CAP4770 Project Proposal

Group 5

Rachel Baryol

Luis Borruel

Heather Burke

Devon Chessher

Problem Statement

Develop a predictive classification model that accurately determines whether a given mushroom is edible or poisonous based on a set of observed physical and environmental attributes.

Data Set

- The data set is from the UC Irvine Machine Learning Repository describing mushrooms.
- The data includes descriptions of samples corresponding to 23 species.
- Each species is identified as edible, poisonous, or unknown edibility.
- The data itself comprises 8,124 instances with 22 categorical features.
- https://archive.ics.uci.edu/dataset/73/mushroom

Approach and Tools

- decision tree model
- python
- pandas
- numpy
- matplotlib
- seaborn
- scikit-learn
- xgboost
- shap
- notebook

Expected Results & Evaluation Metrics

- Expected Results:
  - A trained classification model that achieves at least 95% accuracy in distinguishing between edible and poisonous mushrooms using their categorical attributes.
- Evaluation Metrics:
  - Accuracy, Precision, Recall, and F1-score, with emphasis on Recall for the poisonous class to avoid false negatives.
  - Random Forest Regression, which builds an ensemble of decision trees and aggregates their predictions to improve accuracy and reduce overfitting.

Thank You