Lab-7

07/03/24

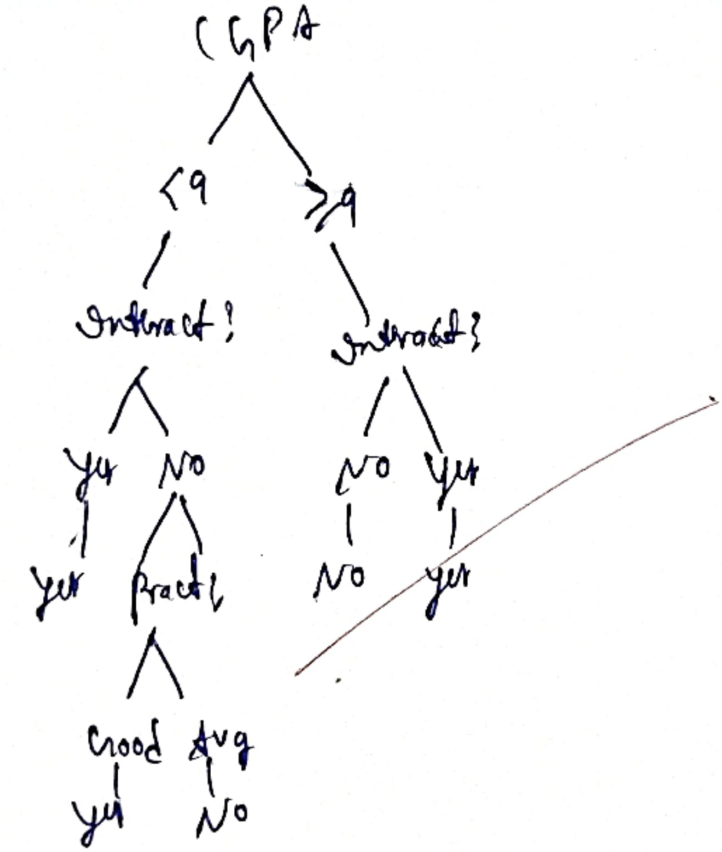
Implement Random Forest ensemble method on a given dataset.

To write)

Sample Sets) Sample Set-1 - CGPA at Root node

S.No	CGPA	Interactiveness	communication skills	practical knowledge	Job offer
1	<9	Yes	Good		
2	<9	No	Moderate	Good	Yes
3	<9	No	Moderate	Good	Yes
4	≥9	No	Moderate	Average	No
5	≥9	Yes	Moderate	Average	No
				Good	Yes

Decision tree (Root: CGPA):

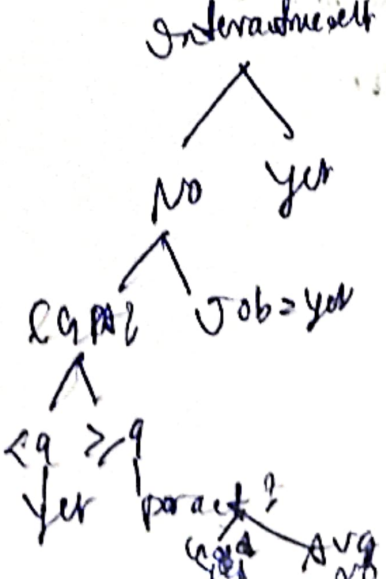


- If CGPA < 9:
  - if interactiveness = yes → Joboffer = yes
  - if interactiveness = No:
    - if practical knowledge = Good → Joboffer = yes
    - else (Average) → Joboffer = No.
- For CGPA ≥ 9:
  - if interactiveness = No → Joboffer = No
  - if interactiveness = yes → Joboffer = yes

Sample 2: interactiveness at Root Node

Sl.No	CGPA	Interactiveness	Communication skills	practical knowledge	Joboffer
2	<9	No	moderate	Good	yes
3	≥9	No	moderate	Average	No
4	≥9	No	moderate	Average	No
5	≥9	yes	moderate	Good	yes
6	≥9	yes	moderate	Good	yes

Decision tree (Root = interactiveness):



* If Interactiveness = yes → Job offer = yes
* If Interactiveness = No:
    * If (CGPA < 9 → Job offer = Yet
    * If (CGPA ≥ 9:
        * If practical knowledge = Good → Job offer = yes
        * Else (Average) → Job offer = No.

To write:

After building the RF model, write the answer for the following questions in your observation book.

1. For "edit.csv" dataset.

What is the best accuracy score and confusion matrix of the classifier you observed and using how many trees?

* Best accuracy Score: 1.0 (or 100%).
* Number of Trees (n_estimators): 1.
* Confusion matrix:

```
[[0  0  0]
 [0  9  0]
 [0  0  1]]
```

Best result was obtained by fine-tuning the number of trees from 1 to 100. The best accuracy was achieved with just 1 tree, which indicates that the dataset is simple and easily separable.