NAME :- KRUNAL RANK

ROLL NO :- U18CO081

CLASS :- BTECH 4TH YEAR

SEMESTER :- 7

DIVISION :- B

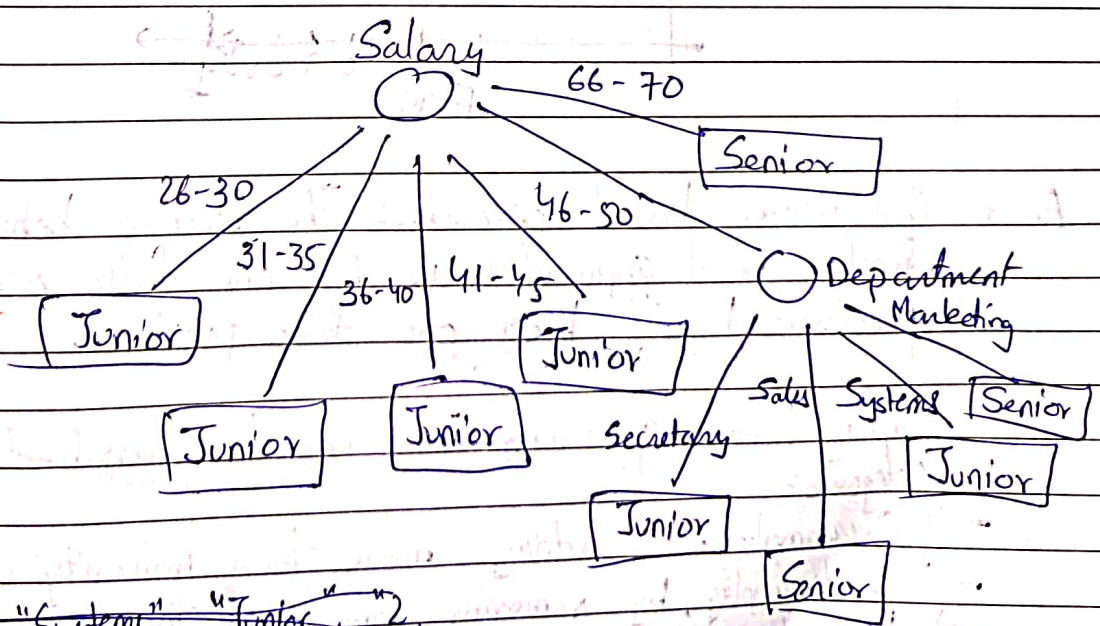
Data Warehousing and Data Mining
Tutorial 5

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 precondition 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.

b)



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

P(X | senior) = 0

P(X | junior) = 0.028.

Hence, Junior is the prediction.

