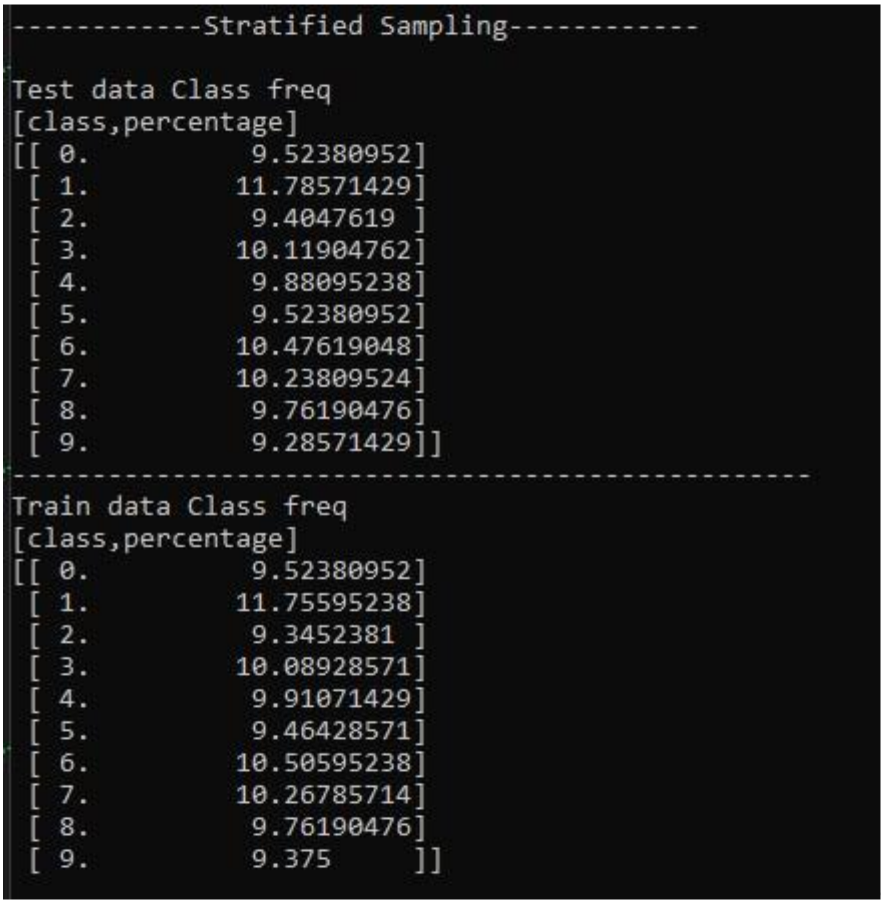
Report

1a) PCA- Principal Component Analysis, is used to reduce the dimensions of the data. Initially the continuous initial variables are standardized, so each one of them has equal contribution. Then each variable’s relationship with other variables is seen. Then principal components are chosen using the covariance matrix.

1b) SVD- Singular Value Decomposition, it is factorizing the input feature matrix into 3 matrices, where two of them are orthonormal and the centre matrix is diagonal with positive entries.

1c) t-SNE, uses local relationships between points to create low-dimension mapping. It creates a gaussian distribution which helps to define relationships.

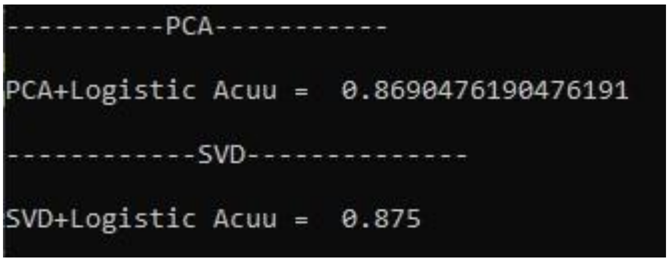
1d)



```
------------Stratified Sampling------------

Test data Class freq
[class,percentage]
[[ 0.          9.52380952]
 [ 1.         11.78571429]
 [ 2.          9.4047619 ]
 [ 3.         10.11904762]
 [ 4.          9.88095238]
 [ 5.          9.52380952]
 [ 6.         10.47619048]
 [ 7.         10.23809524]
 [ 8.          9.76190476]
 [ 9.          9.28571429]]
---------------------------------------------------
Train data Class freq
[class,percentage]
[[ 0.          9.52380952]
 [ 1.         11.75595238]
 [ 2.          9.3452381 ]
 [ 3.         10.08928571]
 [ 4.          9.91071429]
 [ 5.          9.46428571]
 [ 6.         10.50595238]
 [ 7.         10.26785714]
 [ 8.          9.76190476]
 [ 9.          9.375     ]]
```

