Assignment 5

Name	ID
Somaia Ahmed Abdelrahman	9220374

Eng. Samar Alaa

Dr. Inas Yassine

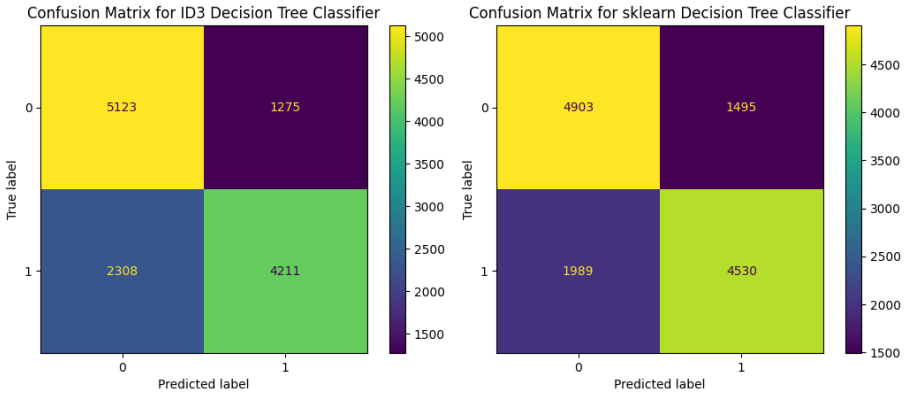
Faculty of Engineering - Cairo University

Comparison between from scratch model and Scikit-learn model

In this comparison, we evaluate the performance of a from-scratch Decision Tree model and the Scikit-learn Decision Tree model on the same dataset.

- **From scratch model**
  - Train Accuracy for from scratch Decision Tree model: 0.7151705183292688
  - Test Accuracy for from scratch Decision Tree model: 0.7226136099713556
  - Train time for from scratch Decision Tree model: 1.7135 sec
  - Prediction time for from scratch Decision Tree model: 0.1568 sec
- **Scikit-learn model**
  - Train accuracy for built-in Decision Tree model: 0.7264351798087717
  - Test accuracy for built-in Decision Tree model: 0.7302779283115275
  - Train time for built-in Decision Tree model: 0.0517 sec
  - Prediction time for built-in Decision Tree model: 0.0066 sec

Metric	From scratch	Scikit-learn
Train accuracy	71.52%	72.64%
Test accuracy	72.26%	73.03%
Train time	1.7135sec	0.0517sec
Prediction time	0.1568sec	0.0066sec



Conclusion:

- The **Scikit-learn Decision Tree model** outperforms the **from-scratch model** in terms of both accuracy and efficiency (both in training and prediction time).
- While the from-scratch model performs reasonably well, Scikit-learn's built-in implementations offer faster execution and slightly better accuracy.
- The **from-scratch model** can be improved further in terms of both optimization and potentially tuning parameters to enhance performance.

Part 3

Part 3

1- $P(A=1) = \frac{8}{14}$, $P(A=0) = \frac{6}{14}$

$H(A) = -\frac{8}{14}\log_2\frac{8}{14} - \frac{6}{14}\log_2\frac{6}{14} = 0.985$

2- IG for each attribute..

→ early registration: 0 [4+ 4-] H(0) = 1

1 [4+ 2-] H(1) = 0.9

$IG = 0.985 - \left(\frac{8}{14} \times 1 + \frac{6}{14} * 0.918\right) = 0.02$

→ Finished homework: 0 [3+ 4-] = 0.985

1 [5+ 2-] = 0.863

$IG = 0.985 - \left(\frac{7}{14} * 0.985 + \frac{7}{14} \times 0.863\right)$

$= 0.061$

→ Senior: 0 [3+ 3-] = H = 1

1 [5+ 3-] = H 0.954

IG = 0.011

→ likes Coffee: 0 [5+ 5-) → H=1

1 [3+ 1-] = 0.811

IG = 0.039

→ Liked the last HW : 0 [3+ 2-] 0.971
1 [5+ 4-] 0.991
IG1 = 0.001

∴ root Node : Finished HW II

Finished HW II = 1

$$S_{ER1} = \frac{-3}{3}\log_2\frac{3}{3} - 0 = 0$$

$$S_{ER0} = \frac{-2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$$

$$S_{S1} = \frac{-3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$$
$$S_{S0} = \frac{-2}{2}\log_2 1 - 0 = 0$$

$$S_{LC1} = \frac{-1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$
$$S_{LC0} = \frac{-4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5} = 0.722$$

$$S_{LH1} = \frac{-3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$$
$$S_{LH0} = \frac{2}{2}\log 2 = \frac{2}{2} - 0 = 0$$

