a) Motivation/background – this should briefly sketch out why you chose this topic and should preferably mention academic work that has taken place within the area.

- I used to go walking in the woods when I was younger
- I would always see different types of mushrooms which I would wonder about
- It’s important to find out about the edibility of some mushrooms since they can be
- Mushrooms are specifically the ‘fruits’ of certain species of fungi that are part of the underground network of mycelium

b) A statement of the research question(s) that will be addressed.

- I mainly want to find out about how edible specific types of mushrooms are
- After observing specific characteristics, I want to be able to find out if any given mushroom is edible or not

c) An overview of the data that you are going to use. This, in most cases, is predefined, but you should give a brief description of how you accessed the data, the variables they contain, and any initial limitations that you have identified with the data.

- I found the data on https://archive.ics.uci.edu/datasets
- It has many different sets of data and is generally quite reliable
- It contains many different ways of visibly defining a mushroom so that a specific type can be identified
- With the data set, there is a brief explanation of it, which tells me that there are some values missing from one of the features. This will have to be dealt with in R, but the most realistic option is to just remove any row with the missing data, assuming there is not too much.

d) A sketched overview of the methodology that you intend to apply to the data to be able to answer your research questions.

- I will use classification and certain machine models, such as GLM and trees, to make a classification system to help identify edible or non-edible mushrooms
- Ideal,ly it will allow someone to look at a mushroom, observe its features and by applying an algorithm that I will fin,d the person will be able to identify how likely it is that this specific mushroom is edible

e) The model or models that you are using should also be compared with the standard classical classification alternatives.

- Certain charts, graphs and tables for an EDA to get a good idea of how the data lines up and interacts
- GML functions will be useful to look at correlations across the data, and can also be used to find a percentage chance of a specific mushroom, with given characteristics, being edible
- Leave-one-out cross-validation will be useful as well as bootstrapping
- Fitting trees and, by extension, forests are very useful. They can be used on categorical data to sort through and find good predictors for a specific factor. Forests are also very robust and reliable within the data.