Risk Prediction of Pancreatic Cancer Using Urine Biomarkers

By: Jenny Harston

Tech specs:

**Development environment**: Anaconda3 - Python 3.9.7

**Packages used:**

* pandas
* numpy
* random
* matplotlib
* sklearn

**Files:**

* project\_JennyHarston.py
* pc\_data.csv

Instructions:

* Include both files in the same folder

Text

Description automatically generated

* Make sure all packages used are installed in your environment

Text

Description automatically generated

* Run the code using python
* If you would like to look at the first 10 samples of the original dataset, uncomment line 113
* The code is currently set up to compare the diagnoses of Control vs. PDAC. If you would like to compare Benign vs. PDAC, change line 119 to

dataset = CompareDiagnoses(diag\_dict, 2, 3)

Text

Description automatically generated

* The code is currently set up to display the ROC curve from an estimator. If you would like to display the ROC curve from the predictions, comment out lines 161-167, and uncomment lines 169-175

Text

Description automatically generated

* Model prediction evaluations will be displayed in the terminal and the ROC curve figure will be produced.

Text

Description automatically generated