|  |
| --- |
| Supervised Learning-1 |
| Case Study |

Case Study

# Objective:

* Learn to handle missing values
* Learn to fit a decision tree and compare its accuracy with 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]

1. 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]

1. Replace the missing values by the most frequent value in each column. [Hint: Refer to Imputer class in Scikit learn preprocessing module]
2. Fit a decision tree classifier and observe the accuracy.
3. Fit a random forest classifier and observe the accuracy.