CS 6375

ASSIGNMENT 2

1. Names of students in the group:
   1. Anuj Audhesh Singh
   2. Shubham Pradeep Kothari
2. Number of free late days used: 1
3. Sources/references that we have used in this assignment are:
4. ID3 algorithm Wikipedia: <https://en.wikipedia.org/wiki/ID3_algorithm>
5. Short paper: <http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm>
6. Indexing and Selecting of Data with Pandas:

<https://chrisalbon.com/python/pandas_indexing_selecting.html>

1. Debugging and error handling: <https://stackoverflow.com>
2. Chapter 3 of Tom Mitchell's book
3. Assumptions:
   1. The dataset has Boolean attributes and Boolean class labels (0 or 1).
   2. There are no missing data and attributes.
   3. The first row of the dataset contains column names and each non-blank line after that contains a new data instance.
   4. The last column contains the class labels.
   5. We assume that log20 = 0
   6. If a leaf node contains data from more than one class, then we have chosen the most frequent class in that node and outputted it. If the leaf node has equal instances of 0 and 1 class labels, we return 1 as the leaf node value.
4. Learning from this assignment: We learnt to use pandas library for data preprocessing. We built the decision tree using ID3 algorithm on the training dataset. Then we printed it as per the given instructions. Below, we have attached a screenshot of the output.

A screenshot of a cell phone

Description generated with high confidence