REPORT
Assignment 04
CSE318

name: Aveerup Chowdhury
id: 2105112

Iris.csv:

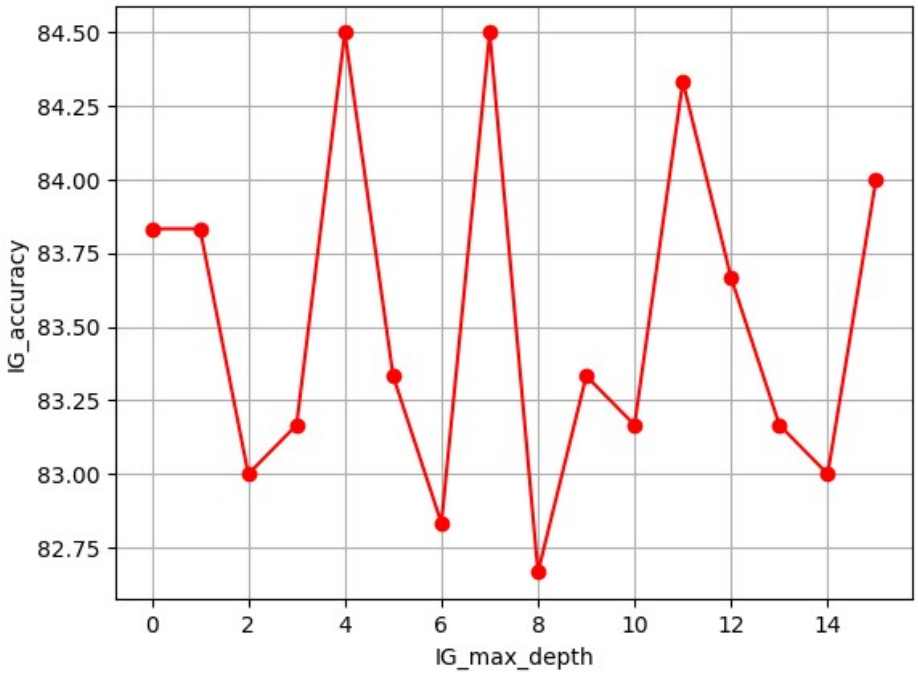
This dataset has less data. This makes it hard for the decision tree to come to a accurate conclusion if there is any mismatch in the given data. This trend can be seen in the graphs provided below.

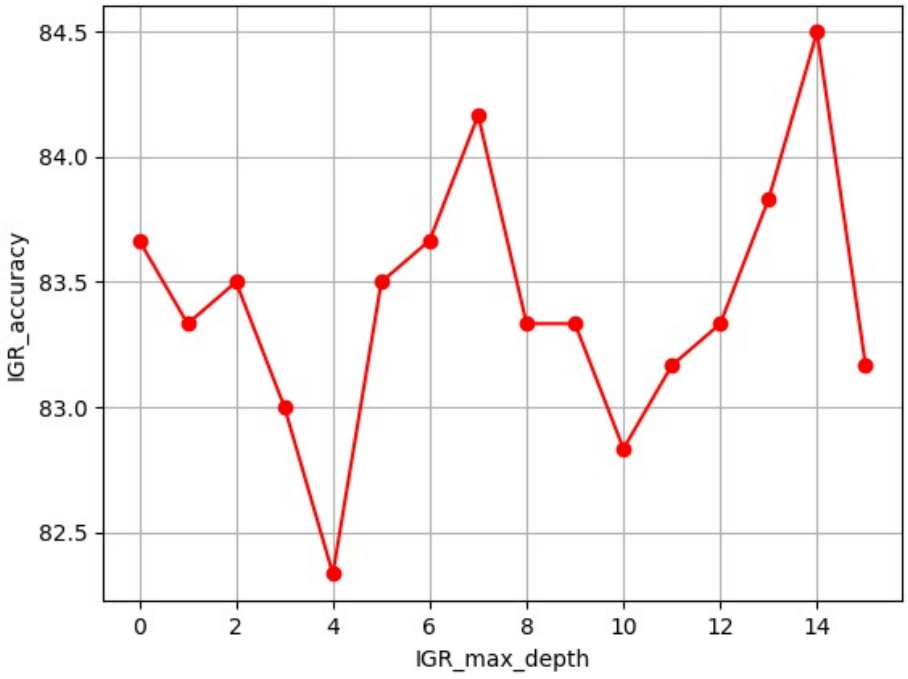
The accuracy vs max_depth graphs for different criterions are fluctuating even when we are using a higher depth. Because even though depth is increasing, small datasets in iris reaches it's full potential depth after a certain iteration where full-tree has less depth than the given max_depth. This makes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