~~d) Every feasible solution is correct.~~

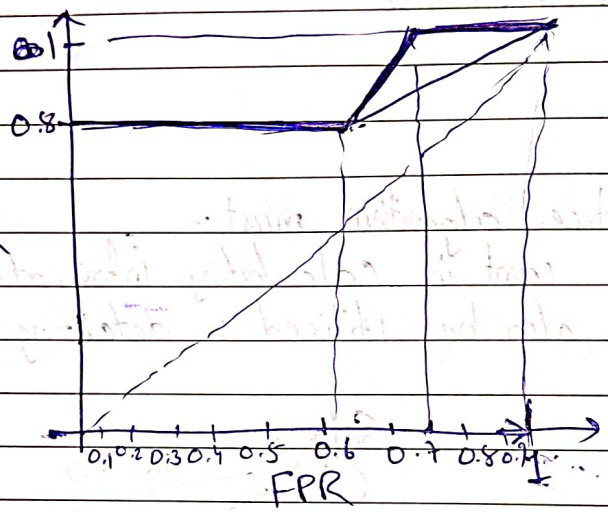
Ans 3:-

$$TPR = \frac{TP}{P} = 0.602 \quad 0.668$$

$TNR = 0.602$

$FPR = 0.332$

$FNR = 0.398$



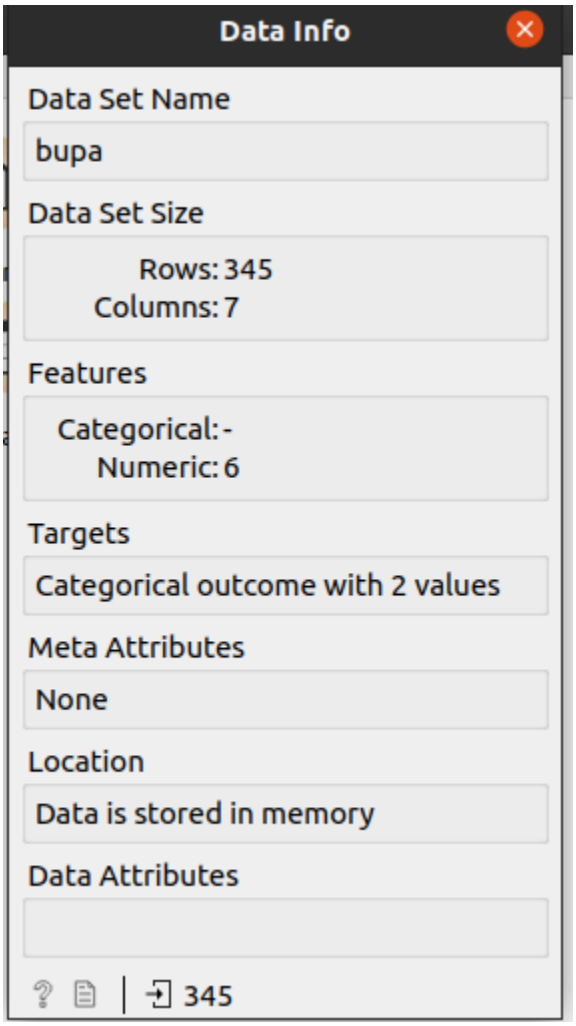
Ans 4:- Cost Function based approaches:- The intution behind cost function based approaches is that one false negative is penalised heavily than one false positive.

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

5.

Dataset Info:



The BUPA dataset contains 345 single male patients with 6 numeric attributes. Five of these attributes are blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption. Each line in the dataset constitutes the record of a single male individual.

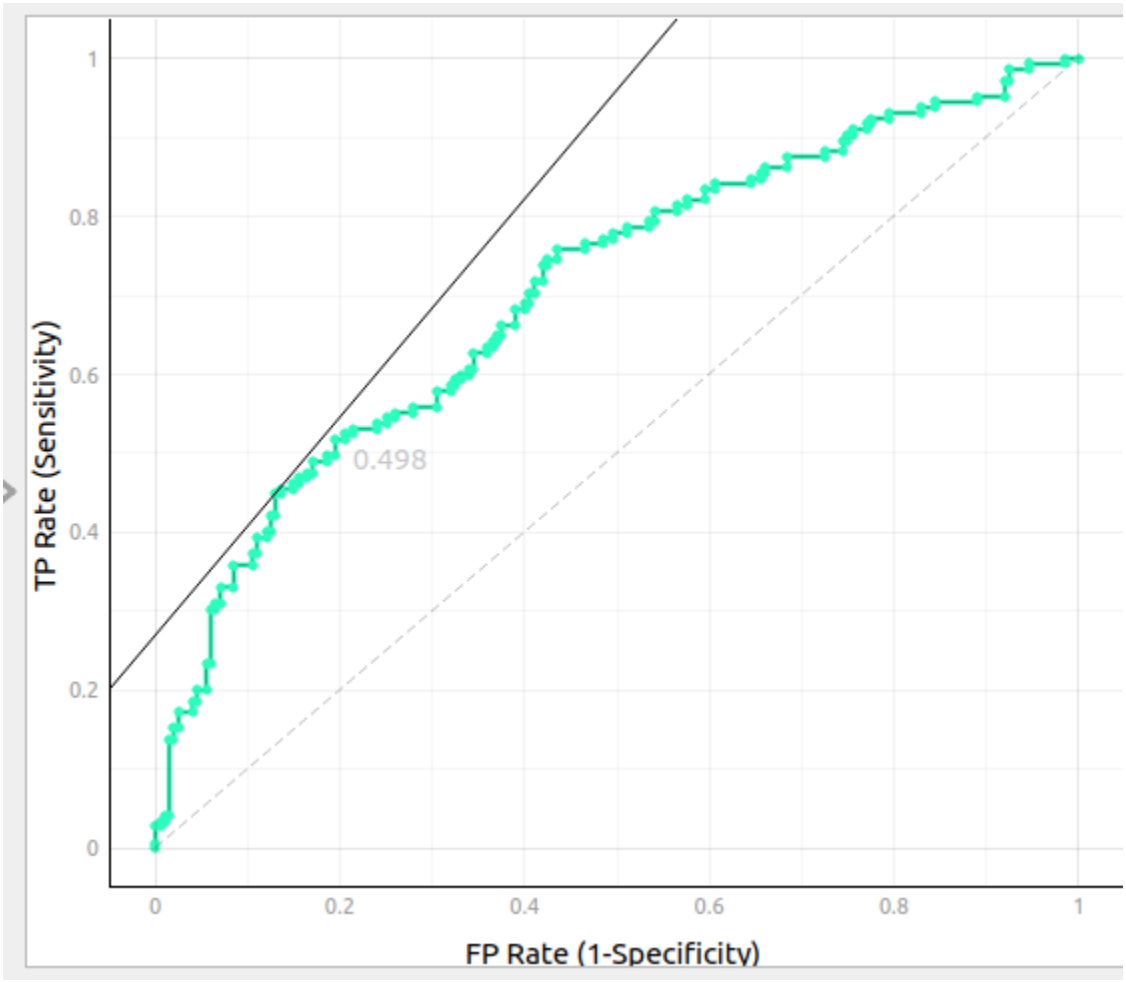
Target Variable : Liver Disorder Prediction (Y/N) [Binary]

