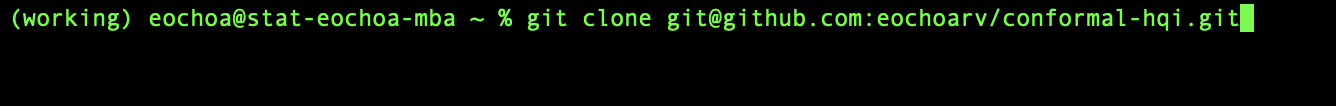
User Guide

This user guide is intended to demonstrate the use of the open-access code on [GitHub](https://github.com/eochoarv/conformal-hqi/tree/main).

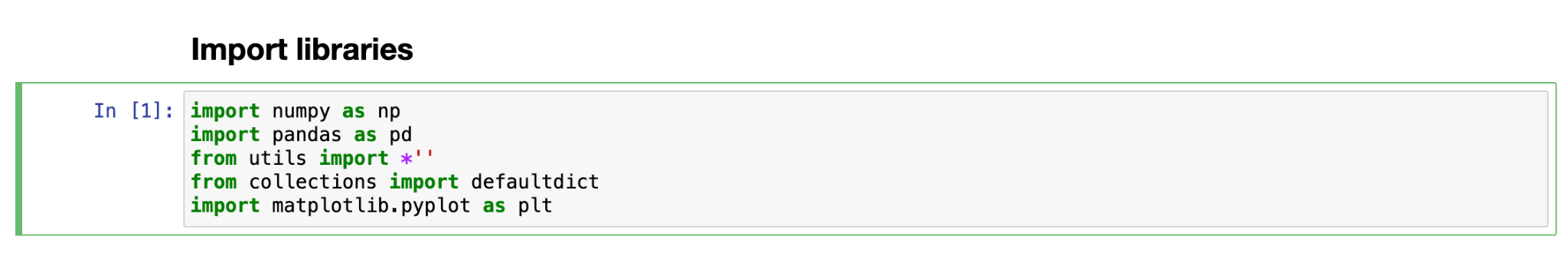
The first step is to install the the git repository on our local computer:



Alternatively, it is enough to download the files on our computer.

Once we have all the needed files, we can open the [usage\_example.ipynb](https://github.com/eochoarv/conformal-hqi/blob/main/usage_example.ipynb) notebook where we will find an example to apply conformal prediction.

Inside the Jupyter notebook, the first step is to import the needed python libraries