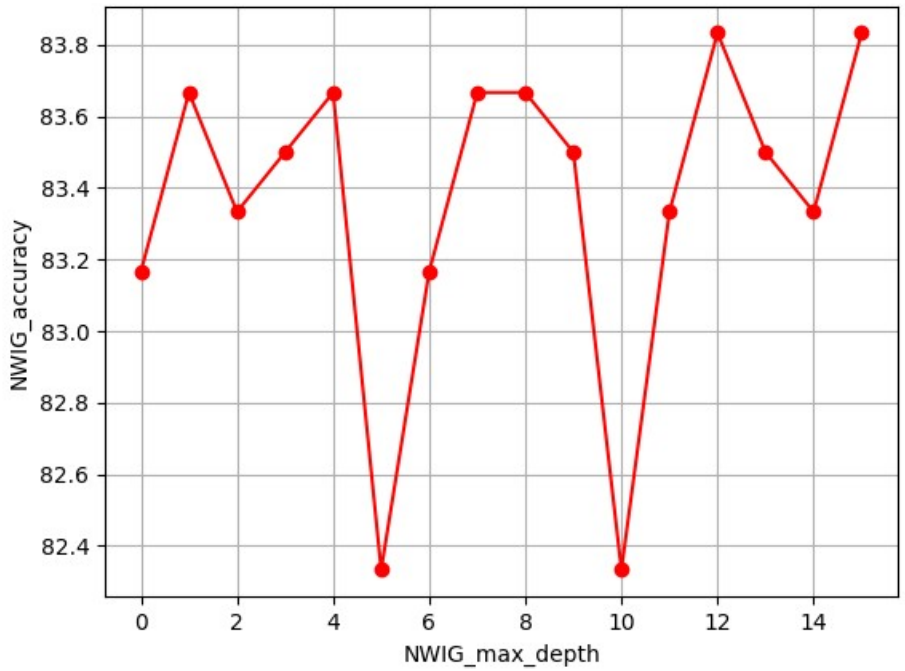
This ineffectiveness of increasing max_depth can be more visible in the nodes vs max_depth graphs. When the full tree depth is less than the max_depth, the tree spreads full each time, making the node number same.

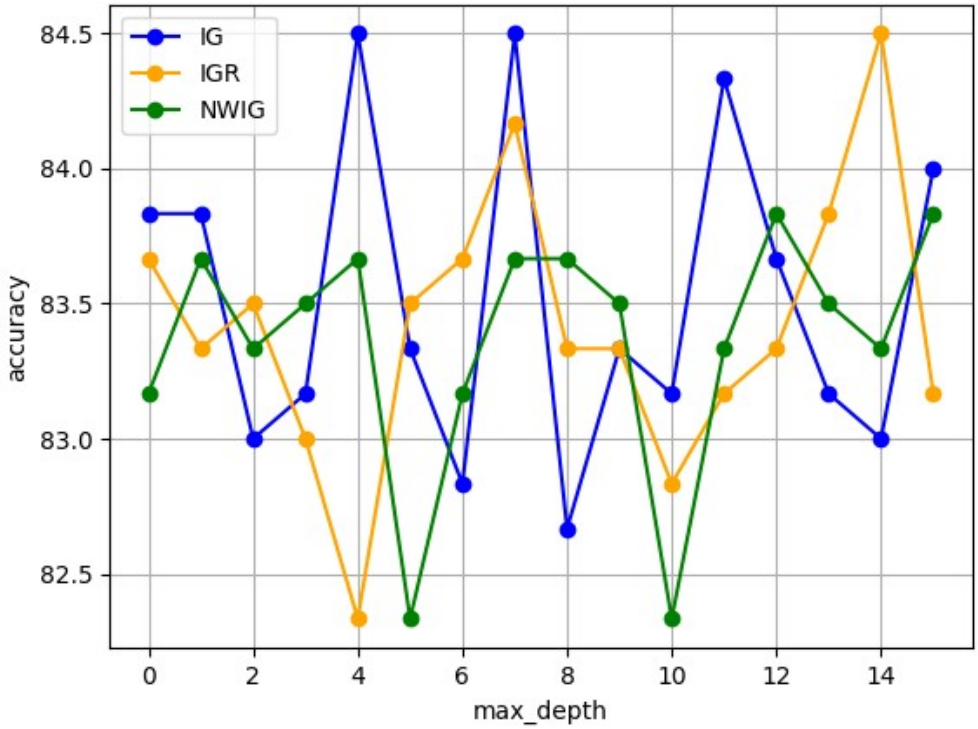
Among the criterions NWIG is the one giving the consistent accracy and IG is the one that is fluctuating the most. This is because IG divides whenever it finds the maximum information gain, not caring about the heavy node numbers. IGR is somewhat between, but it's fluctuation is still as high as IG.

