1400/02/27

**FISH CLASSIFIER**

*ACCORDING TO FEATURES OF LIGHTNESS AND WIDTH*

Benyamin Moadab

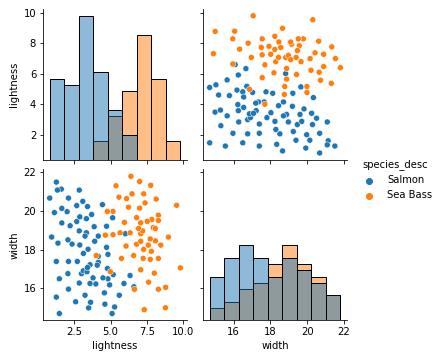
Student ID: 39921441044013

Islamic Azad University Qazvin Branch

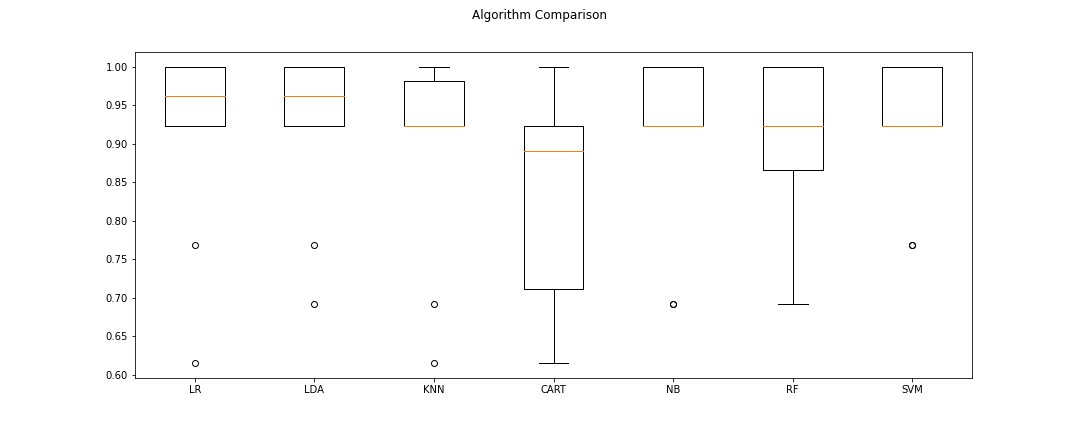
1. **Algorithm Design and Explanation**

In this assignment, several fish classification algorithms were developed and tested for fish data.This dataset includes 130 samples of salmon (Class 1) and sea bass (Class 2) classes. Each class of fish is represented by two unit measurements, which are the length and brightness of a fish.

In order to create a classification, training data were first analyzed and mapped to find a decision rule for classifying fish. For this reason, the data were divided into two groups of fish. Histogram for both salmon and salmon length and brightness As shown in Figure 1 and Figure 2, the box diagram in Figure 3 is also shown. As shown in the histograms, the length of the salmon can be And considered sea fish to be a normal distribution.

****

**Figure 1,2) Histogram & Classfication Data**

****

**Figure 3)Boxplot of Algorithm Comparison**

Boxplots for salmon and seabass fish lengths were plotted separately to support the normality decision. Since there are few exceptionally large or small values and boxplots for both fish lengths look symmetric, it can be said that two fish lengths & Lightness are distributed normally.

After plotting the data and as a result of accuracy and standard deviation for both class were calculated as shown in Table 1.

|  |  |  |
| --- | --- | --- |
|  | **Accuaracy** | **Standard Deviation** |
| **LogisticRegression** | 91.53462% | 0.121383 |
| **DecisionTreeClassifier** | 92.307692 % | 0.103203 |
| **KNeighborsClassifier** | 89.230769 | 0.124985 |
| **LinearDiscriminantAnalysis** | 83.186813 | 0.123258 |
| **GaussianNB** | 90.769231 | 0.113053 |
| **RandomForestClassifier** | 90.000000 | 0.114354 |
| **SVC** | 92.307692 | 0.084265 |

**Table 1. Result of accuracy and standard deviation for all Algorithm**

With descriptive statistics in Table 1, the classifier was implemented as a function. In predict function there is one input parameter which is a dataset to be classified. For every length in the dataset, firstly probabilities of being salmon and seabass for the related length are calculated according to normal distribution. If the probability of being salmon is greater than the probability of being seabass, the classifier labels this fish as salmon, otherwise it labels this fish as seabass.

1. **Results**

Prediction classification algorithms were tested on training data and the accuracy of support vector machine algorithms and linear diagnostic analysis is higher than other methods with 92.307692%.