

**Microsoft: DAT210x Programming with Python for Data Science**

Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▶ 4. Transforming Data
- ▶ 5. Data Modeling
- ▼ **6. Data Modeling II**

**Lecture: SVC**

Quiz

**Lab: SVC**

Lab

**Lecture: Decision Trees**

Quiz

**Lab: Decision Trees**

Lab



## 6. Data Modeling II &gt; Lab: Decision Trees &gt; Assignment 5



Bookmark

**Lab Assignment 5**

Mycology is a branch of biology that generally deals with the study of fungi and mushrooms, and particularly their genetic and biochemical make-up, and their use to humans. Throughout history, fungi have been used for tinder, medicine, and food. For hundreds of years, specific mushrooms have been used as folk medicine in Russian, China, and Japan. Scientists elsewhere have also documented many medicinal uses of mushrooms as well, but not all mushrooms are beneficial--some are quite deadly.

## Dive Deeper



In this lab, you're going to use decision trees to peruse The Mushroom Data Set, drawn from the Audobon Society Field Guide to North American Mushrooms (1981). The data set details mushrooms described in terms of many physical characteristics, such as cap size and stalk length, along with a

classification of poisonous or edible.

As a standard disclaimer, if you eat a random mushroom you find, you are doing so at your own risk. While every effort has been made to ensure that the information contained with the data set is correct, please understand that no one associated with this course accepts any responsibility or liability for errors, omissions or representations, expressed or implied, contained therein, or that might arise from you mistakenly identifying a mushroom. Exercise due caution and just take this lab as informational purposes only.

1. First, visit the data set's page and read through it carefully. Understand what they're saying about missing value representations, and header names, and where the classification column is located. Peek through the data values in a spreadsheet program or text editor and get comfortable with it.
2. Load up the started code in Module6/**assignment5.py**.
3. A copy of the dataset is included in Module6/Datasets/**agaricus-lepiota.data**.
4. You're going to need to review the decision tree code in the SciKit-Learn section of the Decision Tree section. It contains a few calls in there necessary for the completion of the assignment. If you're unable to install graphiz, then complete the assignment by examining the attributes of your classifier.
5. Answer the following questions.

## Lab Question 1

(1/1 point)

What are the top two features you should consider when deciding if a mushroom is eadible or not?

☒ Odor, and Gill Size ✓

☐ Stalk Color Below Ring, and Gill Color

☐ Bruises?, and Cap Color

☐ Population, and Cap Shape

☐ Spore Print Color, and Veil Type

#### EXPLANATION

Once you build out the decision tree, the nodes closer to the top are the most important nodes.

*You have used 1 of 2 submissions*

## Lab Question 2

(1/1 point)

Please enter a numeric value (e.g. 0, 1, 10.5, etc) which correctly answers the question(s) below:

What was your decision tree's accuracy after testing it against the independent testing set? Enter it as printed in the lab

100.0

✓ Answer: 100.0

THIS DATASET WAS PRETTY EASY FOR THE TREE TO CLASSIFY. YOU SHOULD HAVE GOT A 100% SCORE.

*You have used 1 of 2 submissions*

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY  
OPENedX

