

## Big Data Analytics - Homework 5

1. What was the minimum mixed entropy for the final decisions you found?

**Solution:** The minimum mixed entropy obtained is **0.4106332421421127** and the best attribute obtained is '**BrkDnkIsCoffee**'

2. What structure did your final decision stump classifier have? What was the if-else stump you got?

**Solution:**

```
if ('BrkDnkIsCoffee' <= 0):
    target_variable = 0
else:
    target_variable = 1
```

3. What was the accuracy of your resulting classifier on the training data?

**Solution:** The accuracy of the classifier on the training data is **91.3 %**

4. What else did you learn along the way?

**Solution:** The following are a few things I learnt from this homework:

- Dealing with attributes with continuous value: The given dataset had three attributes which has continuous values. The way I handled it in my code is that I have set a flag, if the value is 1 then it includes these attributes in computing the best attribute otherwise, its not included in the computations. I used Otsu's method to decide the threshold, any data point below the threshold is set to 0 and above the threshold is said to 1.
- Meta Programming: This is the first time I wrote a program that writes another program. It did involve a lot of reading for me to understand how its done.
- Importance of entropy: Understood the importance of entropy and how the value of it can signify how pure or impure a node is.

5. What can you conclude?

**Solution:**

- Created decision stumps to predict the target variable.
- Testing the classifier with the training data and then the validation data.
- Finding the accuracy of the results obtained from the classifier for the training data.
- File handling in order to create program that write the classier program.
- Writing the output obtained from the classier into a csv file.
- Learnt how to build a model, train it and then classify the validation set