

Abalone Age Prediction

Jaime Ip 024783334

Toronto Metropolitan University

CIND 820 Big Data Analytics Project

May 5, 2023

Abstract

Abalones are marine snails that primarily inhabit cold coastal waters. They are able to produce pearls by secreting a shell over parasites or gravel that lodge in their flesh. They can also be eaten raw and cooked. Abalones are especially valued in Southeast Asian cuisines and are known as the truffles of the sea in Japan. As the economic value of abalone is positively correlated with its age, estimating the age of abalone can help the farmers and sellers to determine the market price. Similar to estimating the age of a tree, a lab technician can determine the actual age of an abalone by examining a shell sample under a microscope and counting the rings. However, this is a very time consuming and complex process. Therefore, it is necessary to determine the age more efficiently by using other data and methodologies. The theme of this capstone project is forecasting the number of rings based on other physical measurements and then determine the age of the abalone. The dataset that I use for the project is Abalone Dataset from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets/Abalone>). This dataset includes data extracted from 4177 abalone samples that were recorded in December 1995 by Marine Research Laboratories Taroona and Department of Primary Industry and Fisheries in the State of Tasmania, Australia. Each sample has 8 attributes which are the physical measurements of abalone included the number of rings. This project would focus on three questions: 1) Is there any relationship between the number of rings and other physical measurements? 2) How to predict the number of rings based on rules and patterns found in classification and regression 3) Which model is the best for predicating the number of rings? The techniques that are used in solving the above problems include data cleaning, classification, regression, modeling, and model evaluation. Python is the primary tool that is used in this project for coding and data visualization.