Final Year B. Tech. (CSE) – I : 2021-22
4CS462 : PE2 - Data Mining Lab
Assignment No. 5

Group id: DM21G12
Group members:
Abhishek More(2018BTECS00037)
Sushil Wagh(2018BTECS00031)

Title : Design the rule based classifier : Extract the rules from decision tree build in assignment no. 4.

Objective/Aim :
Tabulate the results and evaluate the performance of rules generated using following metrics :

a. Coverage
b. Accuracy
C. Toughness (size)

Use the following categorical data sets from UCI machine learning repository :

a. Balance Scale data set
b. Car evaluation data set
C. Breast-cancer data se

Introduction:

Coverage of a rule:
The percentage of instances that satisfy the antecedent of a rule (i.e., whose attribute values hold true for the rule's antecedent).

Accuracy of a rule:
Rule-Based Classification
The percentage of instances that satisfy both the antecedent and consequent of a rule

- **Rule accuracy and coverage:**

$$coverage(R) = \frac{n_{covers}}{|D|}$$

$$accuracy(R) = \frac{n_{correct}}{n_{covers}}$$

- *where*
  - *D*: class labeled data set
  - |D|: number of instances in *D*
  - n_{covers} : number of instances covered by R
  - $n_{correct}$: number of instances correctly classified by R