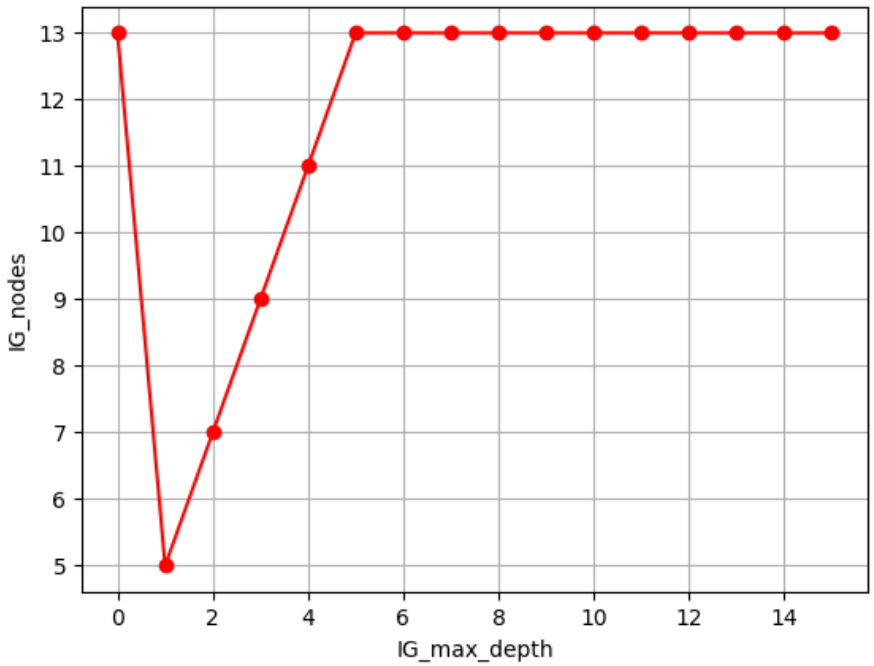
hello

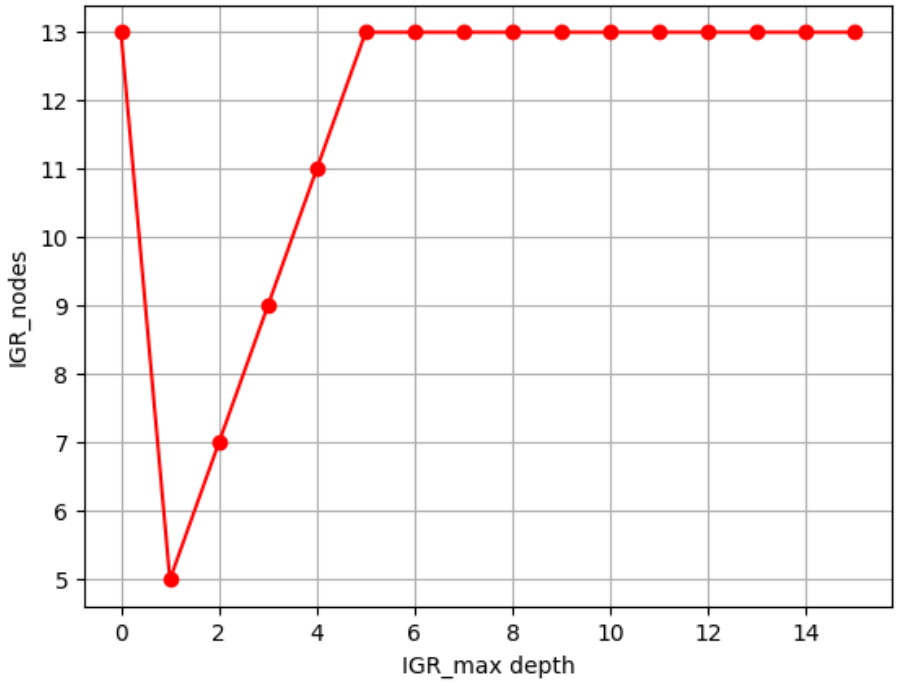

84.50
84.25
84.00
83.75
83.50
83.25
83.00
82.75
IG_accuracy
0
2
4
6
8
10
12
14
IG_max_depth

