I suggest you play a bit with the data and then tell me if you think you can do this task well (e.g., how accurate is your model in a validation sample?).

Thank you,