

ASSIGNMENT DATE	08 OCTOBER 2022
STUDENT NAME	KOMMI CHENCHU KISHORE
STUDENT ROLL NO	110719104024
MAXIMUM MARKS	2 MARKS

ASSIGNMENT:3

ABALONE AGE PREDICTION:

The number of experiments were performed to obtain the optimal parameters for each model. All the experiments were performed on the standard Intel (R) Core (TM) i5-

7200U CPU @ 2.50GHz computer in an Anaconda environment with Python as the programming language. The training dataset consists of 4176 samples. These samples were divided into training consisting of 2923 samples (70%) and testing 1253 samples (30%) subsets.

The Training accuracy with different optimizers is shown fig. 1. BPFFNN model obtained high accuracy for both training and testing. Moreover, Adadelta optimizer scored better compared to other optimizers with BPFFNN (89% training and 88% testing). The figure shows that all optimizers produced similar results except Sgd optimizer.

In fig. 2 the convergence of five different optimization algorithms is illustrated in terms of training loss over the epochs. BPFFNN model had a lower training loss with

Adagrad optimizer. Sgd starts with a rapid descent, but after 150 epoch stops improving.

Rmsprop, Adadelta and Adam optimizers seem to perform almost the same.

Table 1 shows the confusion matrix for a multiclass classification problem with three

classes (1, 2 and 3). As seen in the table, TP₁ is the number of true positive samples in

the class 1, that is, the number of samples that are correctly classified from class 1. E₁₂

samples that were incorrectly classified as class 2 or 3, i.e., is the sum of E_{12} and E_{13} .

The true negative (TN) of any class is equal to the sum of values except row of true class and the column of predicted class.

6

Table 2 summarizes the accuracy of all the classifiers. Generally, all classifiers performed equally well, except the Gauss Naive Bayes, which obtained relatively lesser accuracy (60.88%). Moreover, the Random Forest classifier produced the highest performance on our dataset (87%) followed by SVM, which achieved an accuracy of 86.76%. Furthermore, KNN and Decision tree classifiers reach almost equal accuracy 86.28%, 86.44%, respectively. Compared with the other classifiers, the performance of the proposed model was relatively better. From the obtained results, we can conclude that the BPFNN reached the best accuracy in the abalone age prediction task.

Table 1 The confusion matrix for a multiclass classification problem with three classes.

Pred

True

G1
G2
G3
G1
TP1
E21
E31
G2
E12
TP2
E32
G3
E13
E23
TP3

G1: age < 7, G2: $7 \leq \text{age} \leq$, and G3: age > 16

Fig.1. Training accuracy with different optimizers

Fig. 2. Training loss for each optimizer.

Fig. 3 Mean Square Error for different optimizers

7

Table 2 Comparison of applied algorithms.

No

Method

Accuracy %

1

KNN

86.28

2

Gaus Naive Bayes

60.84

3

Decision Tree

86.44

4

Random Forest

87.00

5

Support Vector Machine

86.76

6

Proposed (BPFFNN)

88.25

The confusion matrix of the Random Forest algorithm is presented in Table 3.

Ob-

tained results demonstrate that it performs the best results for Group-2 class of the da-

taset. There only 8 data in Group-2 that are misclassified. One thousand forty-

one data in Group 2 are classified correctly. Only 3 data classified correctly in Group-3, which

shows it does not perform well.

Table 3 Confusion matrix for Random Forest.

Pred

True

G1

G2

G3

G1

35

31

0

G2

7

1041

1

G3

0

136

3

G1: age < 7, G2: $7 \leq \text{age} \leq$, and G3: age > 16

Confusion matrix of SVM algorithm is presented in Table 4. Obtained results

demonstrates that, the SVM algorithm gave the worst results after the Gauss Naïve

Bayes algorithm for class 2 (between 7 and 16 age of abalone). 59 data in class 2 are

misclassified, that is, FN₂.

Table 4 Confusion matrix for SVM.

Pred

True	Pred
G1	44
G2	22
G3	0
G1	11
G2	990
G3	48
G1	0
G2	85
G3	54

G1: age < 7, G2: 7 ≤ age ≤ 16, and G3: age > 16

Confusion matrix of KNN algorithm is presented in Table 5. Obtained results demonstrates that, 1006 data in class 2 are classified correctly, i.e. TP₂. Only 35 data classified correctly in the 3(above 16 age of abalone) class, that is, TP₃. The KNN and decision tree algorithms gave the worst results for class 1.

Table 5 Confusion matrix for KNN.

True	Pred
G1	41
G2	25
G3	0
G1	11
G2	1006
G3	32

0
 104
 35
 G1: age < 7, G2: $7 \leq \text{age} \leq$, and G3: age > 16

Confusion matrix of Gauss Naive Bayes algorithm is presented in Table 6.

Obtained results demonstrates that, relevant algorithm preform the best results for 1(below 7 age of abalone) and 3 (above 16 age of abalone) class of the dataset. There only 4 data in class 1 (below 7 age of abalone) that are misclassified, that is, FN_1 . The Gauss Naive Bayes algorithm gave the worst results for class 2. Only 629 data classified correctly in the 2((between 7 and 16 age of abalone)) class, i.e. TP_2 . 420 data in class 2 are misclassified, that is, FN_2 .

Table 6 Classification result for Gauss Naive Bayes

True	Pred
G1	0
G2	104
G3	35
G1	62
G2	4
G3	0
G1	67
G2	72
G3	0

G1: age < 7, G2: $7 \leq \text{age} \leq$, and G3: age > 16

The confusion matrix of the decision tree algorithm is presented in Table 7.

Obtained results demonstrate that 1018 data in class 2 are classified correctly, i.e. TP_2 . Only 31

data in class 2(between 7 and 16 age of abalone) are misclassified, that is, FN₂.

The

decision tree algorithm gave the best results after the random forest algorithm for class

2. Only 24 data classified correctly in the Group 3. The decision tree algorithm gave the

worst results after the random forest algorithm for class 3. While for 3 class, it does not

perform well.

Table 7 Classification result for Decision Tree.

	Pred
True	
G1	
G2	
G3	
G1	
41	
25	
0	
G2	
10	
1018	
21	
G3	
0	
115	
24	
G1: age < 7, G2: $7 \leq \text{age} \leq$, and G3: age > 16	
9	

The overall results obtained for abalone classification using the six conventional

classifiers were satisfactory except Gauss Naive Bayes classifier. The proposed

BPFFNN outperformed all other classifiers in terms of classification accuracy.

In addi-

tion, we compared our approach with CNN based method proposed by authors in [8],

which reported 79.09% accuracy. We believe that for simple datasets such as the one we

used in this study, the conventional machine learning approaches are more effective than

deep learning-based approaches. Even though deep learning-based approaches have

shown high classification accuracy for many problems, yet they are data intensive. We prefer conventional machine learning approaches over deep learning methods for both ease of implementation and classification accuracy in scenarios like this where the dataset is small.