lab11_Biology

Batuhan Seyhan S20343

TASK 1

Iris data

TASK 2

The provided file contains data related to the Iris flower dataset. It consists of150 instances, with 5 variables, including 4 numeric features (sepal length,sepal width, petal length, and petal width) and a categorical target variable(iris species). The dataset includes measurements of sepal length, sepalwidth, petal length, petal width, and the corresponding iris species for eachinstance.

TASK 3

Supervised Learning

TASK 4

If the petal length is greater than 1.9, the classification depends on the petalwidth. If it's less than or equal to 1.9, the classification is directly assigned as"iris-setosa." And so on!

TASK 5

5 levels

TASK 6

https://ibb.co/4jGVsP6

TASK 7

https://ibb.co/k3vRBpT

TASK 8

150

TASK 9

Logistic Regression auc—> 0.9976666666666666 CA-0.9666666666666667 F1-0.9666633329999667 PREC-0.9667867146858743 RECALL 0.9666666666666667 MCC0.9500633396673706

TASK 10

Based on the given confusion matrix, the model that performs the best is theone for Iris-setosa since it correctly predicts all instances. However, there aresome misclassifications for Iris-versicolor and Iris-virginica, but thedifferences between the models are relatively small.

TASK 11

https://ibb.co/LQRMf5D

TASK 12

In Orange, after loading the iris dataset and applying the specified settings,the "Data sampler" component is added. The purpose of the "Data sampler"is to create a subset of the data for training and testing purposes.

TASK 13

In the "Test and Score" component, the performance of two models, LogisticRegression and Tree, is evaluated. Both models achieved similar results, withhigh AUC, Classification Accuracy (CA), F1 Score, Precision, Recall, andMatthews Correlation Coefficient (MCC) values. This indicates that bothmodels were effective in classifying the iris dataset.

TASK 14

Based on these results, the performance of both the Logistic Regression andTree models remains similar to the previous evaluation. The AUC valuesindicate that both models can distinguish between different classes with highaccuracy. The Classification Accuracy (CA) also remains high, showing theproportion of correctly classified instances. Overall, the algorithm workssimilarly to the previous evaluation, with both models providing effectiveclassification of the iris dataset. However, there is a slight decrease inperformance across the metrics, suggesting that the models might have aslightly reduced ability to classify the test data compared to the previousevaluation.

TASK 15

5 records were missing

TASK 16

lowest Classification Accuracy (CA) is Naive Bayes, with a CA value of 0.2800.Therefore, Naive Bayes is the worst model at data classification

TASK 17

IRIS-SETOSA

TASK 18

IT'S KNN model with 5 erros.

TASK 19

Iris-virginica Iris-versicolor Iris-versicolor Iris-virginica Iris-versicolor Iris-virginica Iris-versicolor Iris-virginica Iris-virginica Iris-versicolor

TASK 20

Yes, the classification using the "Neural Network" model is different from the"Logistic Regression" and "Tree" models based on the provided data.

TASK 21

Yes the models are different

TASK 22

iris iris(Logistic Regression) sepal length sepal width petal length petal widthIris-setosa Iris-versicolor Iris-virginica Iris-setosa Iris-versicolor Iris-virginicacontinuous continuous continuous continuous class meta Iris-virginica Iris-versicolor 4.9 2.5 4.5 1.7 Iris-versicolor Iris-virginica 5.9 3.2 4.8 1.8 Iris-versicolor Iris-virginica 6.0 2.7 5.1 1.6 Iris-versicolor Iris-virginica 6.7 3.0 5.0 1.7Iris-virginica Iris-versicolor 6.0 2.2 5.0 1.5