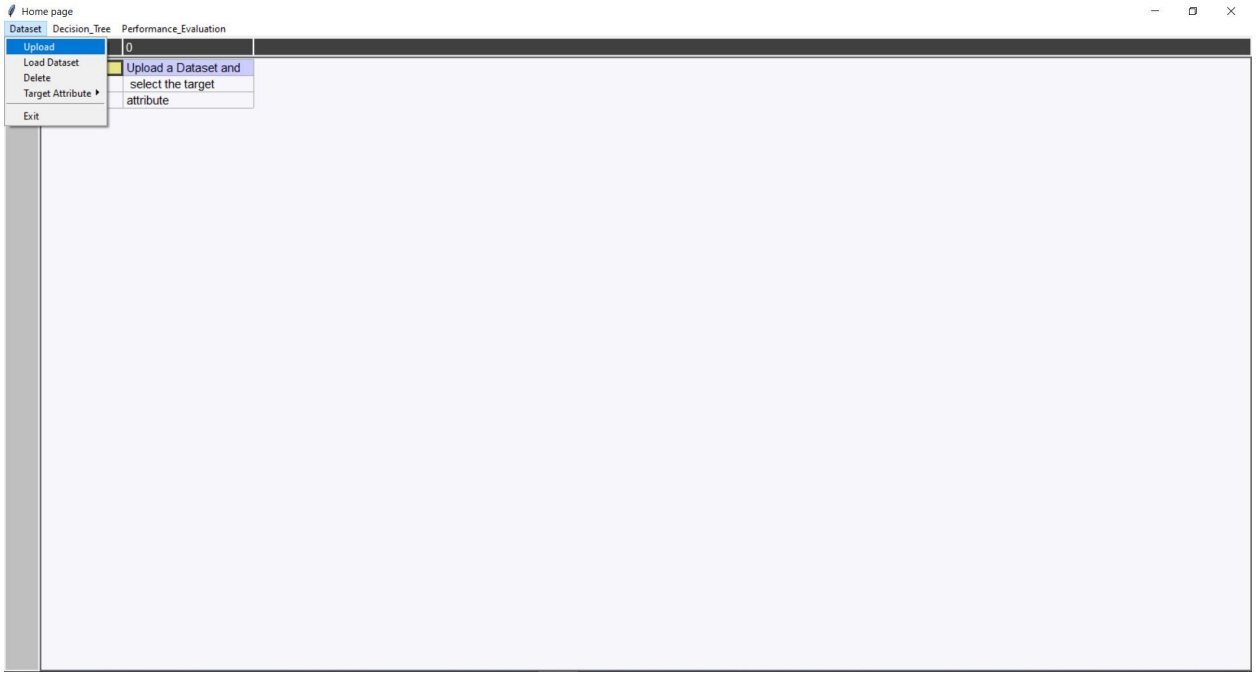
Coverage and Accuracy

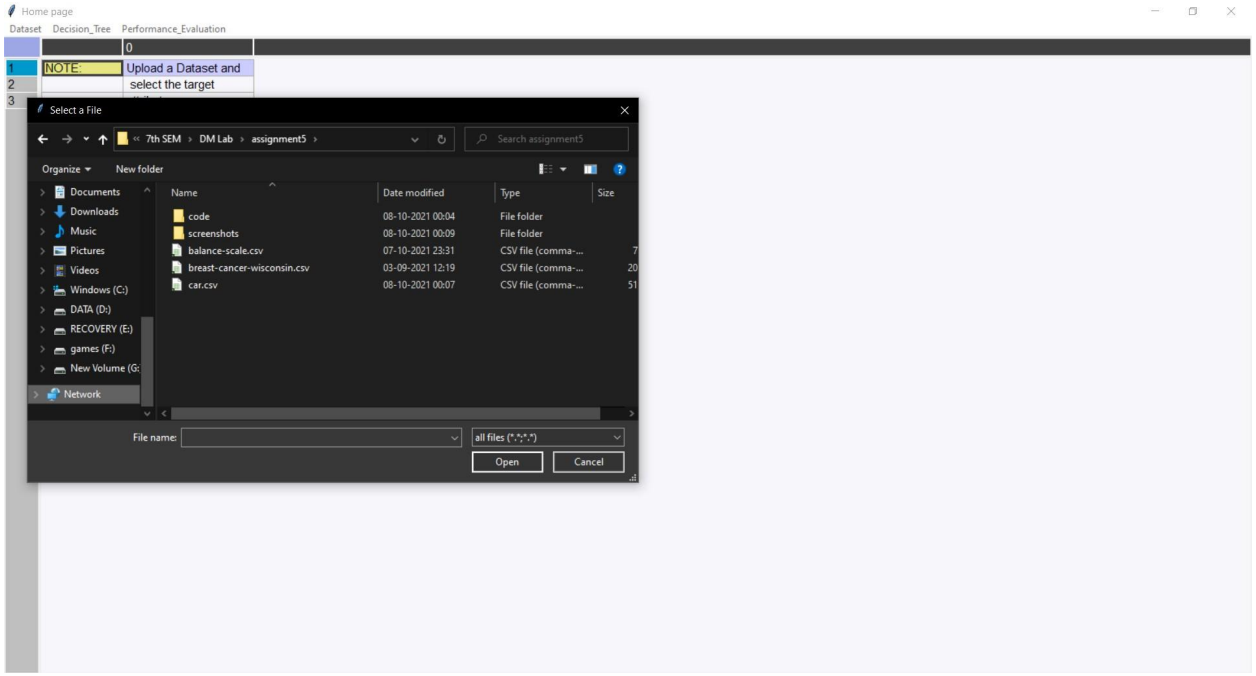
- The rule R1:

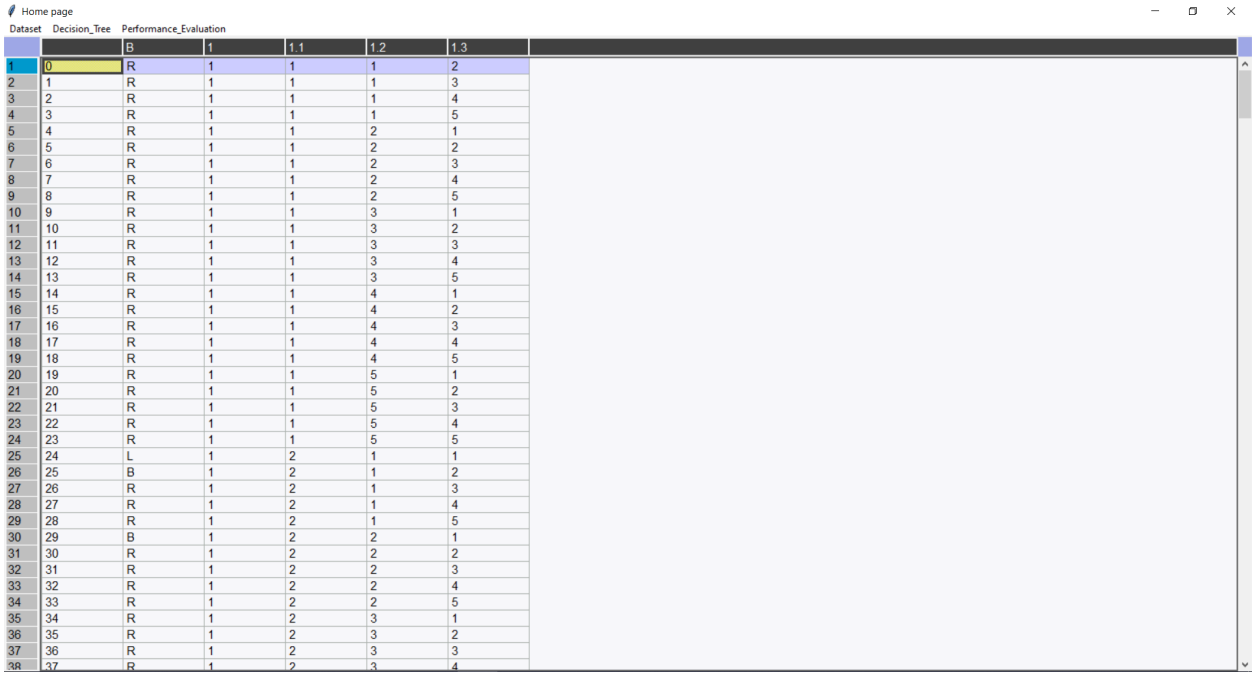
R1: IF *age* = *youth* AND *student* = *yes* THEN *buys_computer* = *yes*

