Comparative Analysis of the Performance of Different Machine Learning Techniques in Predicting the Age of Abalones

Marina Adamović, Pavle Tepavčević

Faculty of Sciences, University of Novi Sad

September 2024

Introduction

- Group of sea snails
- Used in cuisines and it's leftover shell is fashioned into jewelry

Figure: Abalone and its rings.

Defining the problem

- Older abalones have a higher market value
- Task: Determine the age
- Age = Rings + 1.5
- Current approach: Cutting the shell using a microscope and counting the number of rings
- Problem: Time consuming, expensive and it damages the abalone
- Solution: Predicting it's age based on physical measurements using machine learning techniques

Dataset

- Features: Gender, Length, Diameter, Height, Whole weight, Shucked weight, Viscera weight, Shell weight
- Target: Rings
- Good dataset with only a few measurement errors

Gender distribution

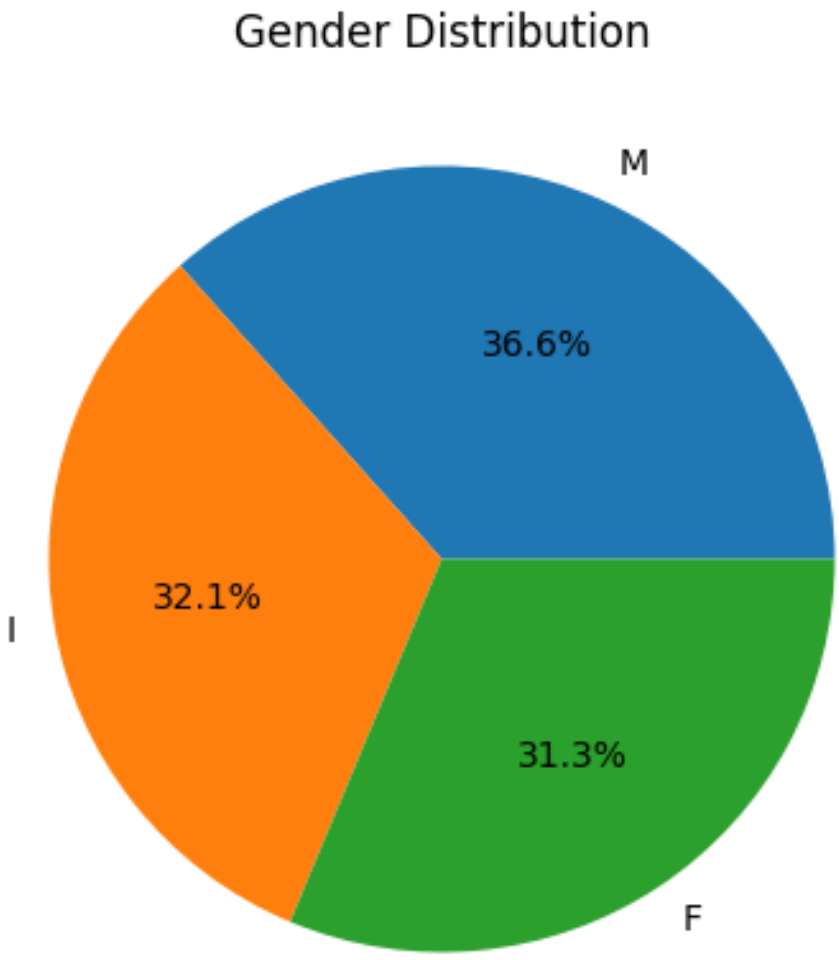


Figure: Gender distribution of abalones.

