Module 7: Supervised Learning-1

Case Study – 2

edureka!

Case Study – 2

Objective:

- Learn to handle missing values
- Learn to fit a decision tree and compare its accuracy with a random forest classifier.

Questions:

1. Let's attempt to predict the survival of a horse based on various observed medical conditions. Load the data from 'horses.csv' and observe whether it contains missing values.
   [Hint: Pandas dataframe has a method isnull]

2. This dataset contains many categorical features, replace them with label encoding.
   [Hint: Refer to get_dummies methods in pandas dataframe or Label encoder in scikit-learn]

3. Replace the missing values with the most frequent value in each column.
   [Hint: Refer to Imputer class in Scikit learn preprocessing module]

4. Fit a decision tree classifier and observe the accuracy.

5. Fit a random forest classifier and observe the accuracy.