

**REPORT**  
**Assignment 04**  
**CSE318**

name: Aveerup Chowdhury  
id: 2105112

**Iris.csv:**

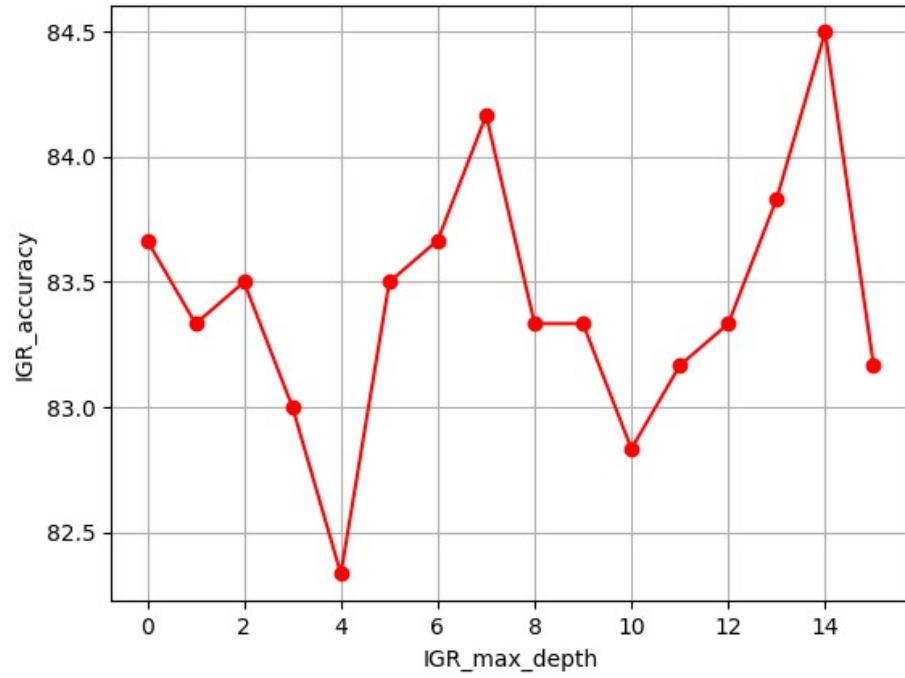
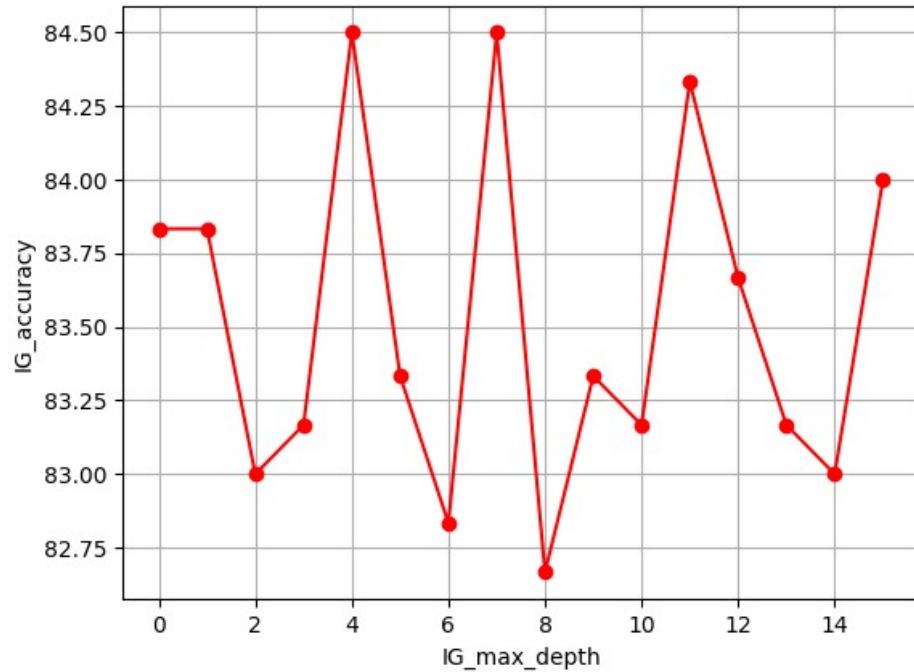
This dataset has less data. This makes it hard for the decision tree to come to a accurate conclusion if there is any mismatch in the given data. This trend can be seen in the graphs provided below.