4

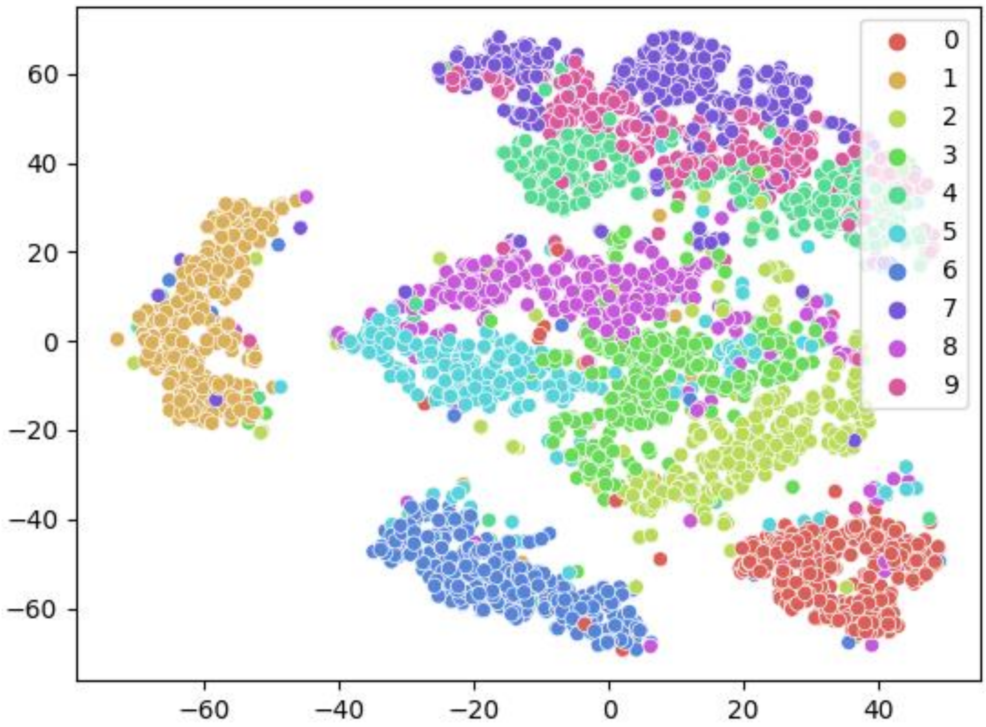
1e,f)



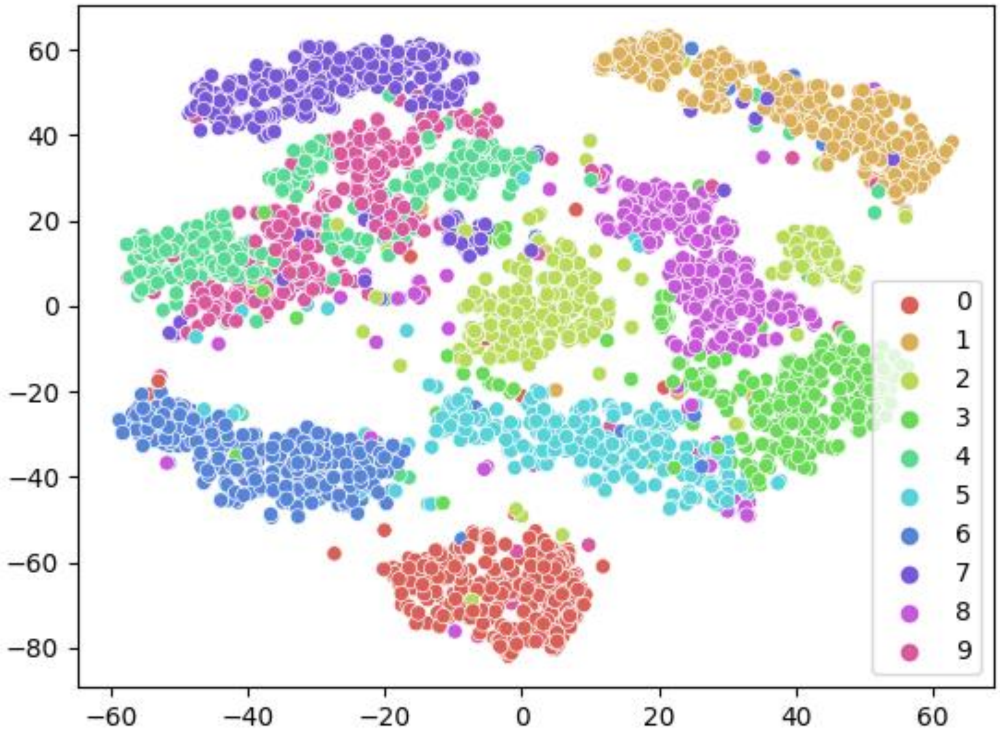
```
----------PCA-----------

PCA+Logistic Acuu =  0.8690476190476191

------------SVD--------------

SVD+Logistic Acuu =  0.875
```