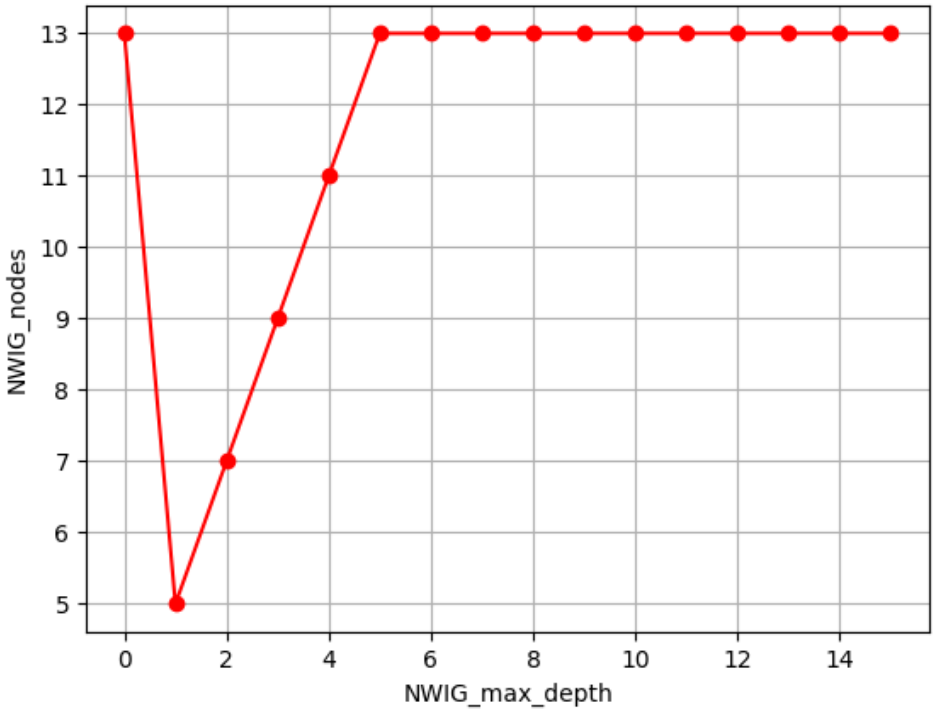

84.5
84.0
83.5
83.0
82.5
IGR_accuracy
0
2
4
6
8
10
12
14
IGR_max_depth

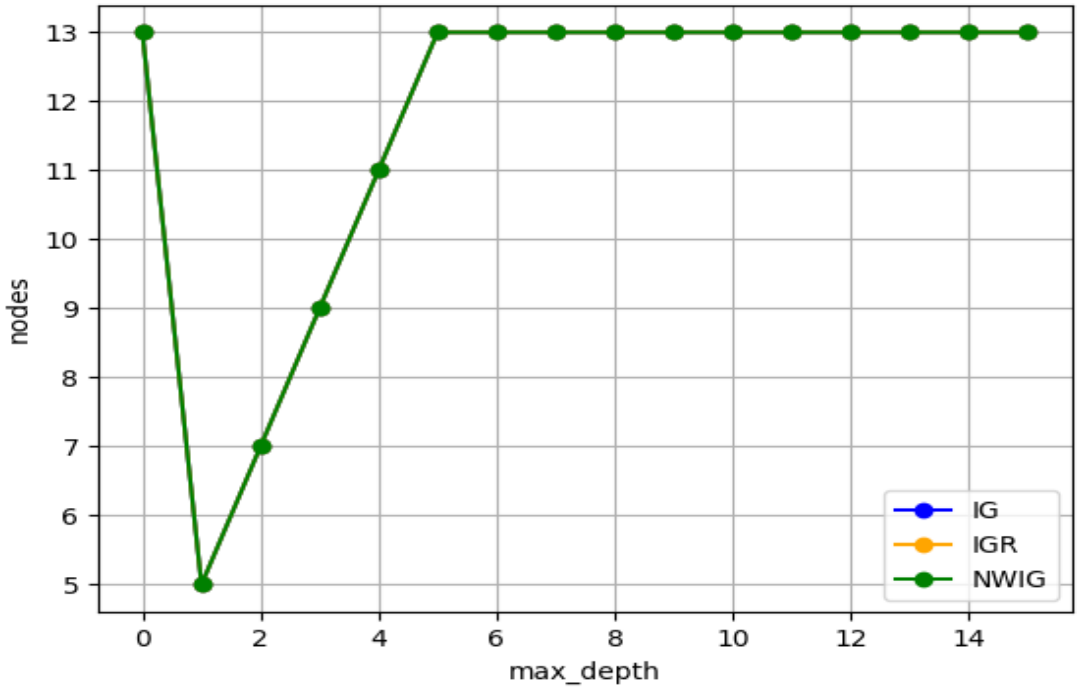

NWIG_accuracy
NWIG_max_depth
83.8
83.6
83.4
83.2
83.0
82.8
82.6
82.4
0
2
4
6
8
10
12
14