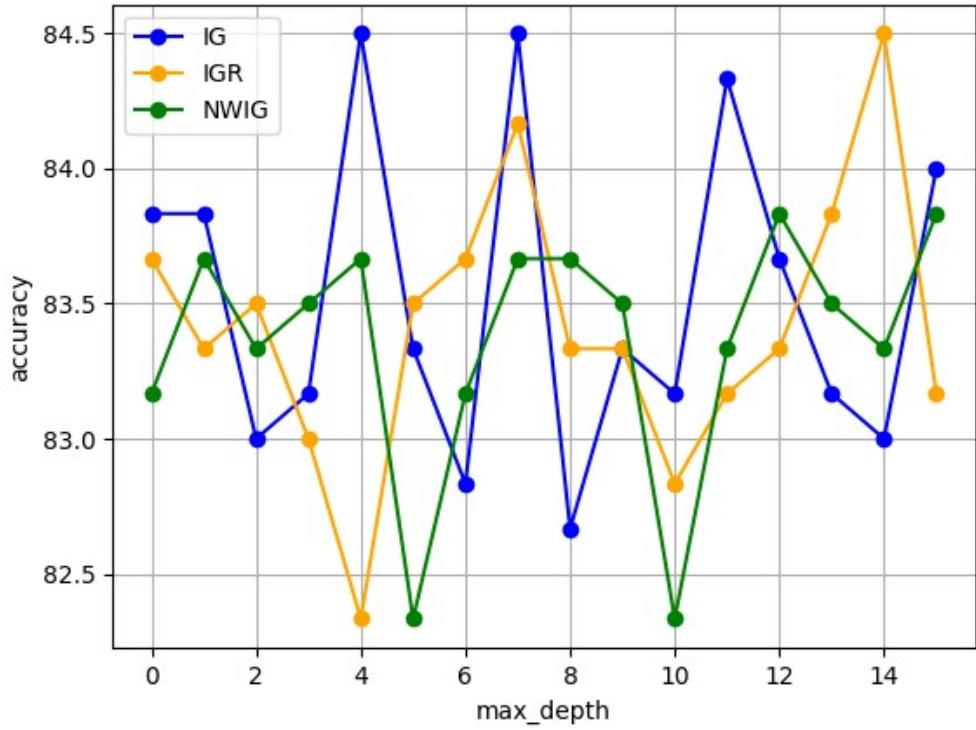
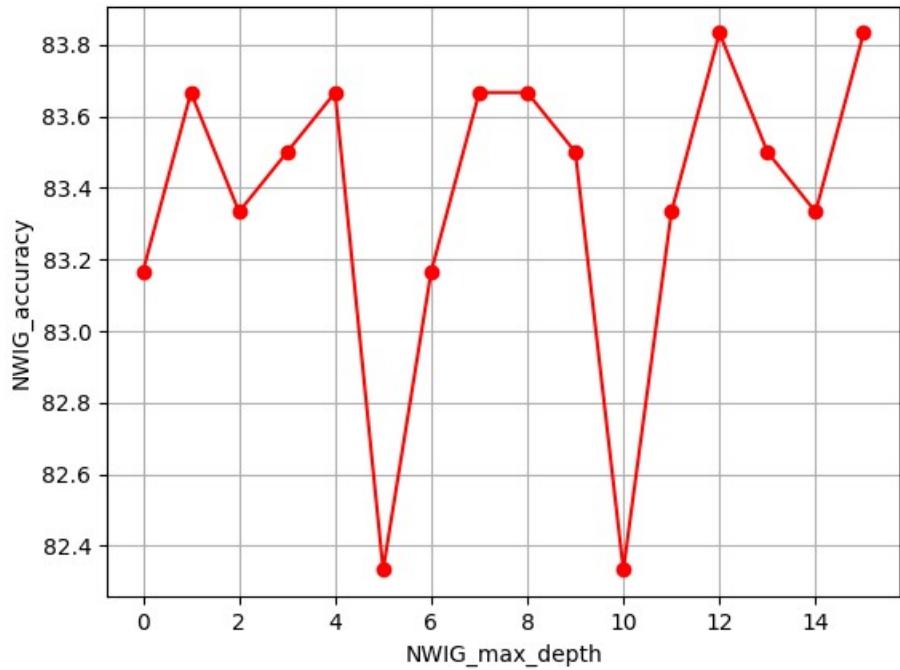
The accuracy vs max\_depth graphs for different criterions are fluctuating even when we are using a higher depth. Because even though depth is increasing, small datasets in iris reaches it's full potential depth after a certain iteration where full-tree has less depth than the given max\_depth. This makes our max\_depth enabling ineffective.

