**Assignment 4**

**Decision tree learning for car evaluation**

**Test Data**

* **Provided**: Car evaluation dataset
* **Total entries**: 1728
* **Class values**
  + unacc, acc, good, vgood
* **Attributes**
  + buying: vhigh, high, med, low
  + maint: vhigh, high, med, low
  + doors: 2, 3, 4, 5more
  + persons: 2, 4, more
  + lug\_boot: small, med, big
  + safety: low, med, high

-------------

**Observation - 1**

* In the research I took 80% of the dataset as training set and 20% of the dataset as the testset.
* Size of trainset = 1382 examples
* Size of testset = 346 examples
* After executing the testset using the Decision Tree that was trained on the trainset, a mean of 92.2832 % and standard deviation of 0.9753 % is found for an average of 20 runs.
* Runs = 20
  + Mean = 92.2832 %
  + Standard Deviation = 0.9753 %

**Observation - 2**

* Further after researching on the gradual learning characteristics of the Decision Tree, I found that if I increase the size of the training set, the performance also increases accordingly. From **68.50%** to **93.35%** for 1382 training examples.
* Starting from –
  + Training set size = 200 examples
* Test set size = 346 examples [**Fixed**]
* Training set size increment = 200 examples
* Percent Correct on Test Set determined by an average of 20 runs on per trainset

The performance graph ------

Report by,

**Apurbo Banik Turjo – 1905096**

Submitted on – 03.09.2023

**Topic:** Decision Tree Learning Algorithm on car dataset