Parent Entropy $= \frac{-5}{7}\log_2\frac{5}{7} - \frac{2}{7}\log_2\frac{2}{7}$

$= \boxed{0.863}$

$IG_{ER} = 0.863 - \frac{3}{7}(0) - \frac{4}{7} * 1 = 0.292$

$IG_S = 0.863 - \frac{5}{7} * 0.971 - \frac{2}{7} * 0 = 0.169$

$IG_{LC} = 0.863 - \frac{2}{7} * 1 - \frac{5}{7} * 0.722 = 0.062$

$IG_{LH} = 0.863 - \frac{5}{7} * 0.971 - 0 = 0.169$

Finished HW II: 0

Parent entropy $= -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7}$

$= 0.985$

$S_{S1} = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918$

$S_{S0} = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.811$

$IG = 0.985 - \frac{3}{7} * 0.918 - \frac{4}{7} * 0.811$

$= 0.128$

$S_{LC1} = 0$

$S_{LC0} = -\frac{1}{5}\log_2\frac{1}{5} - \frac{4}{5}\log_2\frac{4}{5} = 0.722$

$IG = 0.469$

$S_{LH1} = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$

$S_{LH0} = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$

$IG = 0.02$

Early Reg = 1

Parent Entropy = 0

No more Splitting

Early reg = 0

rows = 4 < $\frac{\text{minSamples}}{\text{leaf}}$

likes Coffee = 1

Parent Entropy = 0 No more Splitting

likes Coffee : 0

parent Entropy $= -\frac{4}{5}\log_2\frac{4}{5} - \frac{1}{5}\log_2\frac{1}{5}$

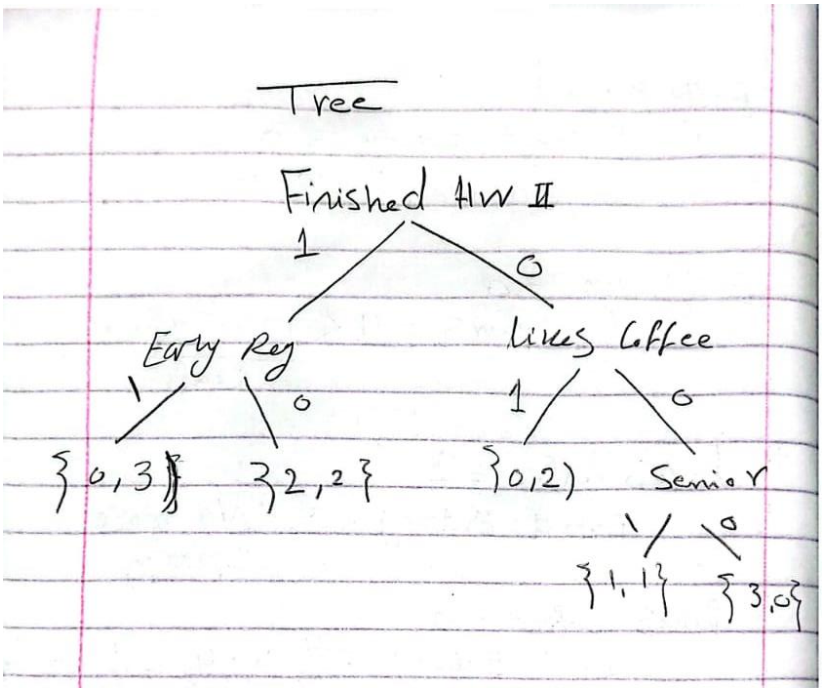
$= 0.722$

$S_{S_1} = 1$

$S_{S_0} = 0$ $\quad IG = 0.322$

$S_{LH_1} = 0.918$ $\quad IG = 0.1712$

$S_{LH_0} = 0$


Tree
Finished HW II
1
0
Early Reg
Likes Coffee
1
0
1
0
{0,3}
{2,2}
{0,2}
Senior
1
0
{1,1}
{3,0}