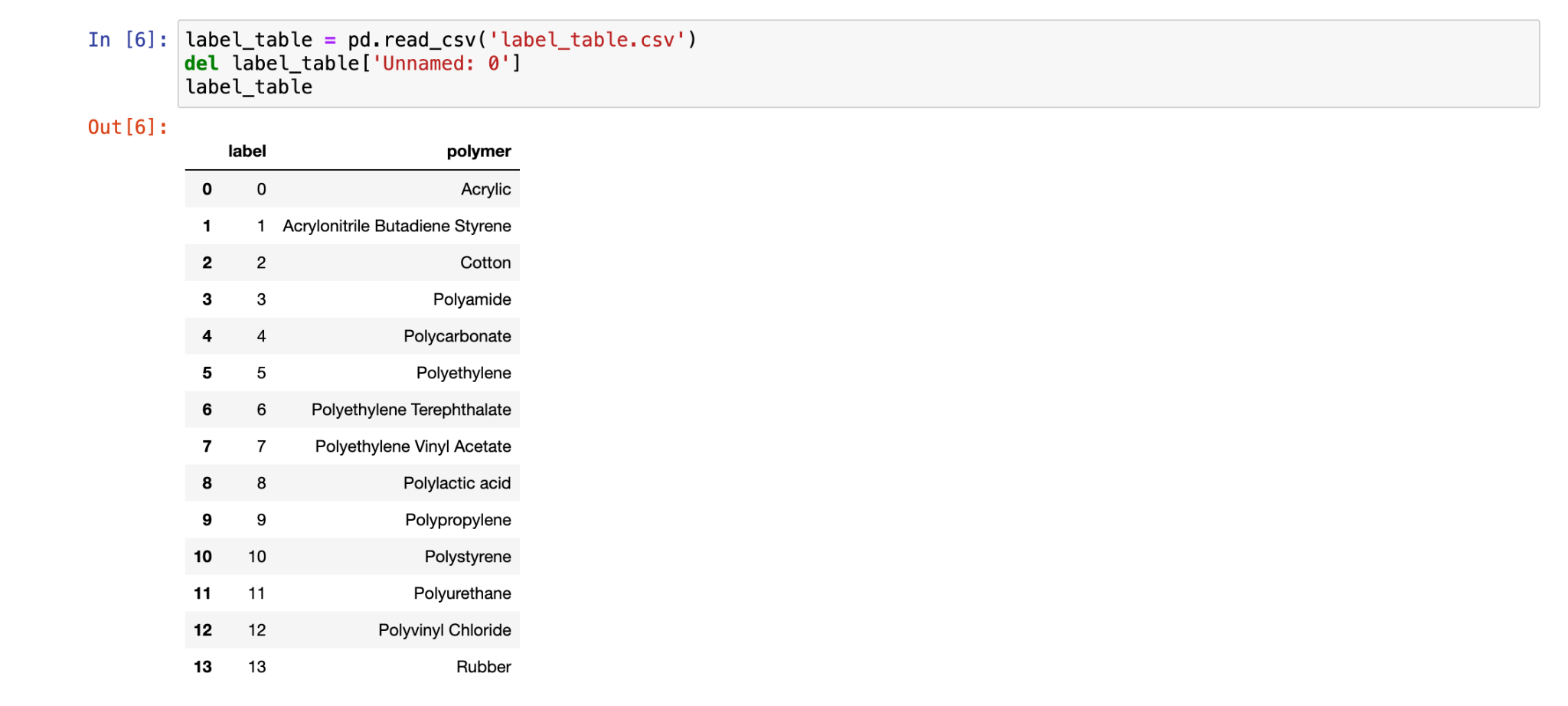


Then, we will read the reference library, calibration set and test set.



The corresponding files need to be in .csv format. An example of the format can be found here

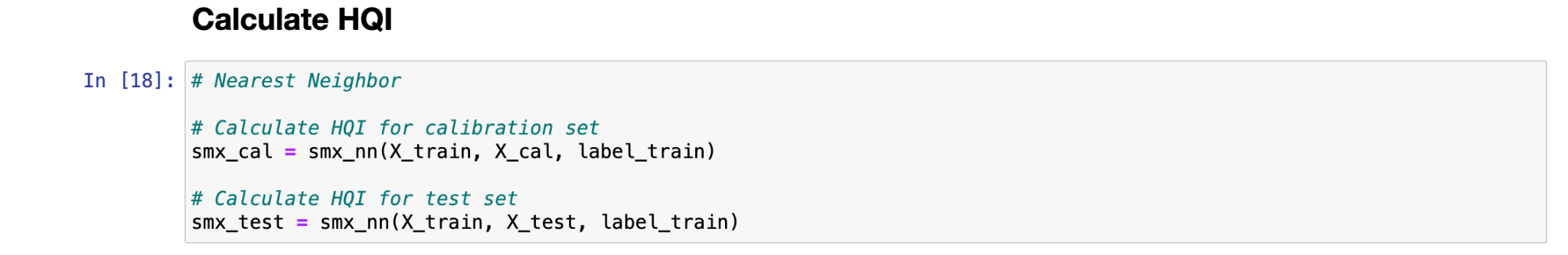
A table with the list of polymers is also needed.