Abalone rings distribution

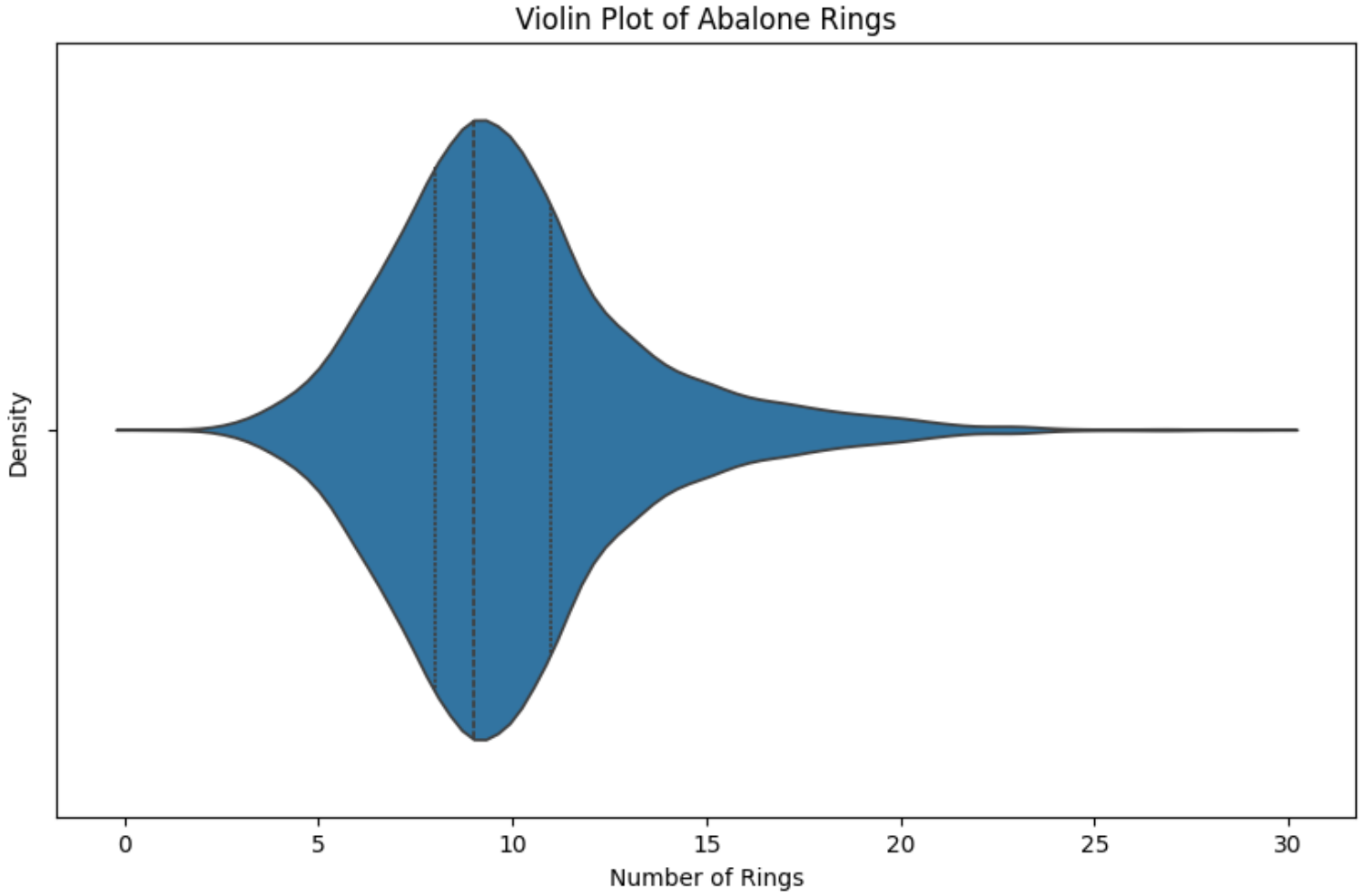
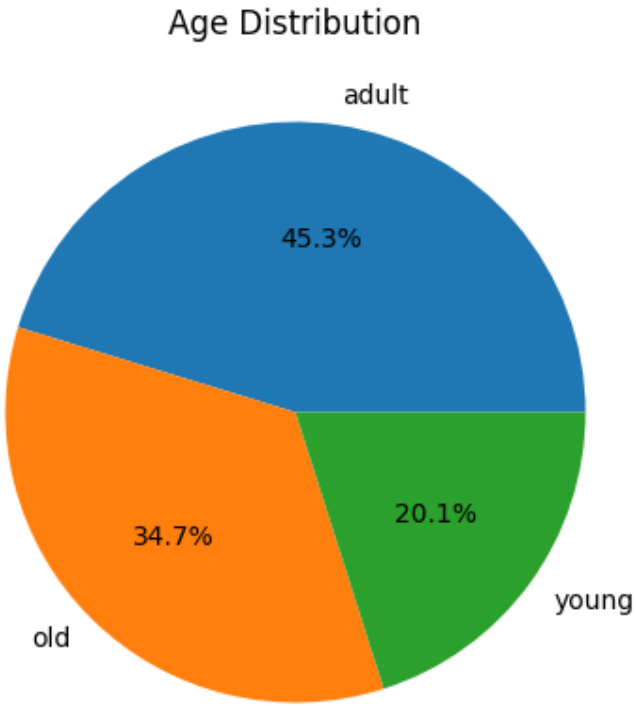


