This is an open-ended task, the purpose of which is to see your approach to a problem. We are providing you with a datafile of instances for which you should build a predictive model. It is a multi-class dataset however you may choose to solve a binary classification problem or a multi-class problem.

There are specific questions we would like you to answer.

Please complete the exercise in a Jupyter notebook. You can use any python libraries you wish. Any discussion can be done in the notebook, or in a separate report if you prefer.

Timeline Guidance

- Please spend up to 3 hours on the task. (if you really want to spend more time, please indicate that you did so).
- Please submit your work within one week.

Dataset

There are approximately 3000 rows in the dataset, in CSV format. Each row has 19 features and one label (called `label`). Each row corresponds to observations recorded for different fish nesting sites. The labels indicate whether the site was home to endangered fish `A`, `B`, `C`, `D`, or none of these.

Many of the features focus around 3 classes of common fish and the activities they do. For example:

- `fish1_activity1`: observation count of a fish of type `fish1` doing `activity1` `fish3_count`:
- observations of unique fish of type `fish3`

Please use the dataset to create a model predicting whether a site will be home to these exotic fish. Reminder: you may chose to focus on individual fish instead of tackling a multi-class problem. Please mention why you chose to proceed as you did.

Questions to Answer

Below we give some specific questions for you to answer. The general areas we are interested in are:

- Exploratory Data Analysis
- Model Training and Evaluation

Please include discussion on:

- What are the characteristics of the dataset? Does anything about the dataset stand out to you?
- What clean-up / data preparation have you done and why?
- What questions would you ask the person who prepared this set?
- Why did you chose the model(s) and training methods you chose?
- Why did you chose the evaluator you chose?
- Does the model performance surprise you?
- If you had one week to work on this problem what would you do?

Submission

Please submit via email. Include in a single archive:

- Your jupyter notebook (*.ipynb)
- A pdf export of your notebook
- Any other supporting documentation