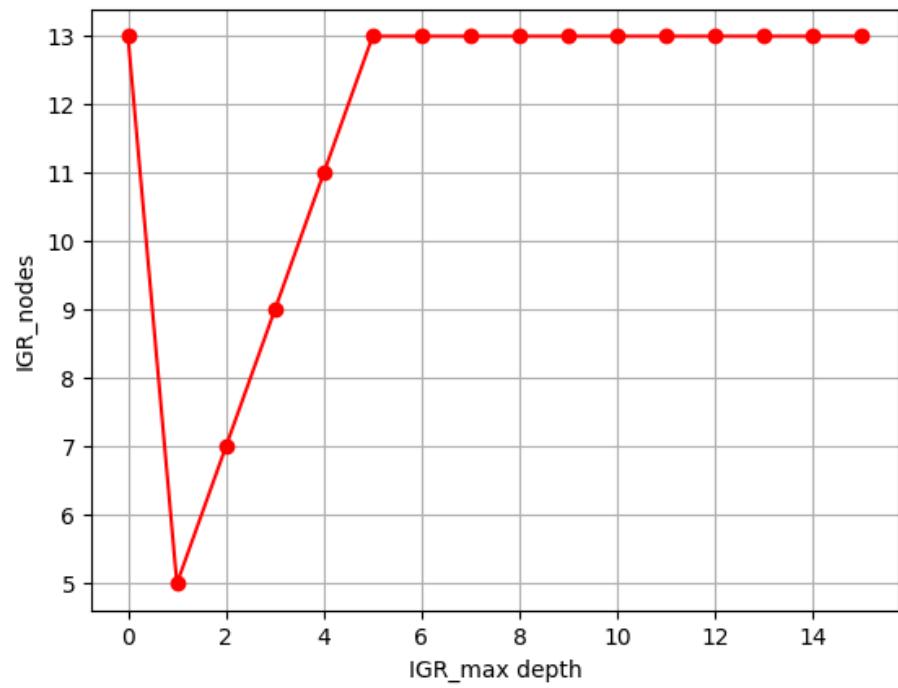
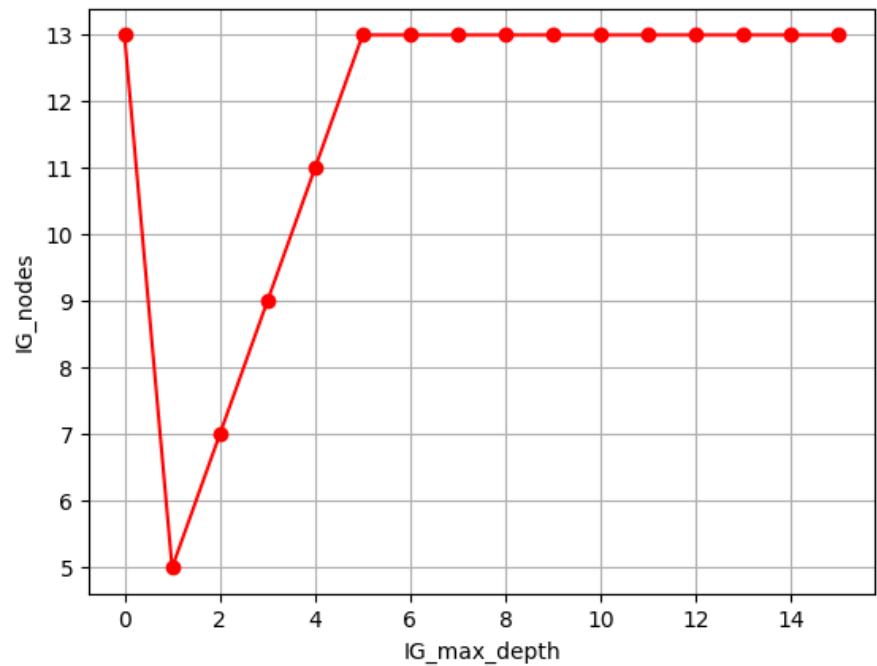
This ineffectiveness of increasing max\_depth can be more visible in the nodes vs max\_depth graphs. When the full tree depth is less than the max\_depth, the tree spreads full each time, making the node number same.

