Decision-Tree-Builder

Builds a decision tree to predict number of deaths using COVID data from dataset-C

Table of Contents

Requirements

Running this script requires python3+ and pandas,math,random,copy libraries of python3 to be installed on running machine

Setup

Run on command line

```
sudo apt-get update
sudo apt-get install python3.6
pip3 install pandas
pip3 install random
pip3 install math
pip3 install copy
```

This will install all required libraries to run this script on the machine If few of the libraries are already installed,run only the required commands from above

Running the Program

Run on command line

```
python3 script.py
```

By default,the program executes all four parts asked in question

Part-1

It first asks for an integer input denoting maximum depth used to build the tree It outputs the train and test dataset accuracy of built decision trees for 10 different random splits of dataset and then the average train and test dataset accuracy and the seed for which maximum test accuracy is achieved

Part-2

Prints the bestDepth of tree for which best test accuracy is achieved among all possible depths as well as the corresponding seed of that split

Part-3

Performs the pruning operation on tree obtained in previous step

Part-4

Prints the tree after pruning Format: Each line denotes a node - may be internal node or a leaf node Each line contains : (Leaf node or Internal node,Depth of node with root indexed as 0,Label of attribute considered to reach this node from its parent,maxGainAttribute for internal node or deaths for leaf node)

Running each part seperately

Open source.py provided in folder, go to the main function(at the bottom of the file) and comment out corresponding lines to check results of only individual parts.(Each part is described in comments in the file)
Part-1 : Comment out Parts 2-4
Part-2 : Comment out Part-1 and Parts 3-4
Part-3 : Comment out Part-1,4 and only print statements in Part-2
Part-4 : Comment out only Part-1 and all print statements in Parts 2-3