IG
IGR
NWIG
accuracy
max_depth
84.5
84.0
83.5
83.0
82.5
0
2
4
6
8
10
12
14


IG_nodes
IG_max_depth
13
12
11
10
9
8
7
6
5
0
2
4
6
8
10
12
14


IGR_nodes
IGR_max depth
13
12
11
10
9
8
7
6
5
0
2
4
6
8
10
12
14


NWIG_nodes
NWIG_max_depth
13
12
11
10
9
8
7
6
5
0
2
4
6
8
10
12
14

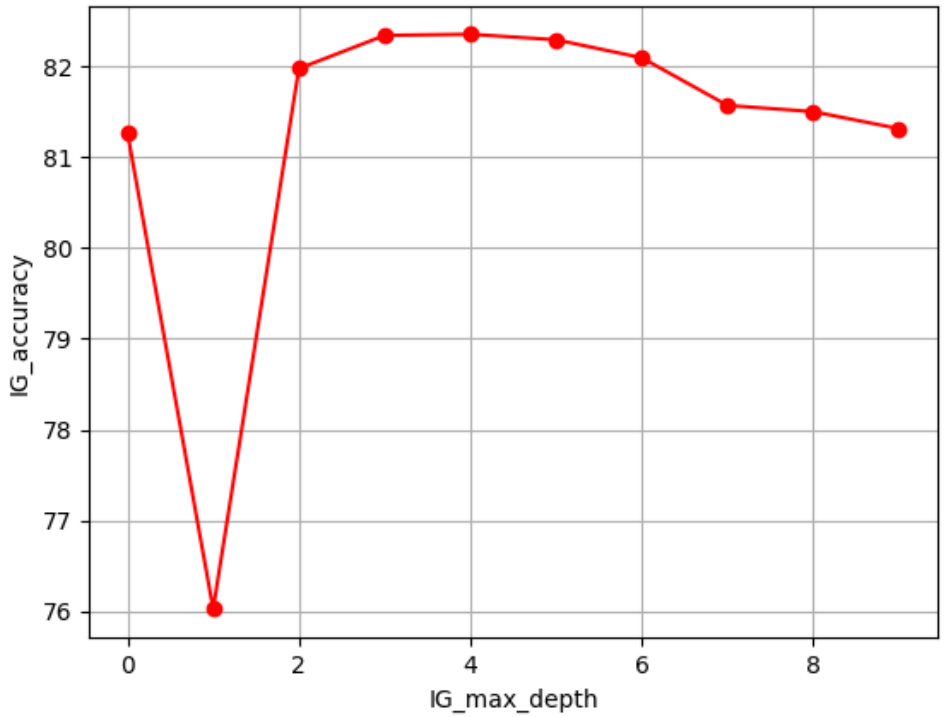

nodes
max_depth
IG
IGR
NWIG
13
12
11
10
9
8
7
6
5
0
2
4
6
8
10
12
14

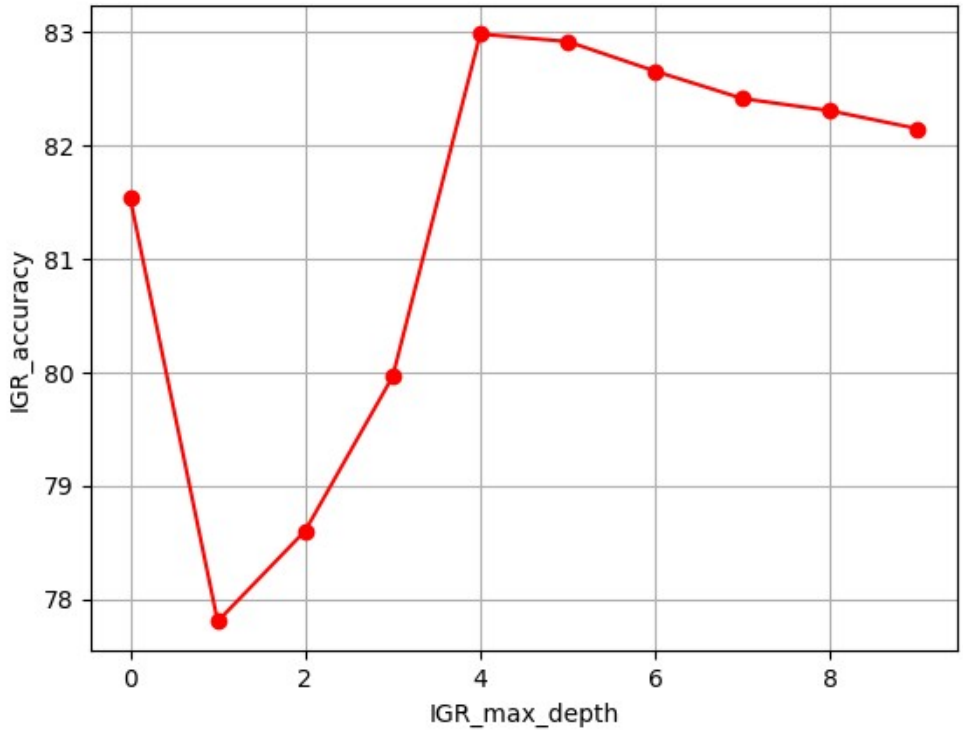
adult.data:

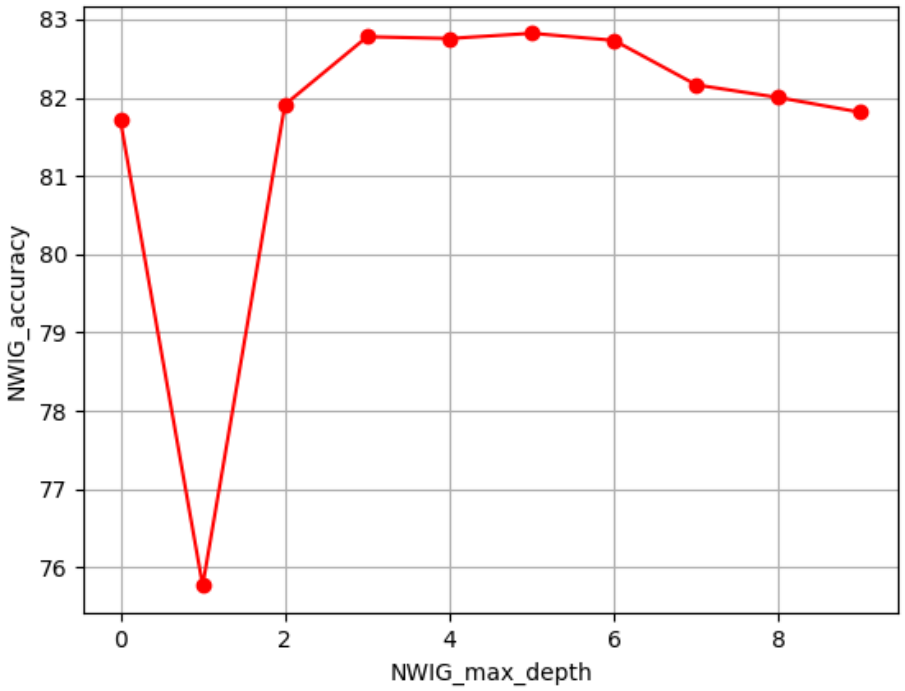
In adult.data the graphs are more stable than iris.csv. Because the dataset is big. And even if some data are wrong it doesn’t have any impact on the overall result.

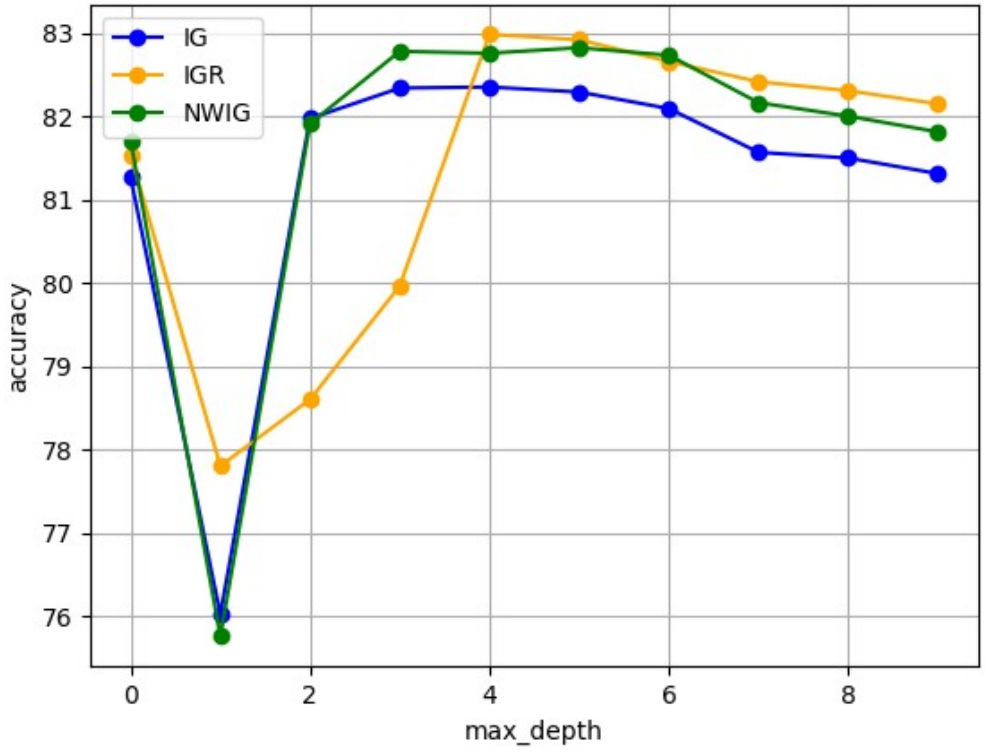
We can see that NWIG has the most accurate predictions, second one is IG and in third place IGR. But In case of consistensy IGR is the top one. IGR fluctuated the least.