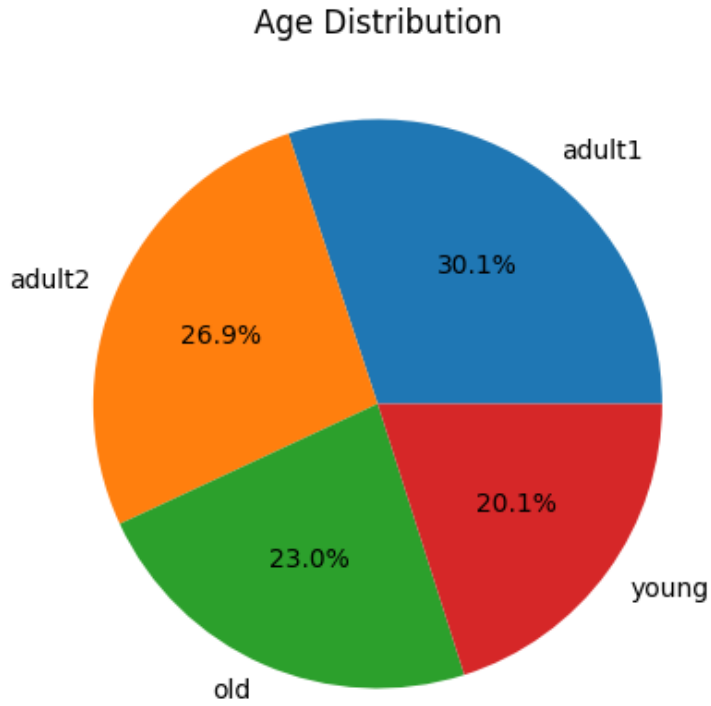
Figure: Violin plot of abalone rings distribution.

Age groups

- Dividing them into age groups
- Two approaches:



Three age groups.



Four age groups.

