

Iris setosa, iris versicolour assignment



**ASSIGN
BUSTER**

Finally, the “ Zoo” data set is a trivial one with 7 classes, which are animal groups, with a total of 101 instances. Each animal instance contains 18 attributes, those of which include the animal’s name or race, 2 numeric for its legs and its type, and 15 Boolean-valued attributes; those that involve simple yes or no answers. The following is an analysis of 4 classification algorithms that can be optimally used for these data sets.

ANN. would be a good decision when simplicity and accuracy are the overwhelming factors, like in the “ Zoo” data set. This classification algorithm does not focus on the prior probabilities, and is very efficient in structure. The primary computation is the sorting procedures in order to figure out the k-nearest neighbors for the test data. There are many advantages. It is structurally trivial, but it’s able to make complex decision boundaries, it doesn’t need much information to be able to work, it naturally gets in tune with our problem-solving techniques, and it learns easily.

The disadvantages are that it takes quite a long time to classify and that it’s somewhat hard to find the best value for k. Decision Tree The Decision Tree algorithm helps solve the problem of classifying data into multiple groups of data. It provides innovative rules for solving large amounts of classification assignments because it works on every different type of data. It’s well-suited for analyzing abundant amounts of info, such as the “ Adult” data set, because it does not need to load all the data in the system’s main memory all at the same time.

It uses a root system to remove the burden of the problem’s difficulty. The Decision Tree exploration engine is used for assignments such as classifying

databases or predicting results. These decision trees should be used when your mission is to assign your records to some ample categories. They help you out with rules that are easy to comprehend, ND those which can also help you pinpoint the best fields in case of future involvement in the project. There are an equal amount of advantages and disadvantages here.

In the bright side, it is easy to comprehend and to generate rules, and it makes your life a whole lot easier when the problem gets degraded in difficulty. On the other hand, once an error has been made on a node at level n , then any and all nodes at level $n-1$, $n-2$, $n-3$,... , $n-k$ will also be wrong. Furthermore, it is not good at handling continuous variables. Nevertheless, being able to work with mass scale database files with Just his algorithm is reputable in itself.