  - R1 covers 2 of the 14 instances
  - It can correctly classify both instances
- Therefore:
  - *Coverage*(*R*1) = 2/14 = 14.28%
  - *Accuracy*(*R*1) = 2/2 = 100%.

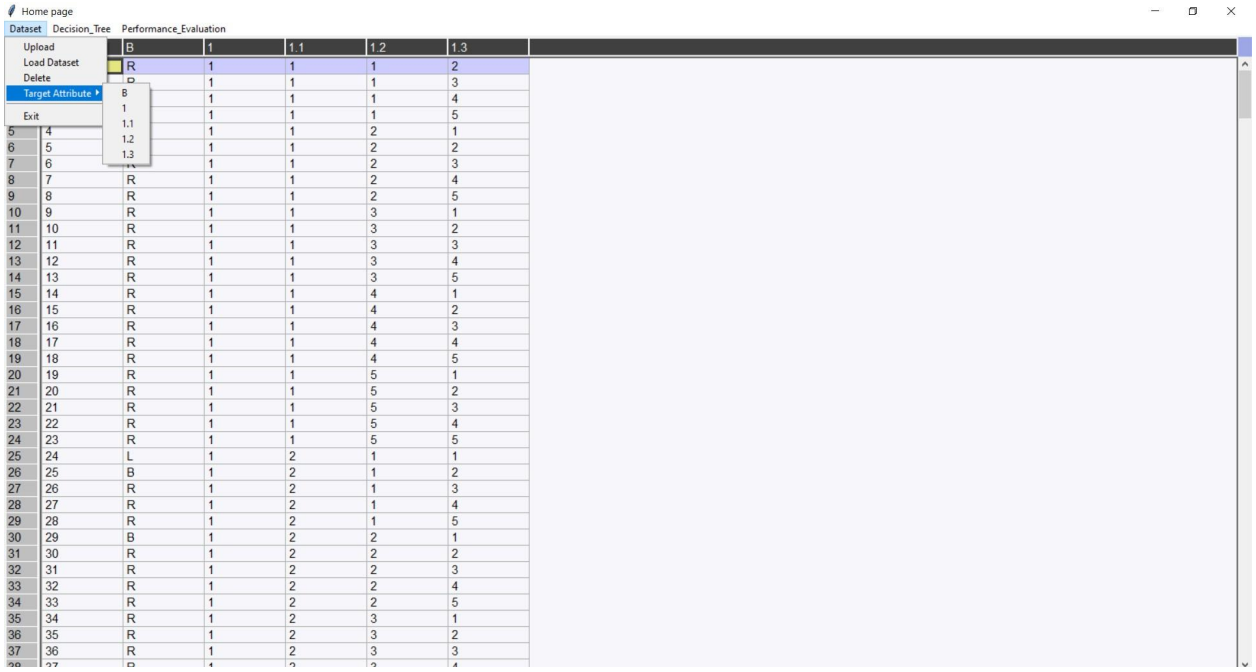
Result/Observations/Screenshots:


Home page
Dataset
Decision_Tree
Performance_Evaluation
Upload
Load Dataset
Delete
Target Attribute
Exit
0
Upload a Dataset and
select the target
attribute