Distribution of gender by each age group

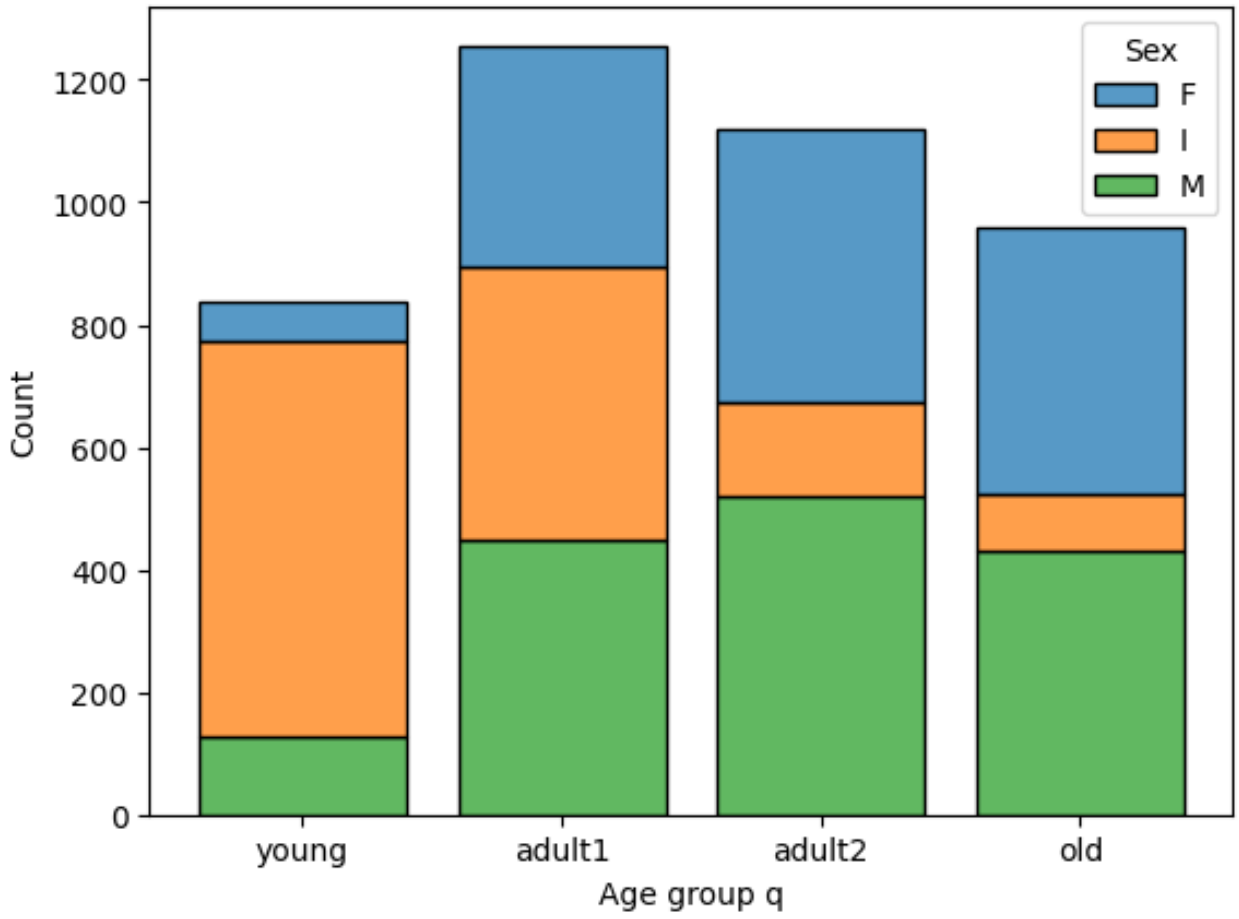


Figure: Distribution of gender of abalones in each age group.

Distribution of features for each gender.