SAKSHAM DHULL 2018186

TSNE after PCA



TSNE after SVD



1g) Accuracies obtained by both the methods are comparable. This is because PCA uses SVD solver to reduce the dimensions.

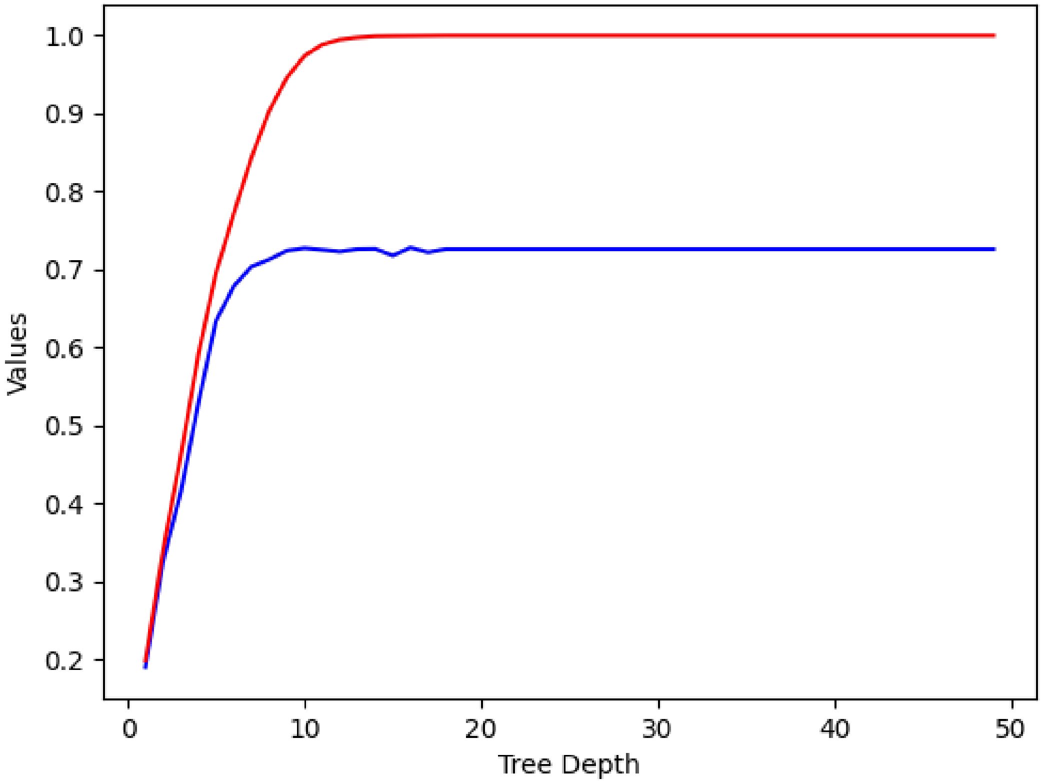
SAKSHAM DHULL 2018186

3)