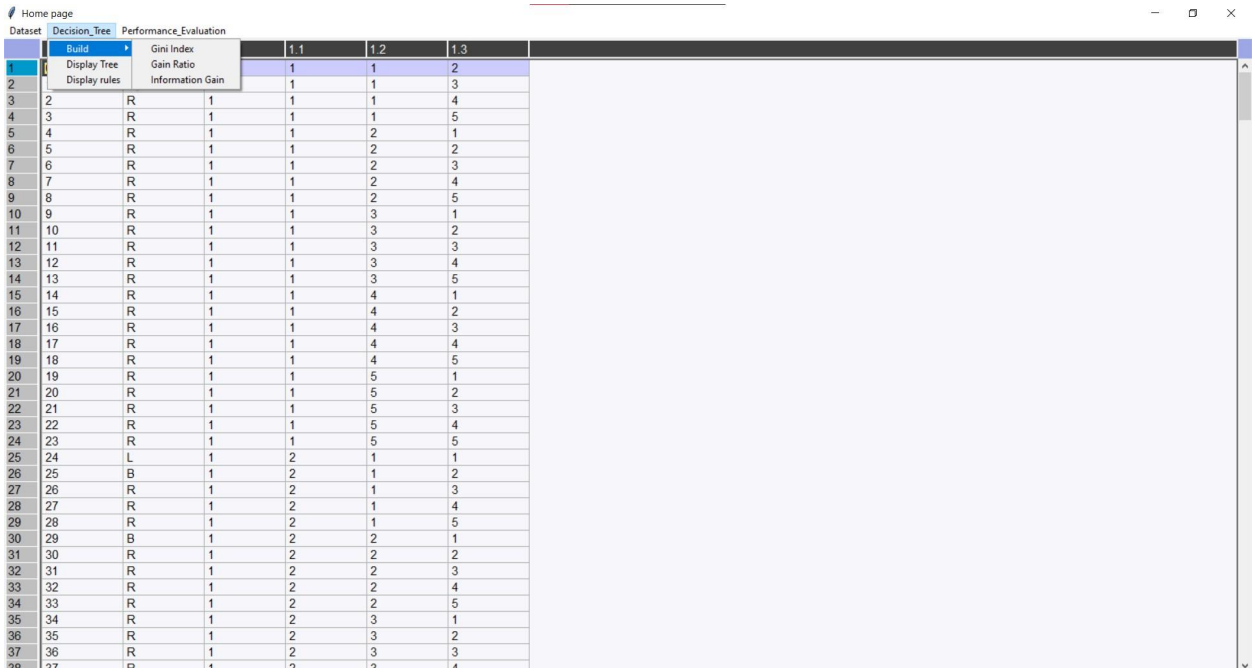

Home page
Dataset
Decision_Tree
Performance_Evaluation
0
1
NOTE:
Upload a Dataset and
2
select the target
3
Select a File
7th SEM
DM Lab
assignment5
Search assignment5
Organize
New folder
Documents
Downloads
Music
Pictures
Videos
Windows (C:)
DATA (D:)
RECOVERY (E:)
games (F:)
New Volume (G:
Network
Name
Date modified
Type
Size
code
08-10-2021 00:04
File folder
screenshots
08-10-2021 00:09
File folder
balance-scale.csv
07-10-2021 23:31
CSV file (comma-...
7
breast-cancer-wisconsin.csv
03-09-2021 12:19
CSV file (comma-...
20
car.csv
08-10-2021 00:07
CSV file (comma-...
51
File name:
all files (*.*;*.*)
Open
Cancel

