Project – Predictive Analysis using scikit-learn

Your assignment is to:

- Start with the mushroom data in the `pandas DataFrame` that you constructed in your "Assignment – Preprocessing Data with sci-kit learn."
- Use `scikit-learn` to determine which of the two predictor columns that you selected (odor and one other column of your choice) most accurately predicts whether or not a mushroom is poisonous. There is an additional challenge here—to use `scikit-learn`'s predictive classifiers, you'll want to convert each of your two (numeric categorical) predictor columns into a set of columns. See for one approach pandas `get_dummies()` method.
- Clearly state your conclusions along with any recommendations for further analysis.

This is by design a very open-ended assignment. You're encouraged to go through the resources in your Week 12 folder. In particular, if you understand the process used in the Kevin Markham videos on Machine Learning with scikit-learn to predict iris species from four predictor variables, you should be able to transfer your learnings to complete this task.

For this assignment only, open collaboration with your classmates is allowed and encouraged. You may ask or answer any related question on the Project 4 discussion forum.

Your deliverable is a Jupyter Notebook that performs these predictive analysis tasks.

If you chose a different dataset for the preliminary processing assignment, you may continue to use that dataset here.

You should post the Jupyter Notebook (.ipynb) file in your GitHub repository, and provide the appropriate URL to your GitHub repository in your assignment link. You should also have the original data file accessible through your code—for example, read directly from the UCI repository or stored in a GitHub repository.