Our goal is to classify the students based on their typical study patterns. Classifying students based on their typical study patterns involves using clustering algorithms to group students who exhibit similar behavior in terms of their quiz and homework attempts. And the basic idea shoul be like this:

1. Feature Selection:

Identify key features that represent a student's study pattern. For example:

- Total attempts in quizzes.

- Total attempts in homework assignments.

- Average attempts per quiz.

- Average attempts per homework.

- Success rate in quizzes.

- Success rate in homework.

2. Data Preprocessing:

- Ensure that the selected features are appropriately scaled.

- Handle any missing data if present.

- Normalize or standardize the features if necessary.

3. Choosing a Clustering Algorithm:

Select a clustering algorithm suitable for your dataset.

4. Determine the Number of Clusters (K):

Use methods such as the Elbow Method or the Silhouette Score to determine the optimal number of clusters for your dataset.

5. Train the Clustering Model:

Use the selected algorithm and the determined number of clusters to train the model on the prepared dataset.

6. Assign Clusters to Students:

Assign each student to a cluster based on the clustering results.

7. Analyze Clusters:

- Visualize the clusters using scatter plots or other suitable visualizations.

- Examine the characteristics of each cluster to understand the study patterns it represents.

8. Interpretation:

Interpret the clusters in terms of study patterns.

9. Evaluate and Refine:

- Assess the quality of the clusters based on domain knowledge and the context of the study patterns.

- If needed, refine the features or consider different clustering algorithms.

10. Use the Clusters:

Utilize the identified clusters to classify students based on their typical study patterns. This information can be valuable for understanding different learning behaviors and tailoring interventions accordingly.

Review on the submitted code file:

1. The categorize\_student function you wrote categorises students according to the values of num\_homework, num\_quiz and Score. In the process of categorising the two categories of average and unmotivated students, you have only excluded the first two, and it is not sufficient to differentiate between the latter two by using the Score.
2. In addition, in order to be able to cover all the students, it is not prudent enough to use the if-else, and the total number of students in the four categories you came up with are duplicated by the way you look at the final data.
3. The classification criteria should be able to adequately reflect students' academic performance and engagement, and consider whether there are other features that can be used in the training of the model to improve the accuracy of the classification.
4. Other classifiers can be tried, such as support vector machines.
5. You've printed the accuracy of the model, but you can also look at other metrics such as confusion matrix, precision, recall, etc. to get a more comprehensive picture of the model's performance.

Summary:

In the period from the 3rd to the 12th of December, not much progress was made on the classification problem, the FEATURES were not clear enough, the criteria for quantifying the FEATURES were debatable, and there was little to no parameter tuning, but we are about to show it on the 20th, and there are still 6 days left, so it is not necessary to put too much effort into presenting the results of the model, and it is sufficient to be able to accurately show what the model is like.