Project Proposal

Machine Learning I

Aaron A. Gauthier and Pedro Uria Rodriguez

October 15, 2018

Project Title

Classification of Mushrooms from the Agaricus and Lepiota Family

Problem Definition

Various guides clearly state that there is no simple rule for determining the edibility of a mushroom. Using data on different features of mushrooms and classification algorithms, we plan to train a model with a high degree of confidence in order to distinguish between edible and poisonous mushrooms for two particular families (namely Agaricus and Lepiota Family).

Motivation

We realized that many people in the USA do not know the difference between an edible mushroom and a poisonous one. We hope that through this project we can help educate people on the top significant features that determine whether a mushroom is safe for ingestion or not. We hope this will help save lives and avoid tragedy, especially with children who are curious.

Method

Data Preprocessing $\longrightarrow$ Model Learning $\longrightarrow$ Model Testing, Understanding and Refining $\longrightarrow$ Comparison between different models $\longrightarrow$ Finding common poisonous traits.

The models will include, but are not be limited to: Logistic Regression, Random Forests, K-Nearest Neighbors and Naive Bayes, all of which are present in the *scikit-learn* package for *Python*. We will be using this library, along with *pandas*, *numpy*, *seaborn*, *matplotlib* and other potential ones. We will write our code on Jupyter Notebooks, and use Skype, Slack, GitHub, Dropbox, email and in-person consultations.

Link to the Data

UCI Machine Learning Repository.

Mushroom Data and MetaData.

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.