



Rensselaer

why not change the world?®

Lab 3: Classification & Clustering with Abalone

Ahmed Eleish

ITWS-4600/ITWS-6600/MATP-4450/CSCI-4960

September 25th, 2026

Tetherless World Constellation
Rensselaer Polytechnic Institute



Lab 3



Files: <https://rpi.box.com/s/126mhpxpn6ef0q2emfi8gpqjq4ld4340>

Exercise 1:

- Using the abalone dataset, select a subset of features as independent variables and with “age_group” as a target variable, do the following:
 - a) Train multiple kNN models and use classification accuracy to find the optimal k (number of neighbors) value.
 - b) Train a random forest model using the same set of features and compare its classification accuracy with the kNN model using confusion matrices.

(optional) try scaling features and examine effect on model predictions

Note: You must split the dataset into training and testing sets for the model to “learn” from the training set and “predict” for the test set. Make sure to use the same split for training both algorithms.

Exercise 2:

- Use k-means and PAM to find clusters in the abalone dataset using the feature subset from the exercise 1.
- Find the optimal value k (number of clusters) for each model and plot silhouette plots for both models with their optimal k .
- (optional) examine effect of feature scaling on clustering outcome

Please upload to LMS:

1. All your code in a *.R or *.MD file

Thanks!
Have a great weekend!*