**Team 2 Project Team Members:** Aaron A. Gauthier & Pedro Urea Rodriguez

**Project Title:** Classification of Mushrooms from the Agaricus and Lepiota Family

**Problem Definition:** Many people do not know how to identify an edible mushroom versus a poisonous mushroom. Using classification algorithms, we hope we can utilize a model with a high degree of confidence in order to distinguish between mushrooms that are edible versus poisonous for two families of mushrooms (namely Agaricus and Lepiota Family). Various guides clearly state that there is no simple rule for determining the edibility of a mushroom. We hope to prove this wrong through the use of data and classification models/algorithms.

**Problem Motivation:** Realized many people in the USA do not know the difference between an edible mushroom and a poisonous mushroom. We hope that through this project we can help educate people on the top significant features that determine if a mushroom is edible or poisonous. We hope this will help save lives and avoid tragedy, especially with children who are curious.

**Proposed Method/Process/Classification Algorithms:**

**Method**

1. Separate X and y variables
2. Use label encoder to replace text data (one hot encoding)
3. Design multicolumn one hot encoder
4. Predict results
5. Review optimal values
6. Finding traits of poisonous mushroom

**Process**

* Data Acquisition – University of California Irvine Machine Learning Data Repository
* Reading in the data
* Partitioning the data into training and testing data
* Extracting features
* Constructing our model
* Measuring the accuracy of the model

**Classification Algorithms**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Language(s)** | **Package(s)** | **Comments** |
| Logistic Regression | Python | Pandas, Numpy, Sklearn, Seaborn, Matplotlib | Initial review shows our model accuracy to be 1 – will explore why this is? We suspect we have an error |
| Random Forest | Python | Pandas, Numpy, Sklearn, Seaborn, Matplotlib | Curious about this Model in the context of this problem |
| K Nearest Neighbor | Python | Pandas, Numpy, Sklearn, Seaborn, Matplotlib | Curious about this Model in the context of this problem |
| Naïve Bayes | Python | Pandas, Numpy, Sklearn, Seaborn, Matplotlib | Curious about this Model in the context of this problem |

**Link to the data:** Main Page for Mushroom Classification Dataset- <https://archive.ics.uci.edu/ml/datasets/mushroom>

Actual Dataset link for Mushroom Classification- <https://archive.ics.uci.edu/ml/machine-learning-databases/mushroom/>

**Responsibility of Team Members:** Duties of each team member will include coding, review of coding, strategy, execution and write ups. Duties of each team member are a shared responsibility in all facets of the project. Team members will utilize Skype, Slack, GitHub, Dropbox, email and in person consultations.