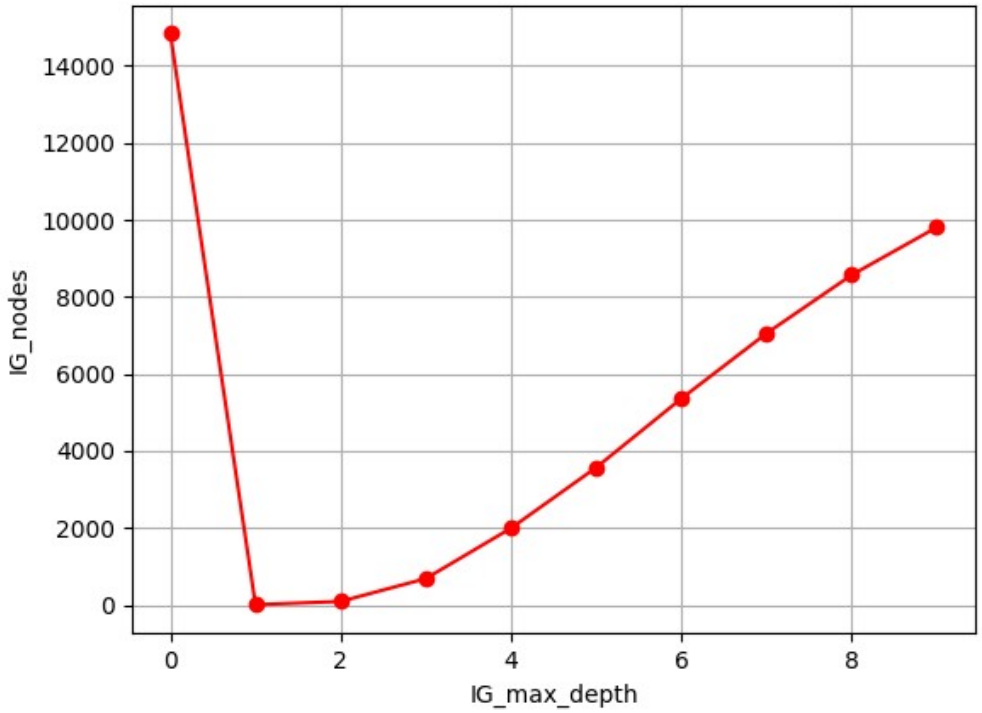
In nodes, everyone follows a trend. The depth 0 or full tree has the highest node. And as the max-depth increases the node number gradually increases but doesn’t reach the full potential of the tree.

