**REPORT**

**TEAM MEMBERS:**

**HARSH PANKAJKUMAR BHALANI (hpb170030)**

**MAITRI DHARMENDRAKUMAR SHAH (mxs172030)**

**Assumptions:**

* Log2 (0) = 0.
* If any leaf node contains data from 2 classes, i.e. if it is not pure node, then we are taking most frequent class label and give output according to that class label, i.e. 0 or 1.

**Best Results:**

Total number of nodes in the tree = 64

Number of leaf nodes in the tree = 33

Number of TrainingInstances = 600

Number of TrainingAttributes = 20

Accuracy of the model on training dataset = 62.83382%

Number of ValidationInstances = 2000

Number of ValidationAttributes = 20

Accuracy of the model on validation dataset = 61.75%

Number of TestingInstances = 2000

Number of TestingAttributes = 0

Accuracy of the model on testing dataset = 60.8%

**What was accomplished:**

We built an effective classifier using ID3 algorithm that learns from training dataset and classifies new testing data correctly.

Pruning the decision tree increases the accuracy for some of the instances, because it takes random number of nodes to prune from the main decision tree.

**What was learned:**

We learnt that ID3 algorithm splits a data set on the attribute with least entropy, reducing the amount of randomness in the decision-making process.

We learned the working of ID3 algorithm and how it helps in choosing the most

effective attributes, thereby, aiding in building the decision tree accordingly.

We also observed that the algorithm learns and performs better when more training data is provided.