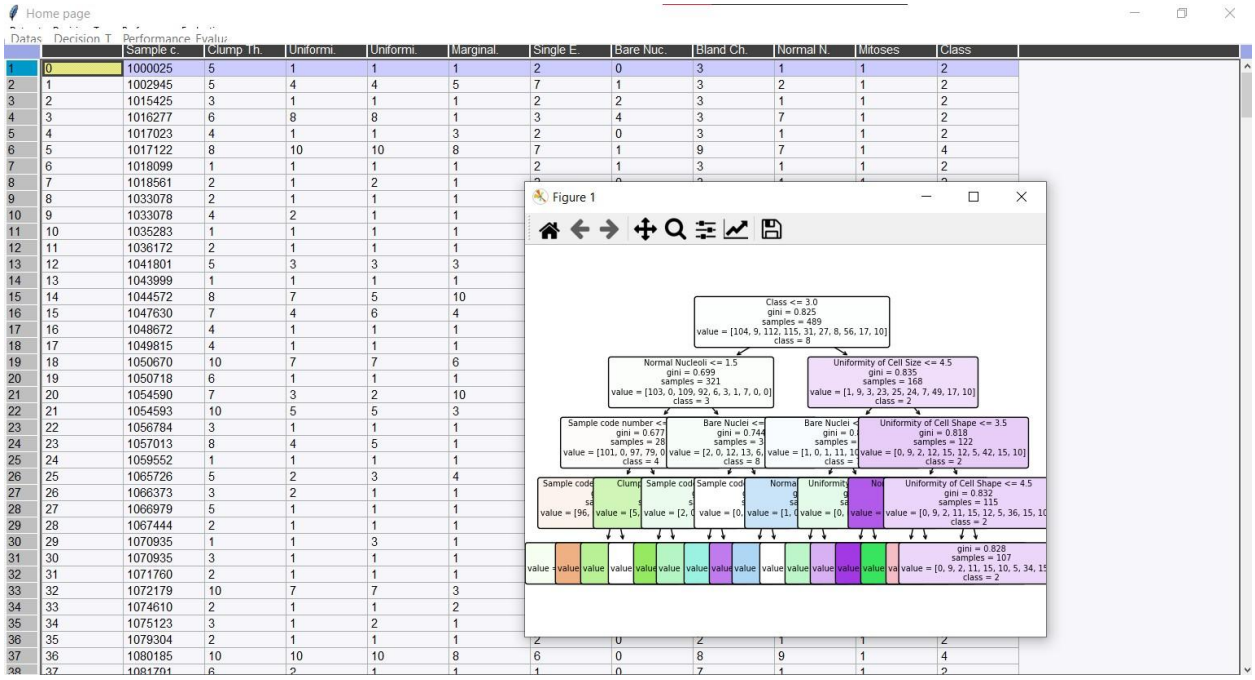


		B	1	1.1	1.2	1.3
1	0	R	1	1	1	2
2	1	R	1	1	1	3
3	2	R	1	1	1	4
4	3	R	1	1	1	5
5	4	R	1	1	2	1
6	5	R	1	1	2	2
7	6	R	1	1	2	3
8	7	R	1	1	2	4
9	8	R	1	1	2	5
10	9	R	1	1	3	1
11	10	R	1	1	3	2
12	11	R	1	1	3	3
13	12	R	1	1	3	4
14	13	R	1	1	3	5
15	14	R	1	1	4	1
16	15	R	1	1	4	2
17	16	R	1	1	4	3
18	17	R	1	1	4	4
19	18	R	1	1	4	5
20	19	R	1	1	5	1
21	20	R	1	1	5	2
22	21	R	1	1	5	3
23	22	R	1	1	5	4
24	23	R	1	1	5	5
25	24	L	1	2	1	1
26	25	B	1	2	1	2
27	26	R	1	2	1	3
28	27	R	1	2	1	4
29	28	R	1	2	1	5
30	29	B	1	2	2	1
31	30	R	1	2	2	2
32	31	R	1	2	2	3
33	32	R	1	2	2	4
34	33	R	1	2	2	5
35	34	R	1	2	3	1
36	35	R	1	2	3	2
37	36	R	1	2	3	3
38	37	R	1	2	3	4