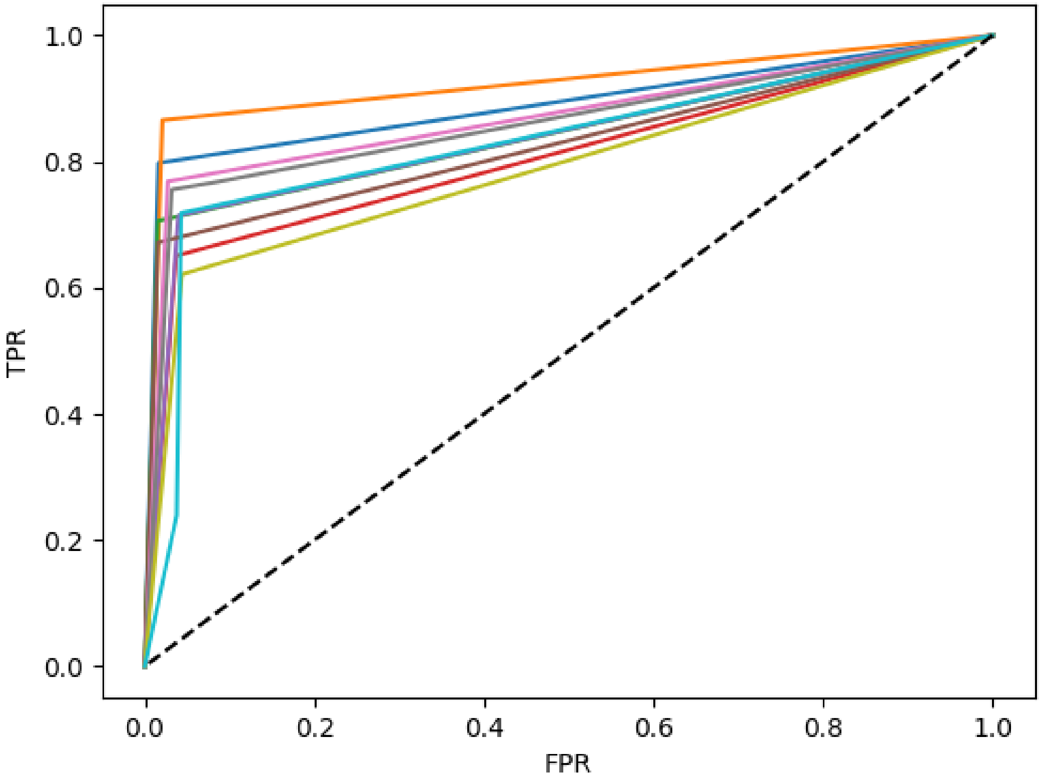
Dataset-A

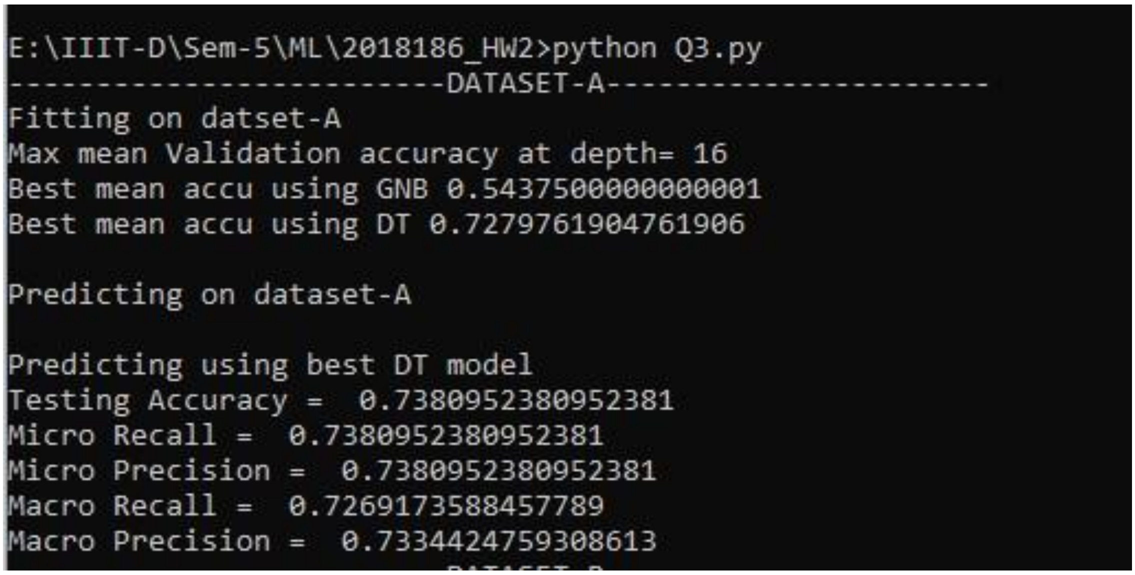
Red→Training Accuracy

Blue→Validation Accuracy



ROC-A



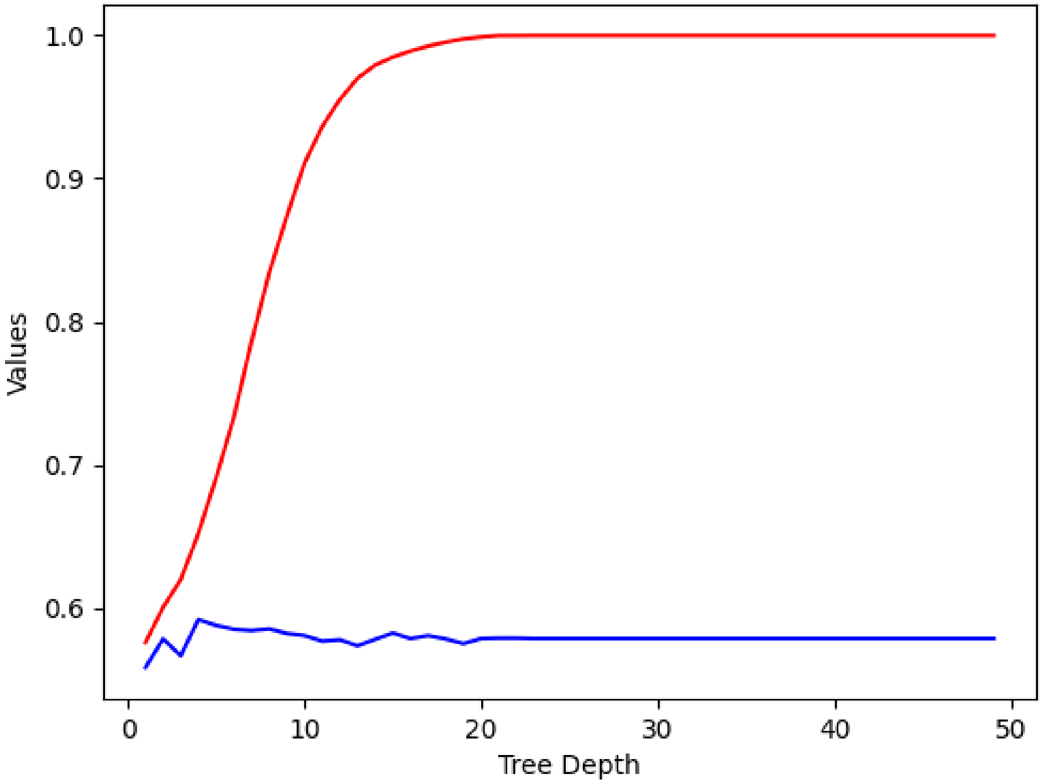


```
E:\IIIT-D\Sem-5\ML\2018186_HW2>python Q3.py
-------------------------DATASET-A----------------------
Fitting on datset-A
Max mean Validation accuracy at depth= 16
Best mean accu using GNB 0.5437500000000001
Best mean accu using DT 0.7279761904761906

Predicting on dataset-A

Predicting using best DT model
Testing Accuracy =  0.7380952380952381
Micro Recall =  0.7380952380952381
Micro Precision =  0.7380952380952381
Macro Recall =  0.7269173588457789
Macro Precision =  0.7334424759308613
```