Logistic Regression:

Evaluation Results

Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.706	0.678	0.671	0.674	0.678

Actual \ Predicted	1	2	Σ
1	75	70	145
2	41	159	200
Σ	116	229	345



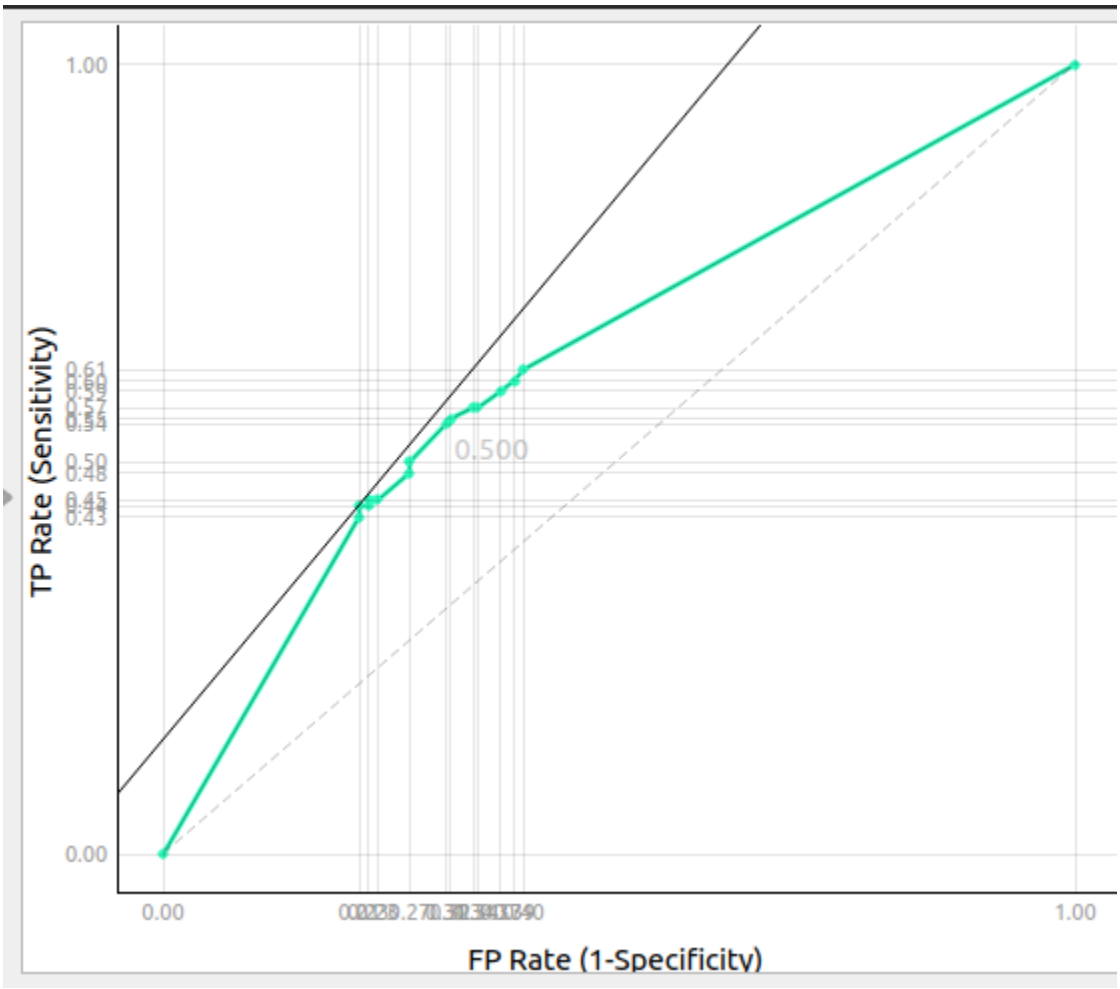
Decision Tree:

Evaluation Results

Model	AUC	CA	F1	Precision	Recall
Random Forest	0.629	0.629	0.628	0.628	0.629

Random Forest with 1 tree = Decision Tree

Actual \ Predicted	1	2	Σ
1	79	66	145
2	62	138	200
Σ	141	204	345

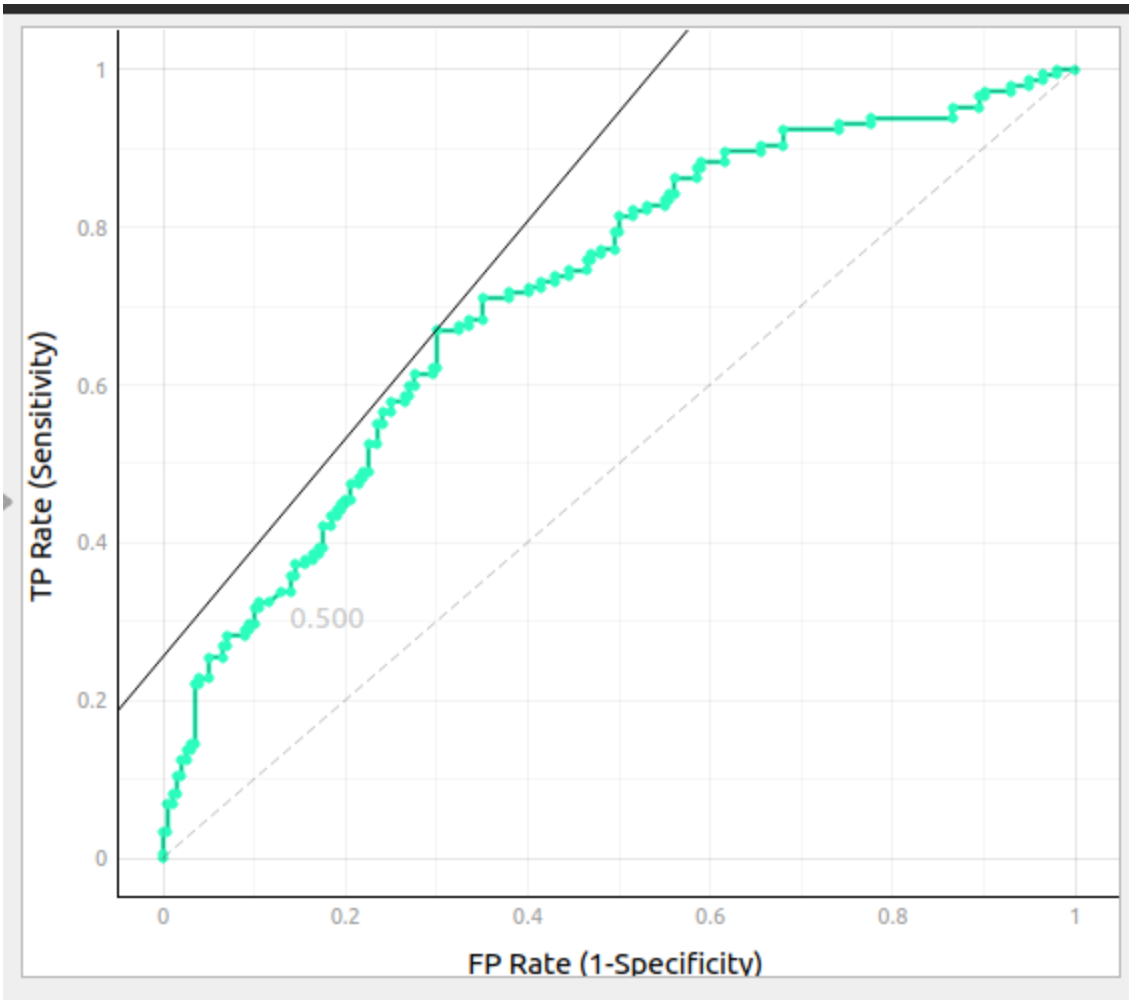


Support Vector Machine :