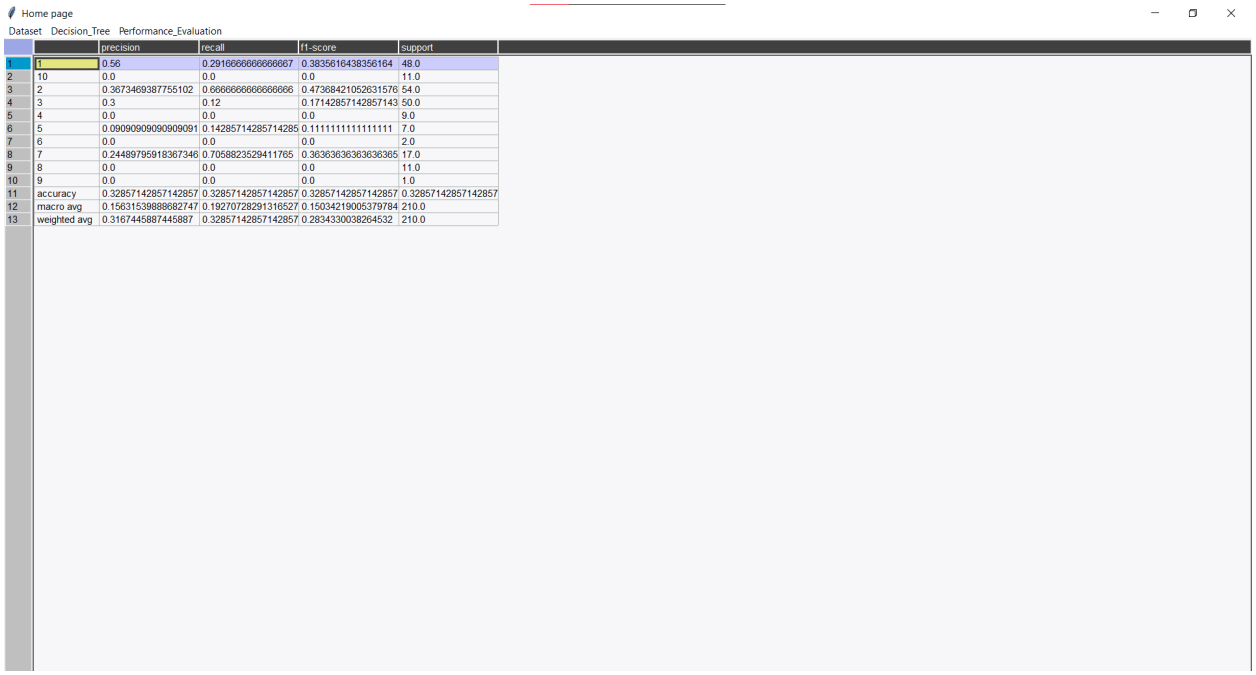


		B	1	1.1	1.2	1.3
1	0	R	1	1	1	2
2	1	R	1	1	1	3
3	2	R	1	1	1	4
4	3	R	1	1	1	5
5	4	R	1	1	2	1
6	5	R	1	1	2	2
7	6	R	1	1	2	3
8	7	R	1	1	2	4
9	8	R	1	1	2	5
10	9	R	1	1	3	1
11	10	R	1	1	3	2
12	11	R	1	1	3	3
13	12	R	1	1	3	4
14	13	R	1	1	3	5
15	14	R	1	1	4	1
16	15	R	1	1	4	2
17	16	R	1	1	4	3
18	17	R	1	1	4	4
19	18	R	1	1	4	5
20	19	R	1	1	5	1
21	20	R	1	1	5	2
22	21	R	1	1	5	3
23	22	R	1	1	5	4
24	23	R	1	1	5	5
25	24	L	1	2	1	1
26	25	B	1	2	1	2
27	26	R	1	2	1	3
28	27	R	1	2	1	4
29	28	R	1	2	1	5
30	29	B	1	2	2	1
31	30	R	1	2	2	2
32	31	R	1	2	2	3
33	32	R	1	2	2	4
34	33	R	1	2	2	5
35	34	R	1	2	3	1
36	35	R	1	2	3	2
37	36	R	1	2	3	3
38	37	R	1	2	3	4