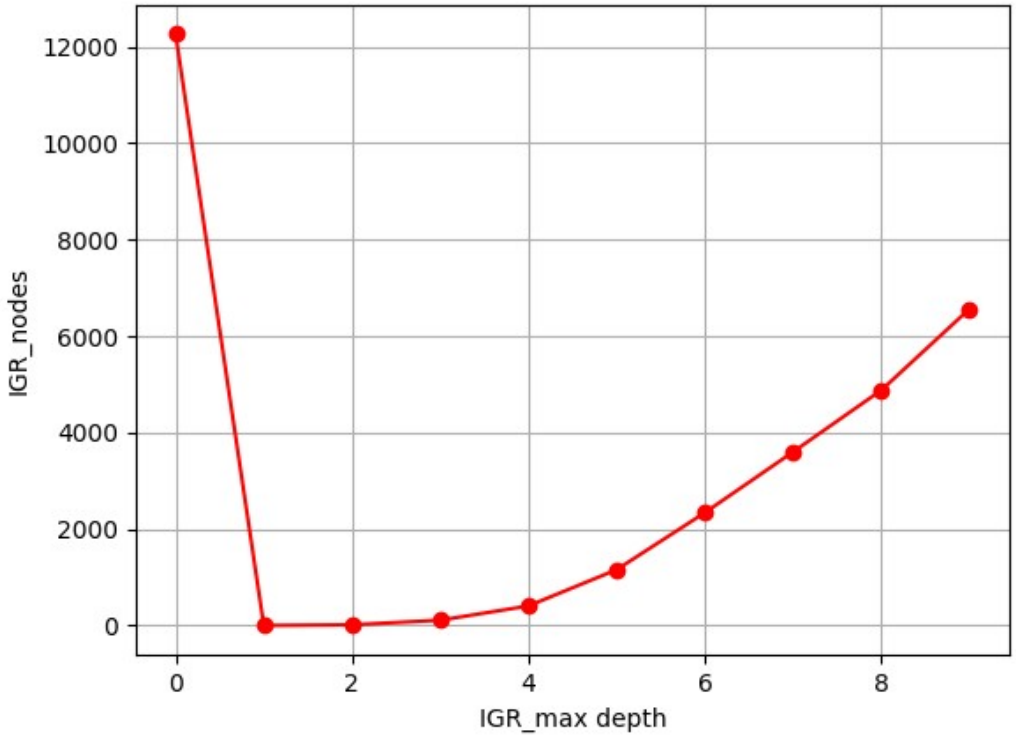

IG_accuracy
82
81
80
79
78
77
76
0
2
4
6
8
IG_max_depth

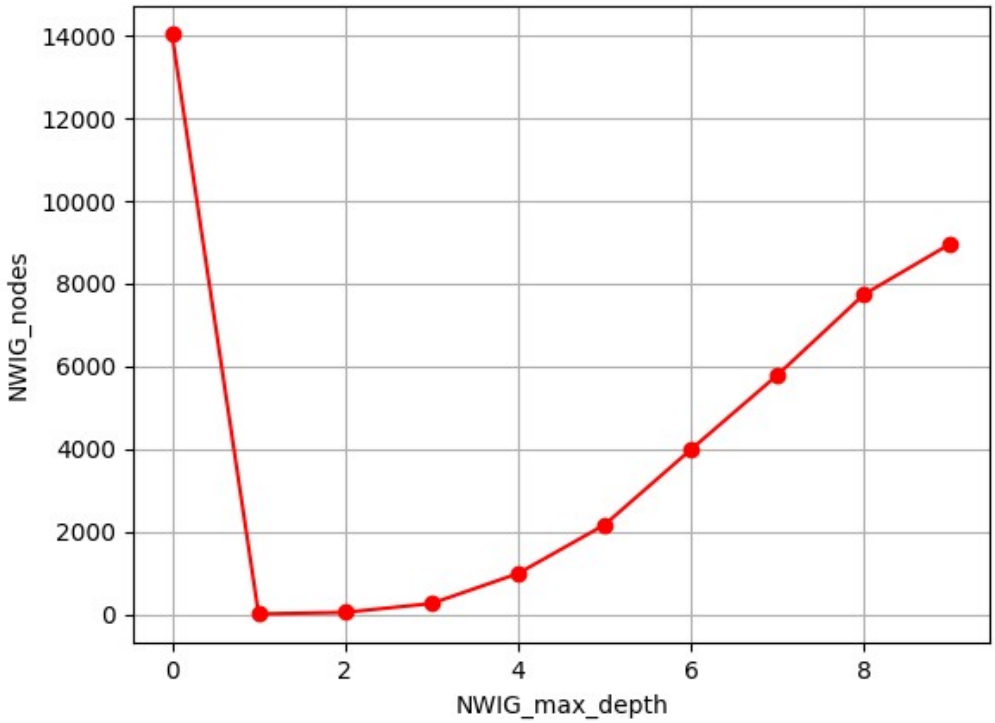

IGR_accuracy
83
82
81
80
79
78
0
2
4
6
8
IGR_max_depth

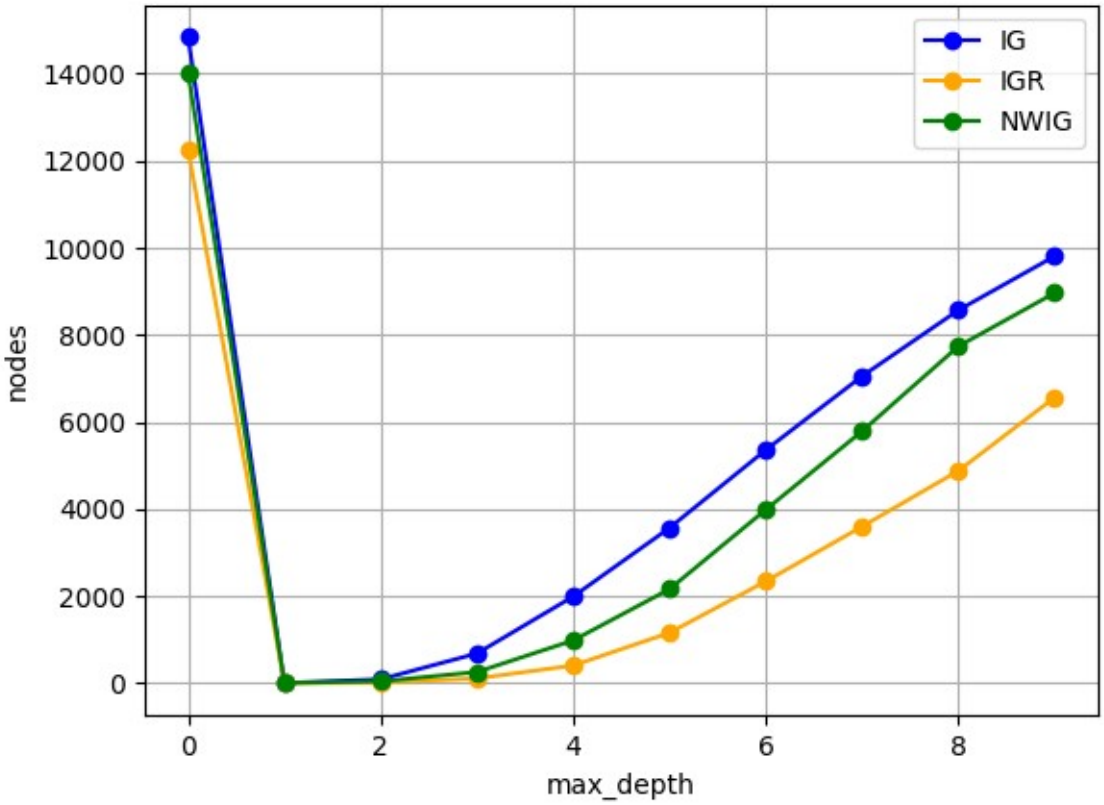

NWIG_accuracy
NWIG_max_depth
83
82
81
80
79
78
77
76
0
2
4
6
8


IG
IGR
NWIG
accuracy
max_depth
83
82
81
80
79
78
77
76
0
2
4
6
8


14000
12000
10000
8000
6000
4000
2000
0
IG_nodes
0
2
4
6
8
IG_max_depth


12000
10000
8000
6000
4000
2000
0
IGR_nodes
0
2
4
6
8
IGR_max depth


14000
12000
10000
8000
6000
4000
2000
0
NWIG_nodes
0
2
4
6
8
NWIG_max_depth


IG
IGR
NWIG
14000
12000
10000
8000
6000
4000
2000
0
nodes
0
2
4
6
8
max_depth