



Financial Crime – Data Science

Technical Exercise

Pre-Requisites

Please go to the following link <https://archive.ics.uci.edu/dataset/73/mushroom>, read the description of the data, then download the accompanying dataset. Take time to note the richness of the data, it will help you with the task of the exercise.

Summary

You have been provided with a dataset which represents 23 species of gilled mushrooms in the Agaricus and Lepiota Family. Each species is categorized as edible, poisonous. There are no simple rules for determining the edibility of a mushroom. Your task is to use this data to understand the differences in characteristics between edible and poisonous mushrooms and write a short guide to help someone out walking foraging for mushrooms decide if a mushroom they come across is likely to be edible or poisonous.

Assessment Requirements

- Code which ingests the data and conducts analysis whose output would inform the forager on the key indicators of a mushroom on if it's edible or poisonous.
 - You should write this code in a form which would allow it to be run on more recent data in the future. i.e. **you should avoid a Jupyter notebook** that isn't written in a form general enough to be applied to new data received in the future.
 - Please upload this code to github, or similar website – alternatively, please return this code as a *.txt file.
- A short doc/presentation which presents the insights to the forager and recommends actions that they can take to improve their chance of surviving their walk!

FAQ's

How will I be assessed?

You be assessed on the following aspects:

- The quality of your code
- The quality of your data science decision making
- The clarity of communication

How long should I spend on this task?

You should aim to spend around 90 minutes on this task.

Do I have to use all the data?

No! Feel free to use a subset of the variables, or a sample of the data if this helps you complete the task in the recommended amount of time.