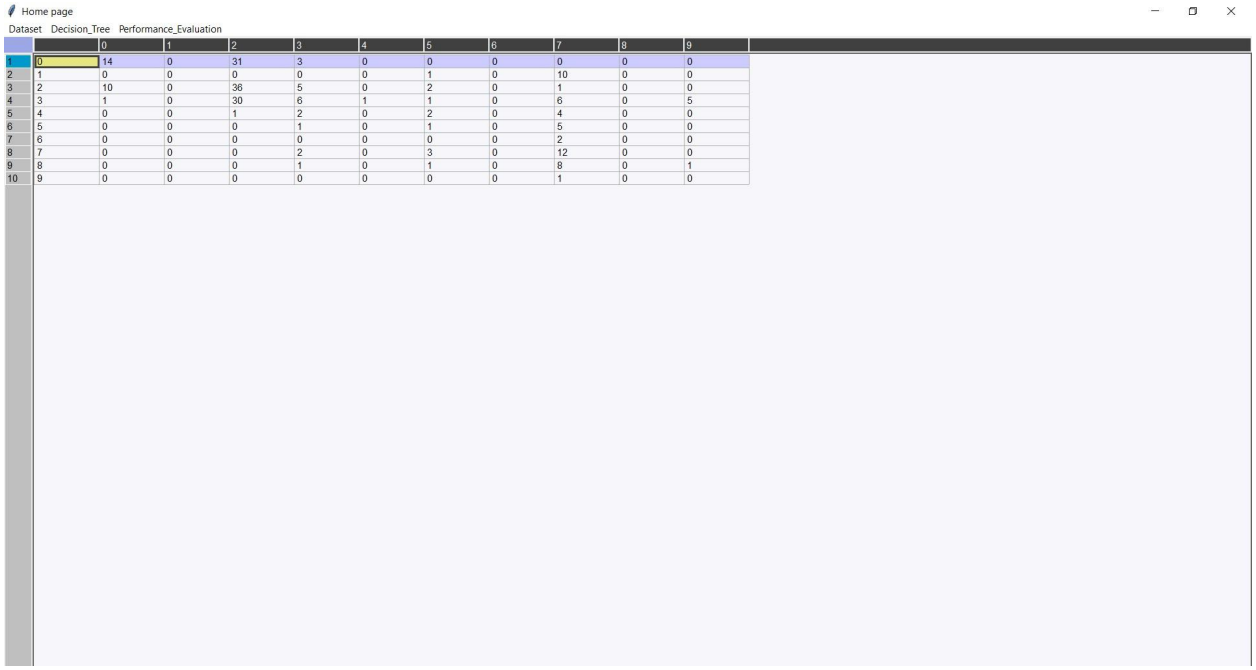

Home page
Dataset
Decision_Tree
Performance_Evaluation
Build
Display Tree
Display rules
Gini Index
Gain Ratio
Information Gain


Home page
Figure 1
Class <= 3.0
gini = 0.825
samples = 489
value = [104, 9, 112, 115, 31, 27, 8, 56, 17, 10]
class = 8
Normal Nucleoli <= 1.5
gini = 0.699
samples = 321
value = [103, 0, 109, 92, 6, 3, 1, 7, 0, 0]
class = 3
Uniformity of Cell Size <= 4.5
gini = 0.835
samples = 168
value = [1, 9, 3, 23, 25, 24, 7, 49, 17, 10]
class = 2

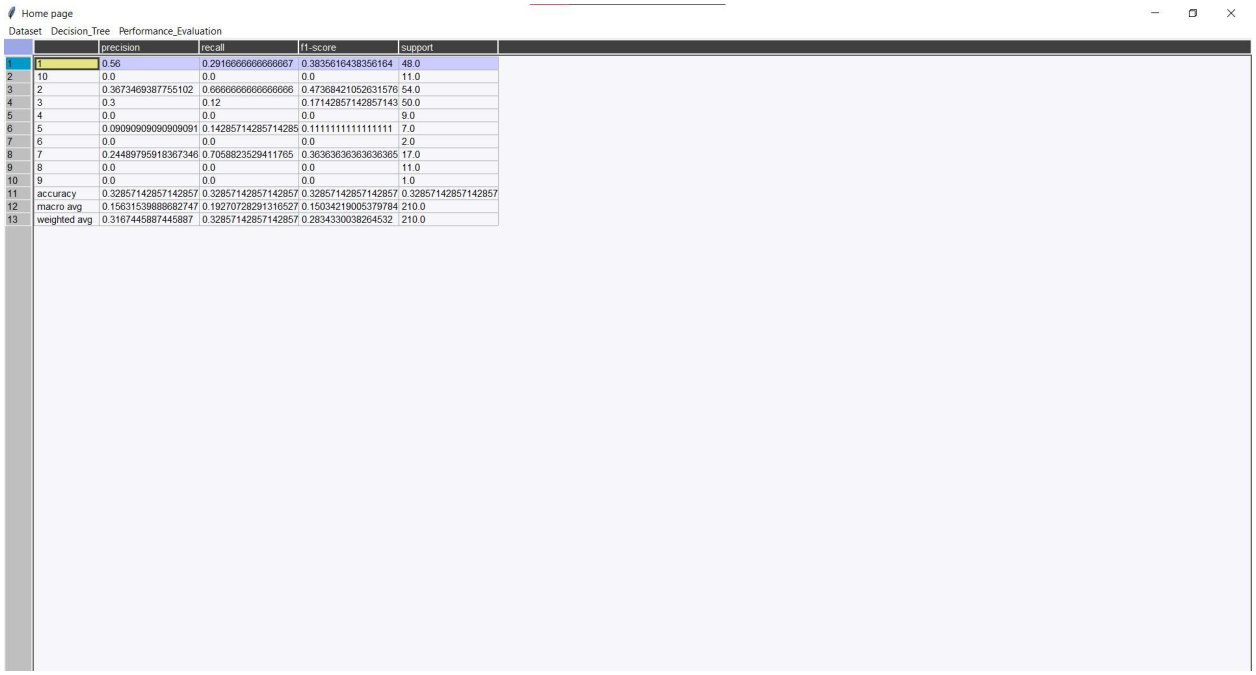
		|--- feature_9 <=
1	0	| |--- feature_7 <= 1.50
2	1	| | |--- feature_0 <=
3	2	| | | |--- feature_0
4	3	| | | | |--- class
5	4	| | | |--- feature_0
6	5	| | | | |--- class
7	6	| | |--- feature_0 >
8	7	| | | |--- feature_1
9	8	| | | | |--- class
10	9	| | | |--- feature_1
11	10	| | | | |--- class
12	11	| |--- feature_7 > 1.50
13	12	| | |--- feature_6 <=
14	13	| | | |--- feature_0
15	14	| | | | |--- class
16	15	| | | |--- feature_0
17	16	| | | | |--- class
18	17	| | |--- feature_6 >
19	18	| | | |--- feature_0
20	19	| | | | |--- class
21	20	| | | |--- feature_0
22	21	| | | | |--- class
23	22	|--- feature_9 > 3.00
24	23	| |--- feature_2 <= 4.50
25	24	| | |--- feature_6 <=
26	25	| | | |--- feature_7
27	26	| | | | |--- class
28	27	| | | |--- feature_7
29	28	| | | | |--- class
30	29	| | |--- feature_6 >
31	30	| | | |--- feature_2
32	31	| | | | |--- class
33	32	| | | |--- feature_2
34	33	| | | | |--- class
35	34	| |--- feature_2 > 4.50
36	35	| | |--- feature_3 <=
37	36	| | | |--- feature_7
38	37	| | | | |--- class
39	38	| | | |--- feature_7
40	39	| | | | |--- class
41	40	| | |--- feature_3 >
42	41	| | | |--- feature_3
43	42	| | | | |--- class
44	43	| | | |--- feature_3
45	44	| | | | |--- class



