CS 6375

ASSIGNMENT \_\_\_\_\_\_2\_\_\_\_\_\_

Names of students in your group:

Pranathi Peri(pxp162530)

Nithin Vasireddy(nxv160230)

Number of free late days used: \_\_0\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_   
Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.

Report

Assumptions:

The following are the assumptions

1)Log20 is taken as 0

2)The nodes to be pruned are selected using a random number generator. The

random number generated may be 1 which indicates the root node. If there is

root node in the randomly generated numbers then we have generated the

random numbers again.

3)While pruning we have seen to that the leaf nodes are not been pruned.

Because if leaf nodes are pruned the same class remains there and it will not

change the tree structure and accuracy.

4)If the node is not pure, then we chose the most frequent class in that node and

output that as the predicted class.

5) If a leaf node contains no data instances, we have labeled the left child as class0

and right child as class1

6)The decision tree tries to improve the validation accuracy by trying the different

combinations of the random nodes. If the validation accuracy does not increase

it gives the best accuracy among them as the accuracy depends both on the

pruning factor and the random number generator function.

Best Results:

If the input is an effective pruning factor, then the decision tree tries to produce the pruned decision tree by selecting the best random nodes which increase the validation accuracy of the decision tree.

If the input is non-effective pruning factor, then decision tree runs for a fixed number of times to find the best accuracy. If the accuracy does not increase after a fixed number of times it gives the best one among them, as the accuracy depends both on the pruning factor and the random number generator function.

**Accomplishments and Learning from this assignment:**

We have learnt about ID3 algorithm.

We have learnt how crucial is pruning factor to a machine learning algorithm.

We have designed and implemented our first Machine Learning algorithm – ID3 on real world data set.

We have implemented the ID3 algorithm in a way that tries to improve the validation accuracy after pruning .