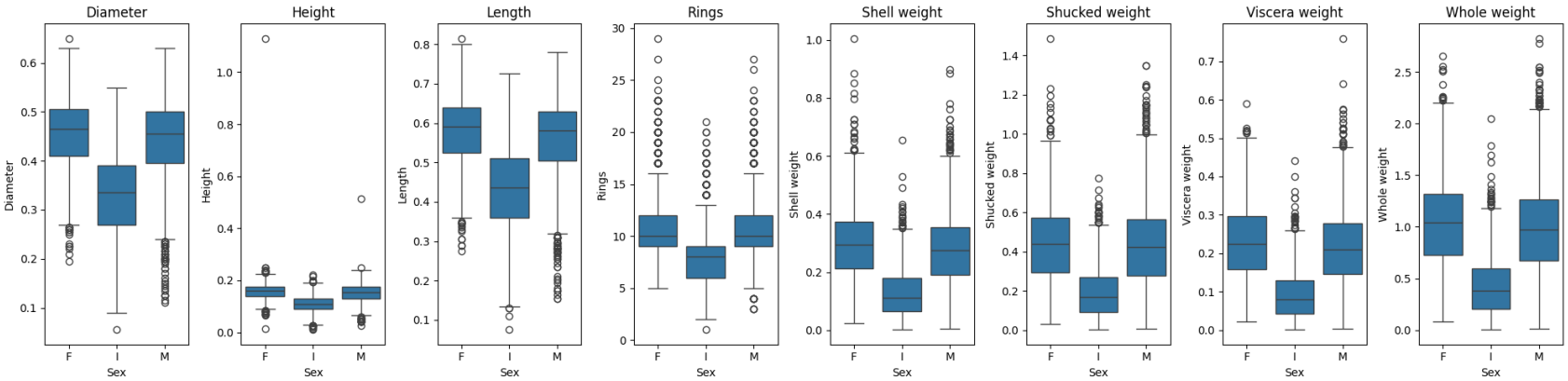
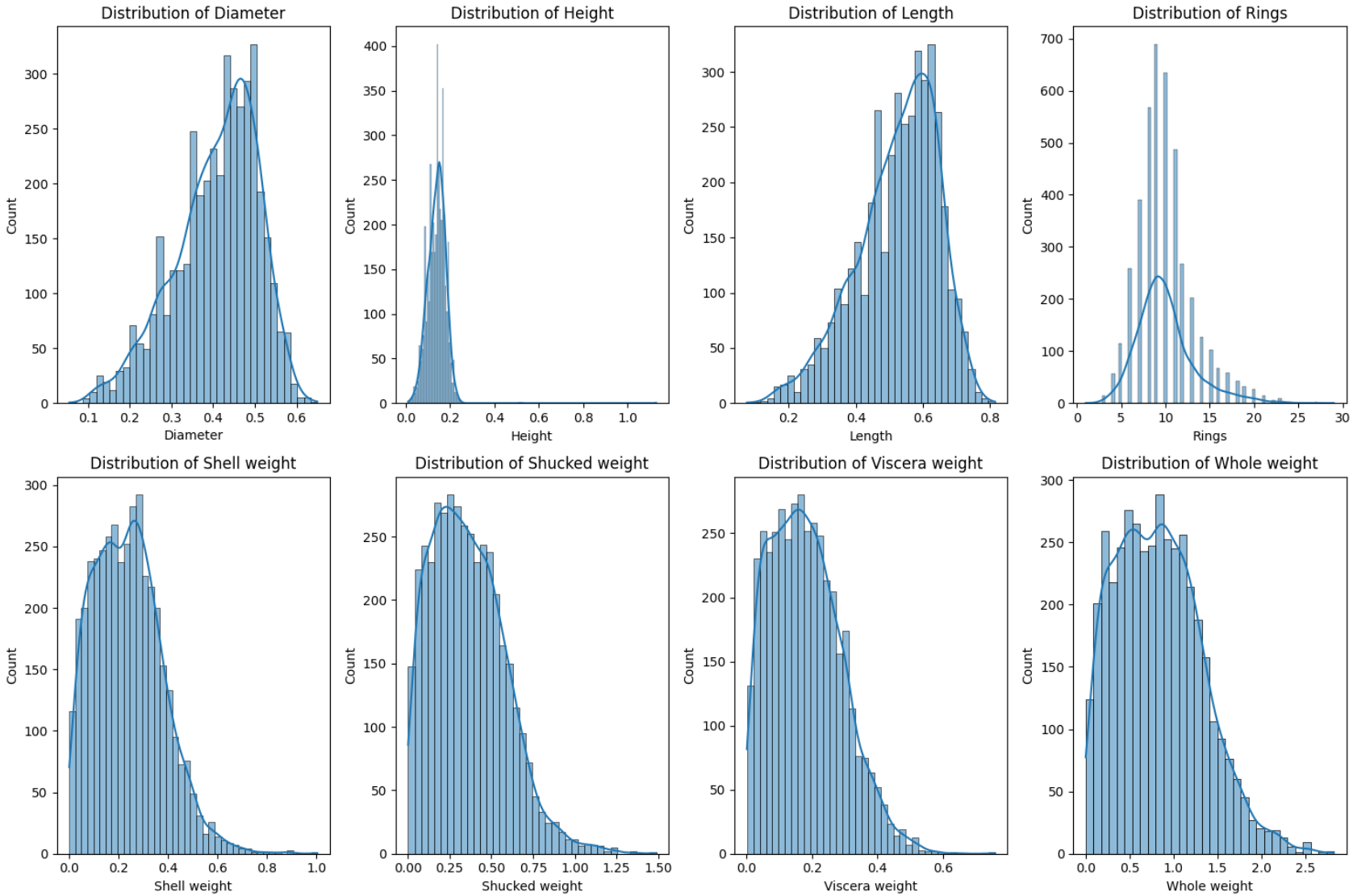


