
Indian Institute of Technology Indore
Discipline of Computer Science & Engineering
CS 403/603 Machine Learning
Assignment -Decision Tree Learning (ID3)

Dataset:

https://archive.ics.uci.edu/ml/machine-learningdatabases/00267/data_banknote_authentication.txt

(It is a txt-file but by changing the txt suffix to csv you can use the Pandas library's read_csv()-function.)

The data is about the **authentication of banknotes**. If you like you can read more about the data here

<https://archive.ics.uci.edu/ml/datasets/banknote+authentication#>

Attribute Information:

1. variance of Wavelet Transformed image (continuous)
2. skewness of Wavelet Transformed image (continuous)
3. curtosis of Wavelet Transformed image (continuous)
4. entropy of image (continuous)
5. class (integer)

In our data the classes are represented by "0" for real and "1" for fake. Divide the dataset into 80:20 Train: Test ratio and perform the following task

Task:

(A) Train your decision tree classifier algorithm implemented in scikit learn on the train-data using the impurity measure:

- a. Information Gain

Draw confusion matrix and report the accuracies on test data.