		precision	recall	f1-score	support
1	1	0.56	0.2916666666666667	0.3835616438356164	48.0
2	10	0.0	0.0	0.0	11.0
3	2	0.3673469387755102	0.6666666666666666	0.47368421052631576	54.0
4	3	0.3	0.12	0.17142857142857143	50.0
5	4	0.0	0.0	0.0	9.0
6	5	0.09090909090909091	0.14285714285714285	0.1111111111111111	7.0
7	6	0.0	0.0	0.0	2.0
8	7	0.24489795918367346	0.7058823529411765	0.36363636363636365	17.0
9	8	0.0	0.0	0.0	11.0
10	9	0.0	0.0	0.0	1.0
11	accuracy	0.32857142857142857	0.32857142857142857	0.32857142857142857	0.32857142857142857
12	macro avg	0.15631539888682747	0.19270728291316527	0.15034219005379784	210.0
13	weighted avg	0.3167445887445887	0.32857142857142857	0.2834330038264532	210.0



		0	1	2	3	4	5	6	7	8	9
1	0	14	0	31	3	0	0	0	0	0	0
2	1	0	0	0	0	0	1	0	10	0	0
3	2	10	0	36	5	0	2	0	1	0	0
4	3	1	0	30	6	1	1	0	6	0	5
5	4	0	0	1	2	0	2	0	4	0	0
6	5	0	0	0	1	0	1	0	5	0	0
7	6	0	0	0	0	0	0	0	2	0	0
8	7	0	0	0	2	0	3	0	12	0	0
9	8	0	0	0	1	0	1	0	8	0	1
10	9	0	0	0	0	0	0	0	1	0	0



		precision	recall	f1-score	support
1	1	0.56	0.2916666666666667	0.3835616438356164	48.0
2	10	0.0	0.0	0.0	11.0
3	2	0.3673469387755102	0.6666666666666666	0.47368421052631576	54.0
4	3	0.3	0.12	0.17142857142857143	50.0
5	4	0.0	0.0	0.0	9.0
6	5	0.09090909090909091	0.14285714285714285	0.1111111111111111	7.0
7	6	0.0	0.0	0.0	2.0
8	7	0.24489795918367346	0.7058823529411765	0.3636363636363636	17.0
9	8	0.0	0.0	0.0	11.0
10	9	0.0	0.0	0.0	1.0
11	accuracy	0.32857142857142857	0.32857142857142857	0.32857142857142857	0.32857142857142857
12	macro avg	0.15631539888682747	0.19270728291316527	0.15034219005379784	210.0
13	weighted avg	0.3167445887445887	0.32857142857142857	0.2834330038264532	210.0

References:

- http://webpages.iust.ac.ir/yaghini/Courses/Data_Mining_882/DM_05_04_Rule-Based%20Classification.pdf
- https://victorzhou.com/blog/information-gain/
- https://www.analyticssteps.com/blogs/what-gini-index-and-information-gain-decision-trees
- https://towardsdatascience.com/the-complete-tensorflow-tutorial-for-newbies-dc

3acc1310f8
- https://towardsdatascience.com/beginners-guide-to-deep-learning-with-tensorflow-ca85969b2f2