Predicting the age of abalone from physical measurements

Project Proposal

Abstract:

Predicting the age of abalone from physical measurements. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope, a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age.

Design:

In this project, I will predict the age of the abalone through its physical measurements, which are represented by a 9 of feathers are:

['Sex', 'Length', 'Diameter', 'Height', 'Whole weight', 'Shucked weight', 'Viscera weight', 'Shell weight', 'Rings'] , The analysis will be based on 4177 abalone measurements information

Data:

Data comes from an original (non-machine-learning) study : The Population Biology of Abalone (_Haliotis_ species) in Tasmania , This project originated from https://archive.ics.uci.edu/ml/datasets/Abalone that offered by uci.edu , which contain 4177 of different abalone information With 9 features for each of them,

Algorithms :

Feature Engineering

- study of the common points between physical measurements
- Create the AGE variable using the rings +1.5
- Dividing the features into numerical and categorical features
- Considering AGE as the goal to be reached after prediction
- Use of numerical and categorical features for forecasting
- Variance Inflation Factor for all feathers
- Split the data into train and test

Models

Lieaner regression, k-nearest neighbors, and random forest classifiers were used before settling on random forest as the model with strongest cross-validation performance. Random forest feature importance ranking was used directly to guide the choice and order of variables to be included as the model underwent refinement

Model Evaluation and Selection

The official metric for Driven Data was classification rate (accuracy) ,Root mean square error (RMSE)

Tools:

- Numpy and Pandas for (EDA) Exploratory Data Analysis

- Scikit-learn for modeling

- Matplotlib and Seaborn for plotting

Done by:

Rawan Mohammad Balubaid

Rawan-balubaid@hotmail.com

0540386411