Model	AUC	CA	F1	Precision	Recall
SVM	0.718	0.649	0.623	0.649	0.649

Predicted

Actual	1	2	Σ
1	52	93	145
2	28	172	200
Σ	80	265	345



Naive Bayes:

Evaluation Results

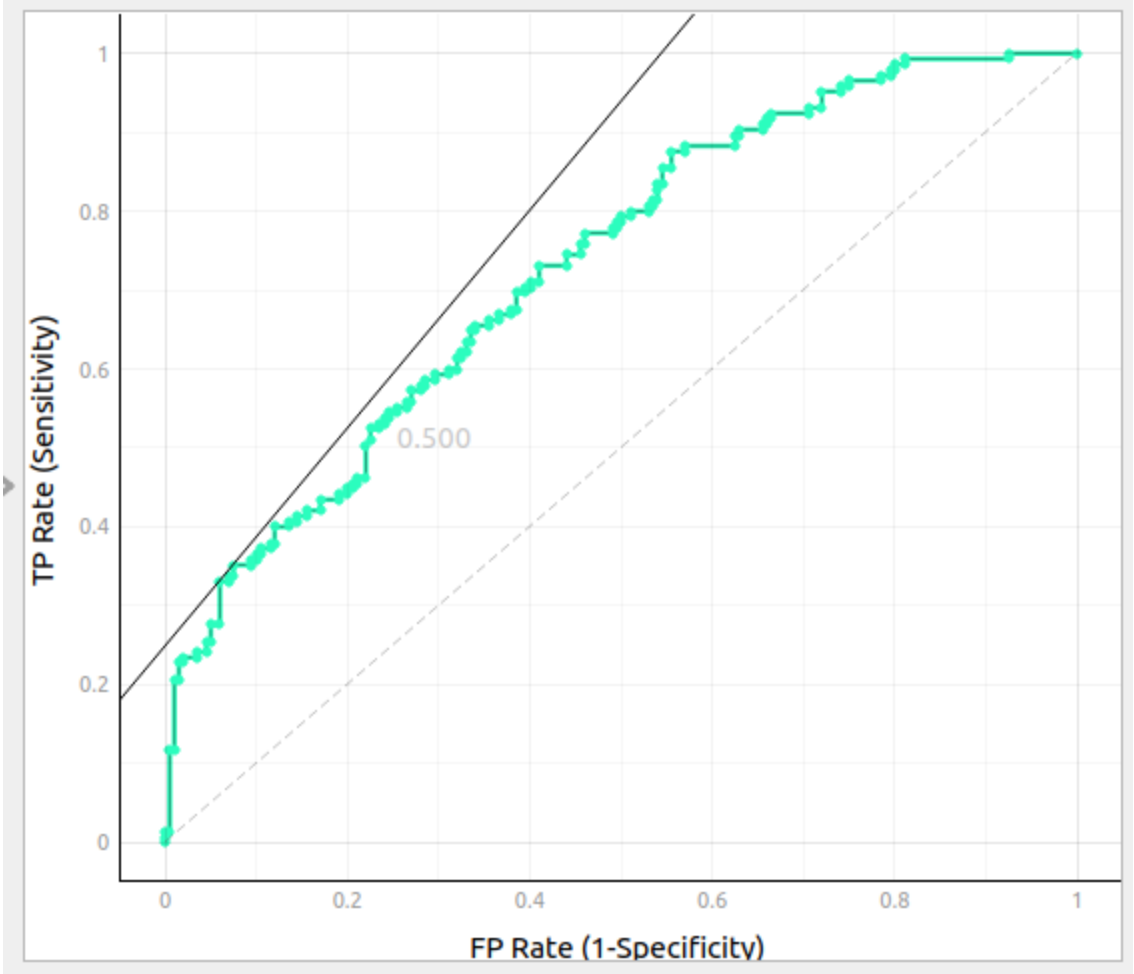
Model	AUC	CA	F1	Precision	Recall
Naive Bayes	0.731	0.667	0.663	0.663	0.667

Model Comparison by AUC

Show: Number of instances

Actual \ Predicted	1	2	Σ
1	79	66	145
2	49	151	200
Σ	128	217	345

Select Correct | Select Misclassified | Clear Selection


TP Rate (Sensitivity)
FP Rate (1-Specificity)
0.500
0
0.2
0.4
0.6
0.8
1