Steven Alvarado, Noor Ashour, Khaiber Amin, Dhruv Arora

ECS 171 Group Project Proposal

Dataset https://archive.ics.uci.edu/dataset/73/mushroom

Problem The problem that our group wanted to address centers around building a classification model that can accurately predict whether or not a species of mushroom is safe for human consumption. From the UCI Machine Learning Repository, we have determined that the Mushroom dataset contains a sufficient number of features to train and validate this model effectively. Within it, we will utilize the physical characteristics that are provided from a sample of 23 mushroom species to analyze any relationships between fruiting body features and human toxicity/edibility. Our goal in developing this model is to help automate the identification of poisonous mushrooms and reduce the risk of misclassification and potential harm. Therefore, our plan is to use machine learning to classify whether a mushroom is edible or not based on a description of its attributes, such as cap shape, gill size, odor, etc.

Roadmap We plan to work in phases by first planning and organizing the dataset within a Jupyter notebook to outline our thought process as we interpret and assemble information. Then, background research will be conducted to study and review the literature behind these mushroom species and any relevant traits that could relate to being poisonous or inedible. With 22 features to work with, thorough exploratory data analysis (EDA) will be necessary to fully understand their meaning within the dataset and identify important patterns—or lack thereof. Once sufficient knowledge has been gathered, we will start developing and testing for precise and effective models. Graphs and plots will be created to visualize the trends in the data to support our findings. Finally, we will showcase our conclusions by providing our source code through GitHub, creating a final demo with an html interface and recorded video, and writing a detailed report paper of the outcomes.