**Project Report**

**Title:** Implementation of ID3 algorithm

**Authors:** Aashaar Panchalan, Priyadarshani Vasudevan

**Assumptions made :**

The user will provide valid paths for training, validation & testing datasets.

**Accomplishments :**

* Successfully created the decision tree
* Successfully calculated the accuracy
* Successfully increased the accuracy of test dataset –
  + for datasets 01: accuracy increased from 75.85% to 80.45% for pruning factor = 0.2
  + for datasets 02: accuracy increased from 72.33% to 76.16% for pruning factor = 0.07

**Best results :**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Pre-pruning accuracy\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
Number or training instances = 600  
Number of training attributes = 20  
Total number of nodes in the tree = 275  
Number of leaf nodes in the tree = 138  
Accuracy of the model on training dataset = 100.0

Number or validation instances = 2000  
Number of validation attributes = 20  
Accuracy of the model on validation dataset before pruning = 75.9

Number or testing instances = 2000  
Number of testing attributes = 20  
**Accuracy of the model on testing dataset = 75.85**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Post-pruning accuracy\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
Number or training instances = 600  
Number of training attributes = 20  
Total number of nodes in the tree = 171  
Number of leaf nodes in the tree = 86  
Accuracy of the model on training dataset = 92.16666666666666

Number or validation instances = 2000  
Number of validation attributes = 20  
Accuracy of the model on validation dataset after pruning = 80.85

Number or testing instances = 2000  
Number of testing attributes = 20  
**Accuracy of the model on testing dataset = 80.45**

**Learning:**

* Learnt programming in python
* Learnt the importance of the data structure tree in accomplishment of ID3 algorithm
* Analyzed the tree before and after pruning, which helped in understanding the concept of overfitting and increasing accuracy.

**Note:** Screenshots and complete output (with printed tree) of a run of the program are present in the folder ‘output screenshots’ for reference.