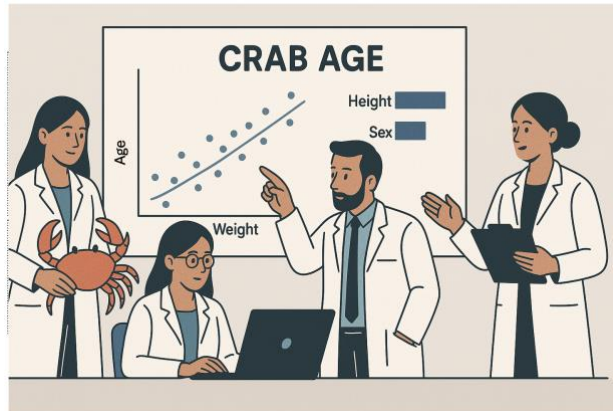


DDS DS 6306 Project 2: Crab Age



Introduction

Wine quality is a multifaceted attribute influenced by many factors, including chemical properties, production techniques, and environmental conditions. The goal of this data science project is twofold:

1. **Objective A:** Your task is to use regression to predict the age of crabs given physical attributes. Good luck!
2. **Objective B:** Analyze the data to uncover key factors or relationships that contribute to predicting crabs' ages. Any other interesting or useful relationships discovered will be valuable as well!

Objective A: Building a Predictive Model

The primary objective is to develop a robust predictive model that accurately predicts crab age based on various features. We aim to achieve the lowest possible Mean Absolute Error (MAE) on the test set, ensuring high precision in our predictions.

Mean Absolute Error (MAE) is a commonly used metric for evaluating regression models. It measures the average magnitude of errors in a set of predictions without considering their direction. The formula for MAE is:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where:

- n is the number of observations.
- y_i is the actual value.
- \hat{y}_i is the predicted value.

Minimizing MAE ensures that our model predictions are as close as possible to the actual values, providing a reliable tool for assessing wine quality.

The student with the lowest MAE on the test set will win an extra 5 points on the project and Bragging Rights! ☺

Objective B: Mining Data for Insights

Beyond prediction, this project aims to delve into the dataset to identify the key determinants of crab age. By analyzing the relationships between various crab metrics and age, we can provide valuable insights to help scientists understand their growth rate, reproductive cycles, and lifespan, which can help commercial markets as well as scientific studies into climate change, effects of ocean pollution, shifts in predatory trends, etc.

Feature Details:

Input variables:

Sex

Length

Diameter

Height

Weight

Shucked

Weight

Viscera

Weight

Shell Weight

Output variable (based on sensory data):

Age

Dates:

Final Project Deadline: Sunday, April 21st at 11:59 pm CST

Deliverables:

7 min recorded group presentation with the link posted on this Google Doc:

https://docs.google.com/document/d/1rTSFZJa8RA4H7wCYanFYfO_bLDQOimvGxMOUdZs7L2M/edit?usp=sharing

Predicted ages of crabs for the competition set. This should be in CSV format with two columns: ID and age.

RShiny Dashboard that at least shows the predicted quality of a wine, given specific specs.

Rmarkdown that describes your analysis.

GitHub Repo with all files that pertain to the deliverables above.

Groups:

Group 1

Barker, Chloe

Nunnally, Drew

Group 2

Dower, Tracy

Camara,

Sumtana

Group 3

Davisson, Kyle

Rocha,

Jonathan

Group 4

Karuri, Titus

Streeter, Devin

Group 5

Vogt, Johnny

Flores, Michael

Please contact your partner via email as soon as possible to establish a line of communication. 😊