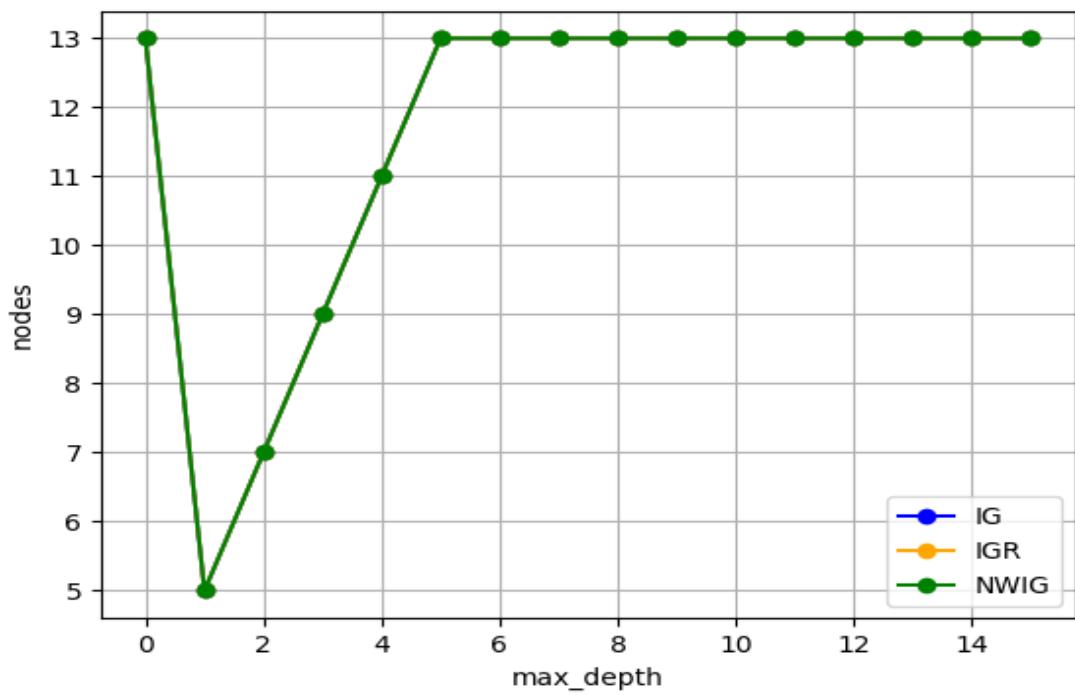
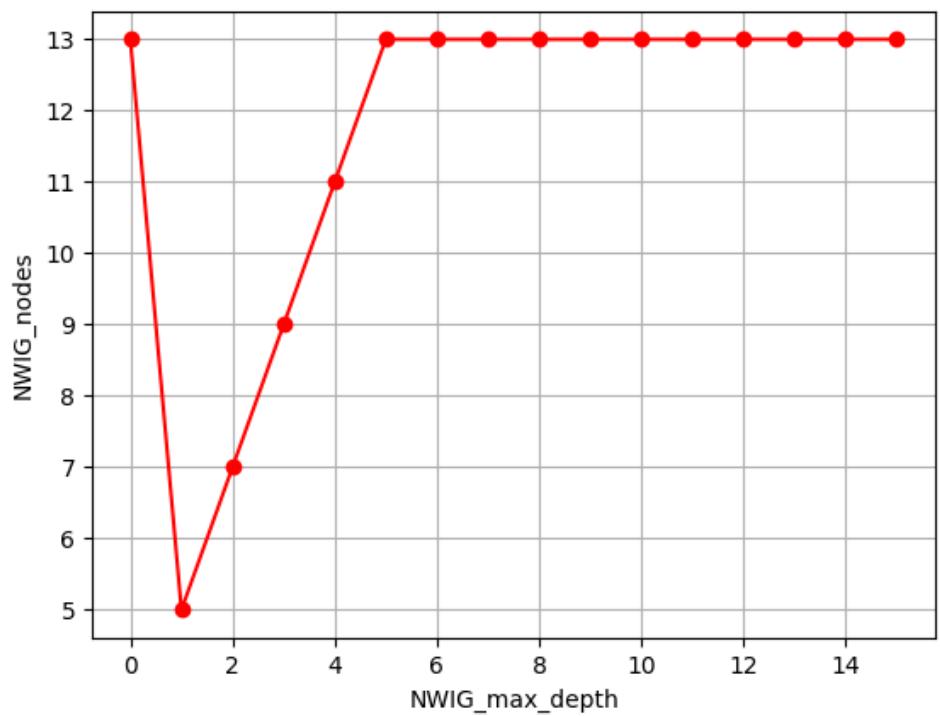
Among the criterions NWIG is the one giving the consistent accracy and IG is the one that is fluctuating the most. This is because IG divides whenever it finds the maximum information gain, not caring about the heavy node numbers. IGR is somewhat between, but it's fluctuation is still as high as IG.

hello









## adult.data:

In adult.data the graphs are more stable than iris.csv. Because the dataset is big. And even if some data are wrong it doesn't have any impact on the overall result.

We can see that NWIG has the most accurate predictions, second one is IG and in third place IGR. But In case of consistency IGR is the top one. IGR fluctuated the least.

In nodes, everyone follows a trend. The depth 0 or full tree has the highest node. And as the max-depth increases the node number gradually increases but doesn't reach the full potential of the tree.

