CS5542 Big Data Apps and Analytics

LAB ASSIGNMENT #4

Name:Sidrah Junaid

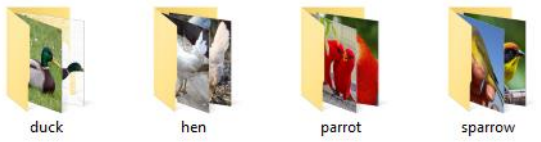
Class Id:15

1.Write a spark program for the following Machine Learning Tasks

Answer

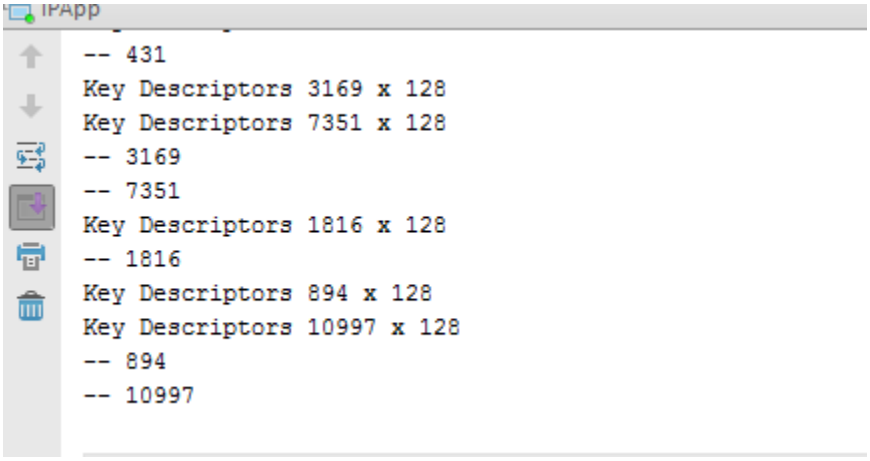
Selected data set contain different images of birds.By classifying the images we can easily identify the different birds.

Dataset:

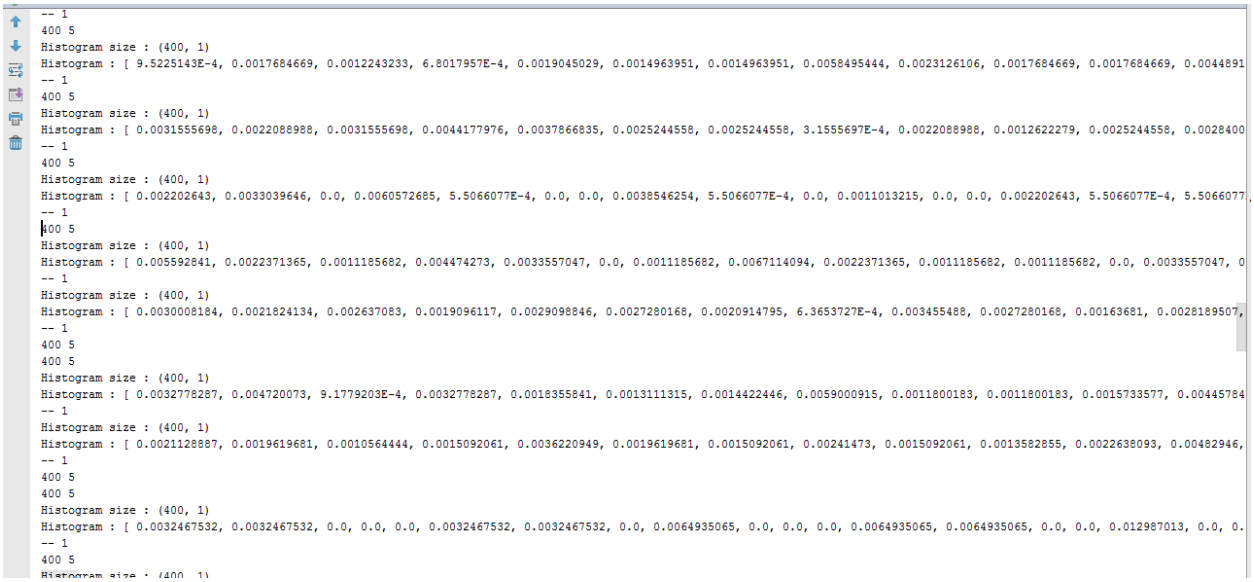


The dataset is split into 70% of Training data and 30% of test data.

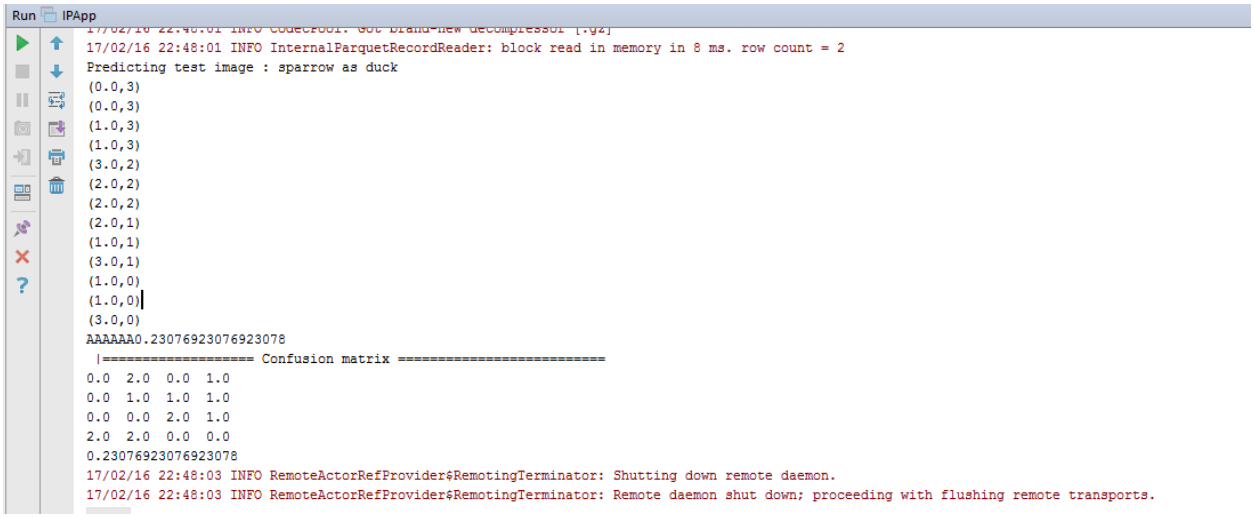
The key descriptors generated by Decision Tree Algorithm are:



Histogram :



Confusion Matrix



Accuracy

23.07%

Error Rate