CSCI4181_6802 Assignment 4: Getting Started

All instructions, code, and questions for this assignment will be provided in a Jupyter notebook. These are interactive notebooks that can interleave code with text and results, facilitating code sharing, exploration, and data analysis. The most straightforward way to run this assignment is in a browser, but there are also local installation instructions.

In a browser (recommended)

Navigate to the JupyterLab site (https://jupyter.org/try-jupyter/lab/). Then **drag and drop** the notebook "**AMR-prediction_assignment_Apr2026.ipynb**" and the data file "**FilteredEfaecium_5.csv**" files into the file tab on the left. Double clicking on the assignment will open it in the main frame of the webpage. The rest of the instructions for the assignment (and some code to check if the installation worked) can be found in this notebook.

In our tests we were able to run the notebook in the Chrome and Edge browsers. JupyterLab is theoretically meant to run in Firefox as well but it was taking infinite time to run, for some reasons. We have not tested Safari or any other browser.

Local installation

Alternatively, you can install software locally starting from https://jupyter.org/install.

Software Pre-requisites

Python 3.6 & the Jupyter Notebook (https://test-jupyter.readthedocs.io/en/latest/install.html) are required, with the seaborn, pandas, and scikit-learn libraries.

Windows/Mac

If you are using **Windows** or **Mac**, I highly recommend anaconda: https://www.anaconda.com/download/. It is easy to install and gives you all the bits you will need for these interactive tutorials. Anaconda installs Python, the Jupyter Notebook, and other commonly used packages for data science.

Linux

If you are using **Linux**, I recommend using anaconda/miniconda to set up an environment for Python 3.6, especially if your system interpreter is Python 2.

We require the Python library scikit-learn (http://scikit-learn.org/stable/), a machine learning library for the Python programming language. If using conda/miniconda, this can be installed with:

```
$ conda install scikit-learn
```

Or otherwise with pip using the command:

```
$ pip install scikit-learn
```

We will manage some of the data using the Pandas package. It is included with anaconda but can be installed with **$ conda install pandas** or **$ pip install pandas**.
Finally, we need Seaborn, which is a Python data visualization library based on matplotlib. The library can be installed with **$ conda install seaborn** or **$ pip install seaborn.**