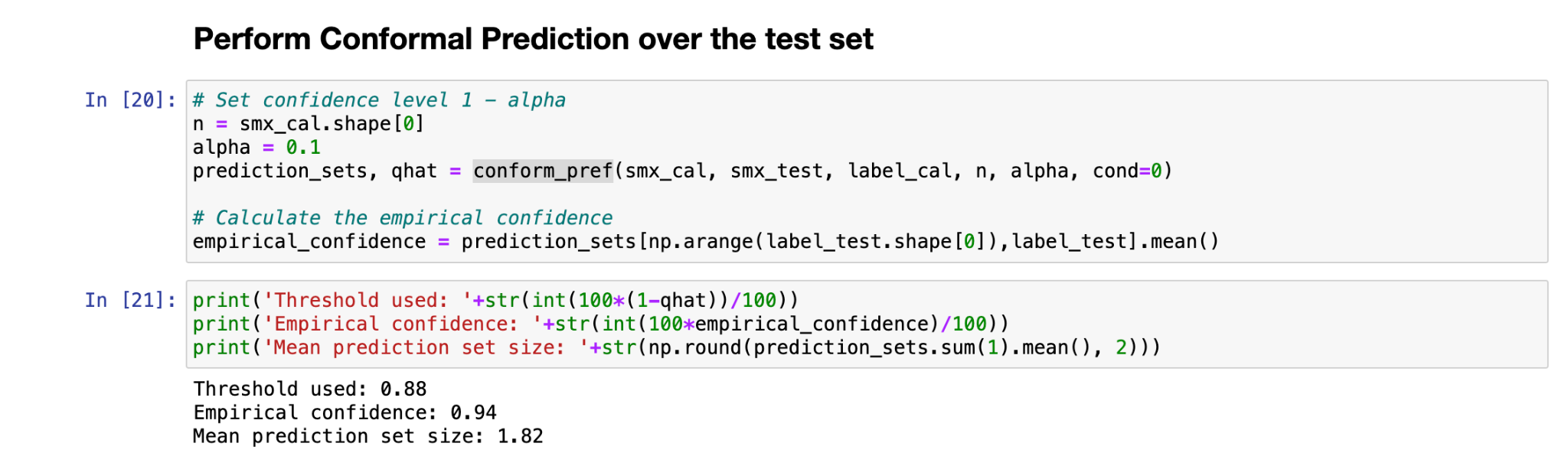
When we have read all the files, we will process the spectral data to be able to calculate the HQI. In this step we simply separate the spectral data from the label/polymer information and apply the SNV normalization.



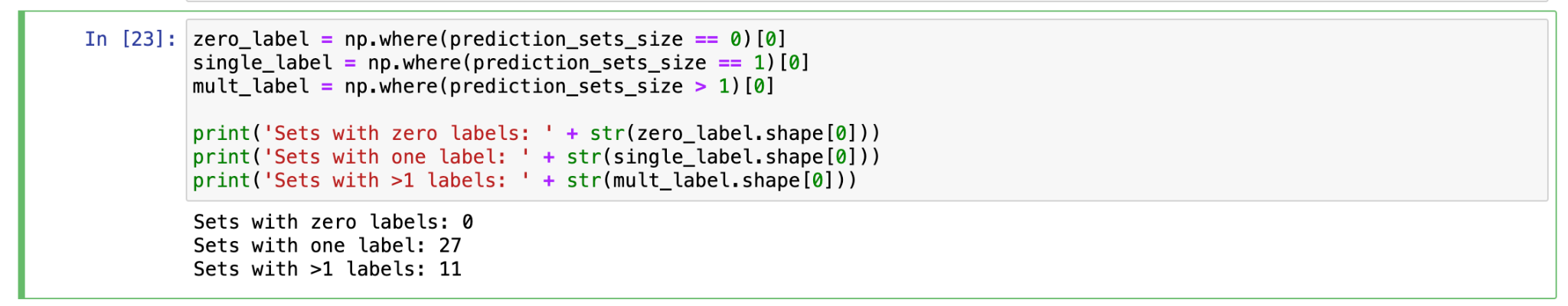
Then, we can calculate our favorite metric for HQI. We will calculate the HQI for the calibration set and for the test set. Remember that the calibration set will help us to construct a distribution of the HQI that will define the threshold.



Now we are ready to perform the conformal prediction for the test set. We need to input the HQI of the calibration set, the HQI of the test set, the label list of the calibration set, the size of the calibration set and the theoretical confidence level alpha to the function conform\_pred(). It will output the corresponding prediction set for each spectra in the test set and the threshold used based on the confidence level. In this case we have access to the true labels of the test set, so we can calculate the empirical confidence.



We can then separate the prediction sets in 3 categories: zero labels, one label and more than one label.



As discussed in the paper, we will consider the prediction sets with zero labels as outliers, meaning that they don’t belong to any of the polymers in the reference library; we will accept all the predicted labels of the predictions sets with just one label and we will perform a manual inspection of the rest of the prediction sets.

