

## [Bonus] Buffalo Trees Classification

Welcome to the second week of the semester! We hope you enjoyed our Meet & Greet Activity and claimed your bonus. We have another activity for you to complete this week.



### TASK

In this challenge, we will be performing classification using the real-world Tree Inventory dataset ([link](#)). This dataset comprises of all street trees within the City of Buffalo.

- The target variable is **Common Name**
- You need to build a classifier, that will predict at least 5 classes, e.g., Maple Norway, Linden .... As part of this task, you do not have to use the whole dataset.
- Input: you need to use at least 5 features as input. e.g., you can use all the 16 columns or drop some.
- You can use any in-build machine learning tools to do this classification or build something from scratch, e.g., Logistic Regression, SVM, Decision tree. Submission that uses neural network or a deep learning framework will not be used for the bonus points.
- Main motivation of this task is to explore various ML tools to solve a real-world task.

### Suggestions:

- Preprocess your dataset, e.g., drop the vacant entries and normalize the dataset.
- You can use a combination of various ML tools in one pipeline
- Check the accuracy using the following snippet:

```
scores = model.evaluate(X_test, Y_test, verbose=0)
print('\n Test accuracy is %.2f%%' % (scores[1]*100))
```

## SUBMISSION

- Share the following details as a comment to the piazza post:

**Target:** # classes to predict

**Input:** # of features used

**Model used:** (e.g. logistic regression)

**Test accuracy:** (e.g. 74.59%)

- Submit a Jupyter notebook or py file with your code and saved output at UBlerns (UBlerns > Challenge)
- Save and submit weights for your trained model.
- Each student can make a max of 5 submissions for this challenge. Submission with the highest accuracy will be considered.
- Only top 5 submissions from different students with the highest accuracy will be considered towards the bonus points.
- In your Jupyter Notebook or py file include all the references used, e.g., scikit-learn library.

## Evaluation

The submission with the highest test accuracy will get a bonus point and the code will be presented as a presentation during one of our OH. Time complexity and overall performance of the model will be considered as well.

**Due date:** Monday, Feb 13 at 11:59 pm

### Notes:

- If the file and/or piazza comment is submitted after the due date, e.g., 00:00am, it will not be considered for the bonus.
- Only submissions with a publicly shared accuracy on the piazza post will be considered.
- No late days can be applied for bonus-based tasks.
- This challenge must be done individually. e.g., submissions from different students with a similar structure or similar setup or same answers will not be considered. This will affect the initial submission and all other similar submissions.