

# Intro to Machine Learning

## HW07

### Part 1 Interpretations:

The model's accuracy of 65% is just above the proportion of the majority class in the dataset, which is 56% for Category 1, indicating potential class imbalance effects. The presence of 13% for Category 2 and 31% for Category 3 further highlights the imbalance.

Below are the some of the reasons why Random Forest might not perform well for the initial features.

- Small dataset size may cause the model to learn poorly and show inconsistent performance.
- Outliers or irrelevant data points might be misleading the learning process.
- Chosen features may lack strong relevance to predict the target effectively.
- Standard scaling was applied, yet alternative scaling or feature work might improve results.
- The Random Forest model might not be fitting the data correctly, needing hyperparameter adjustments.
- Imbalanced classes could be causing a bias towards the more prevalent class in the dataset.
- The randomness in the train-test split may create a non-representative test set, affecting model evaluation.

### Part 2 Interpretations:

- The model performed well with an accuracy score of 1.
- It was able to predict the unseen data correctly to its categories.
- This was due to more information stored in the new features of the dataset.
- The new features included the feature information of its own and information of the next feature.
- The new features capturing growth rate over time played a crucial role in achieving a perfect accuracy score on the test data.