

NAME :- KRUNAL RANK

ROLL No :- U18C0081

CLASS :- BTECH 4TH YEAR

SEMESTER :- 7

DIVISION :- B

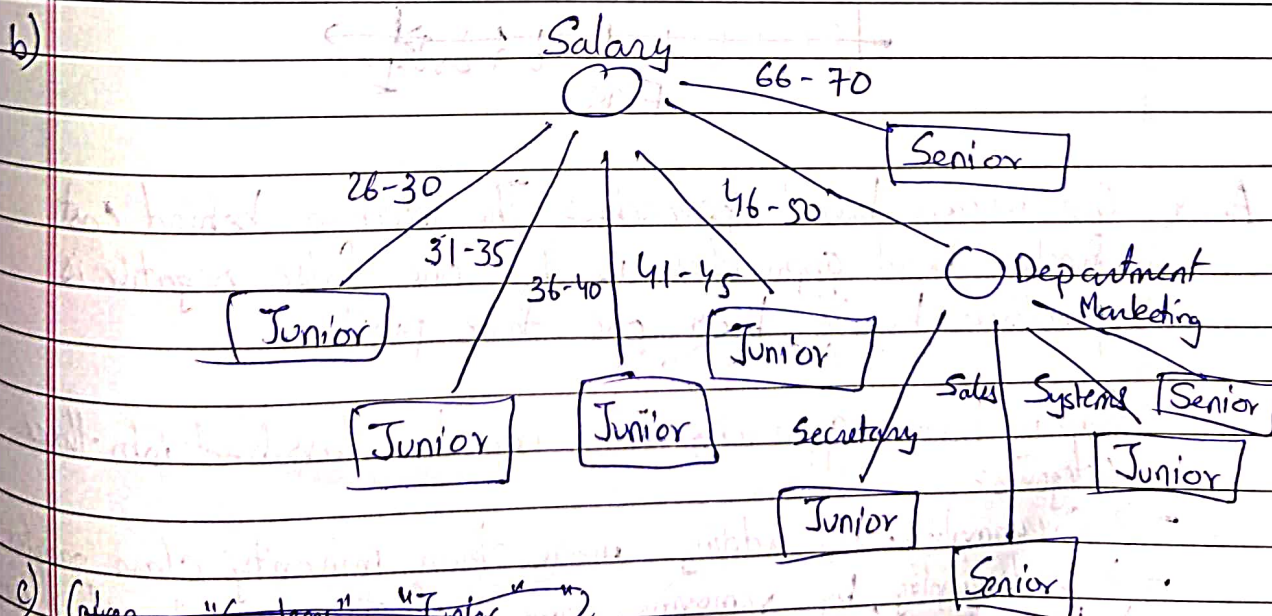
Data Warehousing and Data Mining Tutorial 5

Ans 1. If pruning a subtree we would remove the subtree completely with method (b). However, the method (a), if pruning a rule, we may remove the pre-condition of it. The latter is less restrictive.

Ans 2.

a) The basic tree algorithm must:-

- consider the count in calculating information gain.
- count must also be utilised in detecting common classes.



c) Given "Systems", "Junior", "2"

$$P(X|\text{senior}) = 20$$

$$P(X|\text{junior}) = 0.018$$

Hence, Junior is the prediction.

~~Every feasible solution is correct.~~

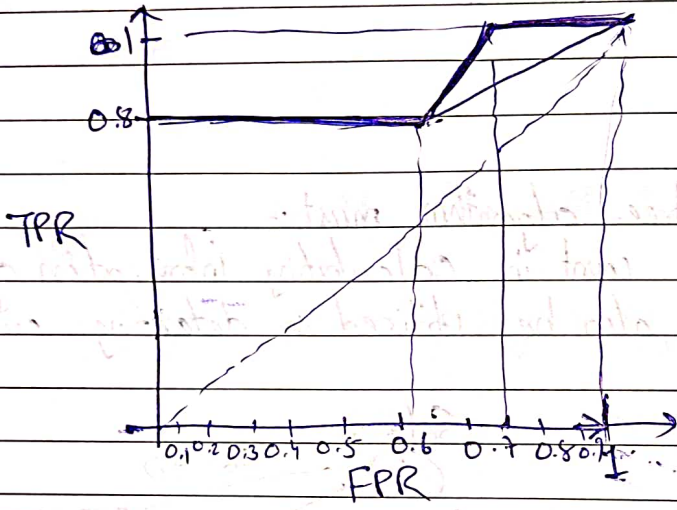
Ans 3:

$$TPR = \frac{TP}{P} = 0.602$$

$$TNR = 0.602$$

$$FPR = 0.332$$

$$ENR = 0.398$$



Ans 4: Cost Function based approaches:- The intuition behind cost function based approaches is that one false negative is penalised heavily than one false positive.

Sampling based approaches:- This can be classified into three categories:-

- Oversampling, by adding more than minority class samples
- Undersampling, by removing some of the majority class samples
- Hybrid