A Short Report on predicting age of abalone using Abalone Dataset

Title: Predicting Abalone Age using KNN Machine Learning Algorithm

Problem Statement: The age of abalone, a type of sea snail, is an important factor in determining its commercial value. However, determining the age of abalone is a time-consuming and costly process, which involves cutting the shell and counting growth rings. As a result, there is a need for an accurate and efficient method for predicting the age of abalone based on easily measurable physical attributes such as weight, length, and diameter.

Objectives: The goal of the project is to develop a model that can predict the age of an abalone based purely on the other physical measurements. This would allow researchers to estimate the abalone's age without having to cut its shell and count the rings.

Here we applying the K-nearest neighbor (KNN) machine learning algorithm to find the closest prediction score possible.

Dataset: The Abalone Dataset. This dataset contains age measurements on a large number of abalones with eight characteristics of sex, length, diameter, height, whole weight, shucked weight, viscera weight, and shell weight.

Methodology: The dataset used for this study contains information on the physical characteristics and age of abalone, collected from the Tasmanian coast. The dataset contains 4,177 observations with eight attributes, including sex, length, diameter, height, whole weight, shucked weight, viscera weight, and shell weight.

The EDA(Exploratory Data Analysis) is done using dataset for known the dataset closely and read the data nature by viewing at dataset using different plots.

Preprocessing is performed before to perform this KNN classification as needed for outcome with a better result.

The KNN algorithm was used to develop the prediction model. The dataset was split into training and testing sets, with 80% of the data used for training and 20% for testing.

The optimal value of K was determined by Hyper Parameter Tuning using GridSearchCV. The model was evaluated using the root mean squared error (RMSE) metrics.

Findings: The KNN algorithm was able to accurately predict the age of abalone based on its physical characteristics. The optimal value of K was found to be 22, based on GridSearchCV. The prediction model achieved an RMSE of 2.15 on the test dataset.

Conclusion: The KNN machine learning algorithm can be an effective method for predicting the age of abalone based on its physical characteristics. The developed model achieved a relatively high level of accuracy in predicting the age of abalone, with an RMSE of 2.15. This model can be used by abalone farmers, fishermen, and seafood processors to estimate the age of abalone and to make informed decisions about their management and harvest.

References:

1. https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/

2. https://github.com/mdsaiful13/Machine-Learning-JU

3. https://colab.research.google.com/github/mdsaiful13/Machine-Learning-JU/blob/main/Abalone_size.ipynb

4. https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/abalone.data