Figure: Distribution of each feature for three genders.

Distributions of physical features



Distributions of physical features

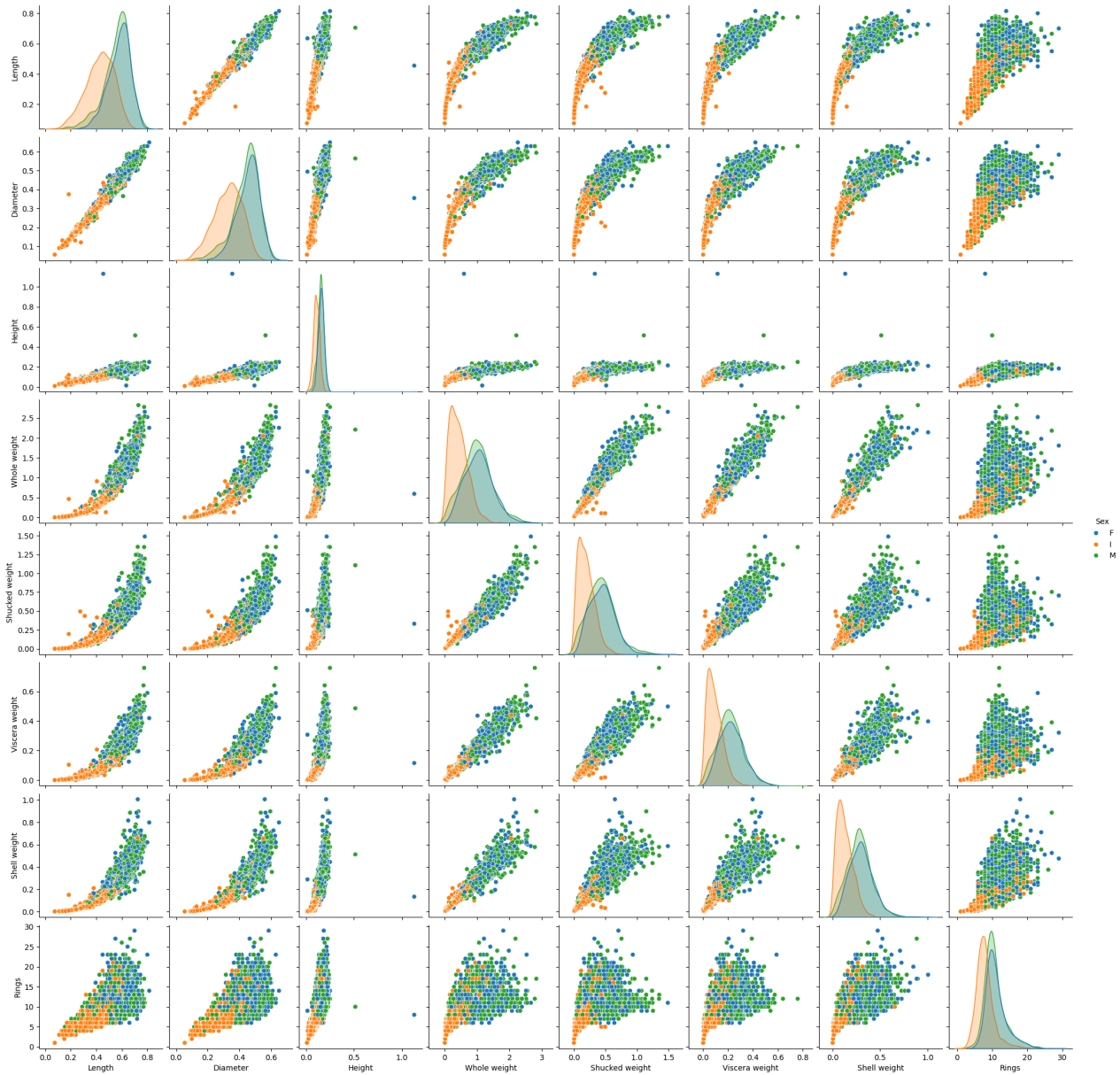


Figure: Pairplot of all physical features.

Correlation matrix

Figure: Correlation matrix of all physical features.

Explaining the variance

- The features are very correlated

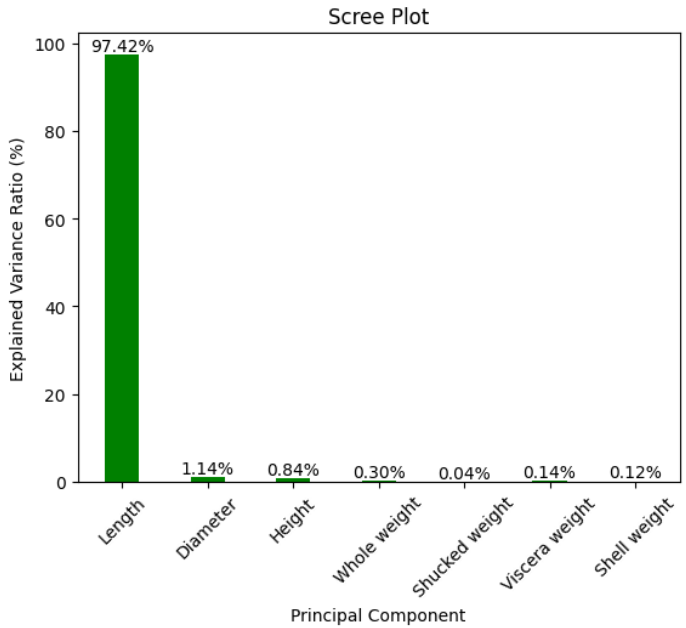


Figure: Percentage of variance that each feature can explain

Machine learning

- Classification - predicting the age group
  - k-NN (three or four age groups)

Machine learning

- Classification - predicting the age group
  - k-NN (three or four age groups)
- Regression - predicting the number of rings
  - Linear regression (univariate and multivariate)
  - k-NN
  - Neural networks

Classification

- Method - using k-NN (based on Length) to estimate the PDF of each group and then using the ML criteria
- Three age groups:
  - Error of classification $= 41\%$
- Four age groups:
  - Error of classification $= 52\%$

Linear regression

- Two approaches: Univariate and Multivariate

Linear regression

- Two approaches: Univariate and Multivariate
- Univariate - taking the length
  - MAE $= 2.02$
  - $R^2 = 0.30$

Linear regression

- Two approaches: Univariate and Multivariate
- Univariate - taking the length
  - MAE $= 2.02$
  - $R^2 = 0.30$
- Multivariate - taking all physical features
  - MAE $= 1.63$
  - $R^2 = 0.55$

Regression based on k-NN

- Method - the prediction of number of rings is the mean of number of rings of k nearest neighbors (based on Length)
- MAE $= 1.88$

Neural networks

- Method - a dense neural network which predicts the number of rings is trained
- MAE $= 1.53$ (based on everything)

Comparison of Regression Methods

Method	**MAE**
Linear Regression (Univariate)	2.02
Linear Regression (Multivariate)	1.63
k-NN Regression	1.88
Neural Networks	1.53

Table: Comparison of MAE for Different Regression Methods.

Thank you!