You will have to form a group of 3 to 4 students.

Choose a dataset from the UC Irvine Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php) with at least 5000 instances and 20 attributes for classification or regression. Compare how the different approaches seen in class perform on this dataset to predict accurately the classes or the values of the unlabeled data. You should determine what are the best hyper-parameters for each approach you are using. You could use any Python libraries.

<https://archive.ics.uci.edu/dataset/73/mushroom>

First Project: Machine Learning Analysis

This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family (pp. 500-525).

Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended.

This latter class was combined with the poisonous one.

The Guide clearly states that there is no simple rule for determining the edibility of a mushroom; no rule like ``leaflets three, let it be'' for Poisonous Oak and Ivy.

Abstract

Have you ever been walking down the streets of Windsor and been plagued by the question of determining the edibility of a mushroom. [0] While as a forager, mushroom hunting, is an exhilarating outdoor adventure that transforms you into an outdoor adventurer. This research endeavor delves into the mixture of machine learning methods with the natural world. The aim is to classify different species of gilled mushrooms in the Agaricus and Lepiota Family. The main goal of the project aims seeks to provide a reliable method for classifying the edibility of gilled mushrooms.

1. Introduction
2. Dataset
3. Data Preprocessing
4. Model Selection
5. Model Training and Evaluation
6. Hyper-Parameter Tuning
7. Comparison of Results
8. Conclusion and Findings
9. Description of Student Participation

References

[0] – fun guyz

dataset

<https://archive.ics.uci.edu/dataset/73/mushroom>