JimH 7.1 Capstone 2 Proposal

Problem Identification The recent resurgence of individuals interested in the adventure of wild mushroom hunting, commonly called 'Shrooming', brings with it the danger of harvesting mushrooms that are poisonous. Some mushrooms have a simple process to identify while other mushrooms have several different factors to determine if it is edible or poisonous and there can be two or more that look very similar but only have a couple identifiers that help the individual determine if it is edible, this produces a major challenge as the end results are literally life or death! The predictive model that we will try to develop will enable the 'Shroomer' to recognize the edible mushrooms from the poisonous mushrooms which will facilitate the 'Shroomer' to have more confidence but encourage new individuals to try out the activity. In a way we are demonstrating the ability to use this model in a retention campaign that will attract new participants but also continue to retain those pre-existing and loyal customers.

Hypothesis: There is a wide and varied selection of mushrooms that are both edible and poisonous. With so many different mushrooms available and some can be encountered in relatively close proximity and during the same season, it can be a relatively dangerous outdoor adventure. It can be easy to identify some edible mushrooms that are unique and don't look like any other poisonous ones, while other edible mushrooms have a poisonous twin that has only a few distinguishing factors. We hope to leverage a machine learning model to enable a test that can be trained to identify if a given mushroom is edible or poisonous based on its physical characteristics.

By leveraging machine learning models, the confident identification of the mushroom and determining if it is edible or poisonous. We are going to and give the 'Shroomer' a solid method to stay safe on their outdoor adventure.

Data Source: We will focus on a data set that includes 23 different species of gilled mushrooms that will enable the model to be trained in a way that will provide a foundation for larger data sets. The data set has been curated by UC Irvine Machine Learning Repository, and is available from kaggle.com.

Potential client and deliverable: The hypothetical client could be a company looking to make an app that a 'Shroomer' could use to assist in determining if a mushroom found in the wild is edible or poisonous. The deliverable would include a report detailing the model and training involved and a few examples of how the model could be deployed.