Dataset-B

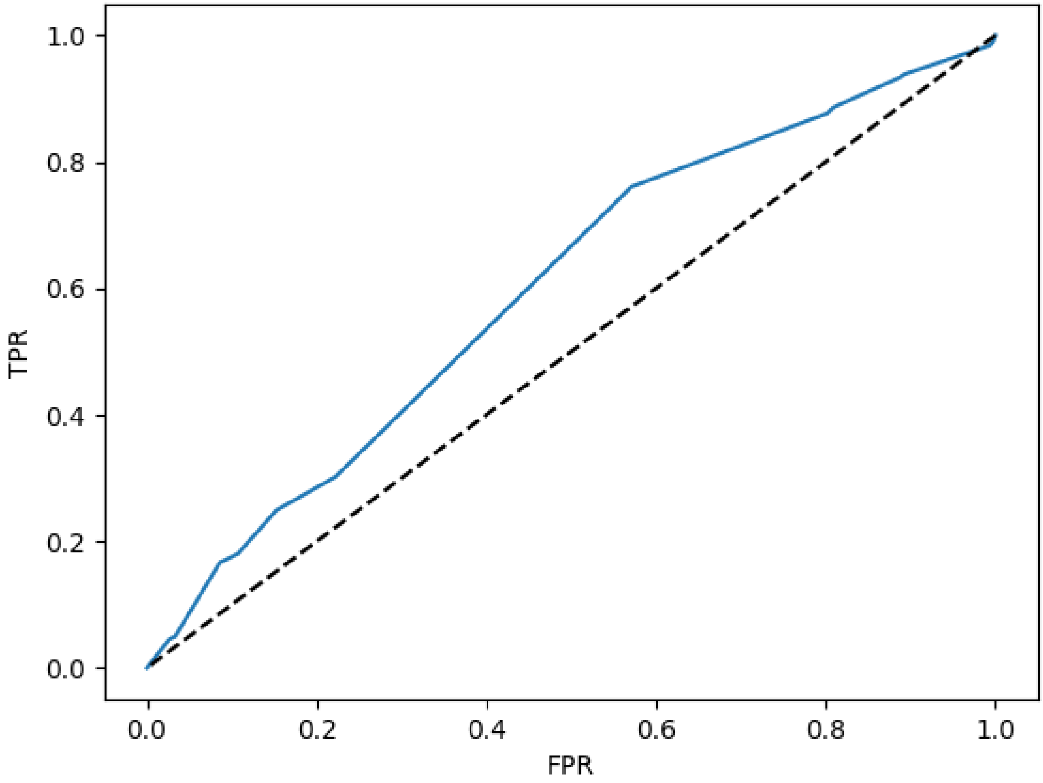
Red→Training Accuracy

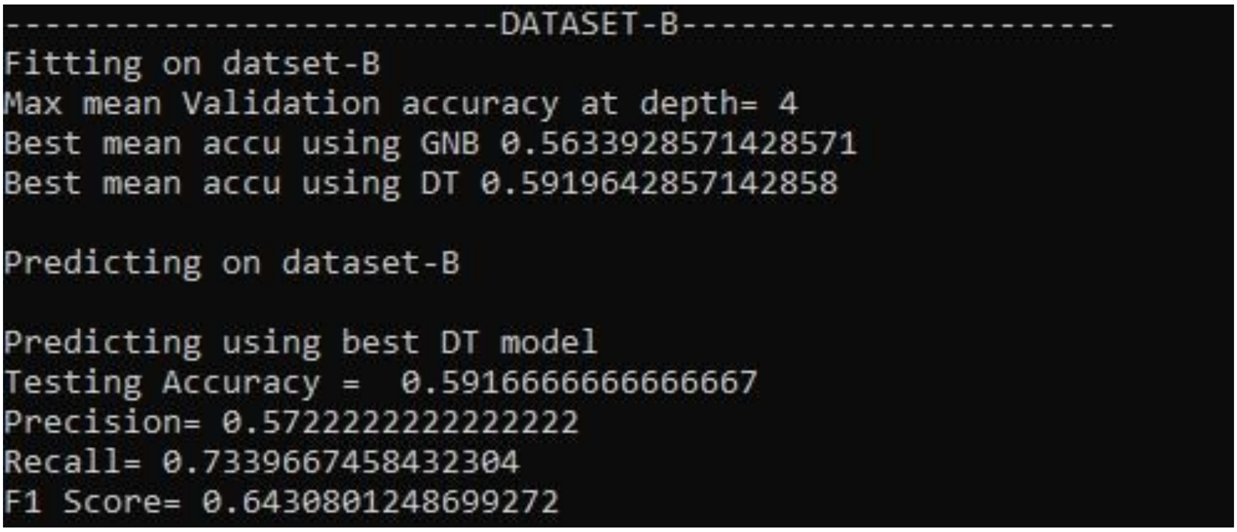
Blue→Validation Accuracy



SAKSHAM DHULL 2018186

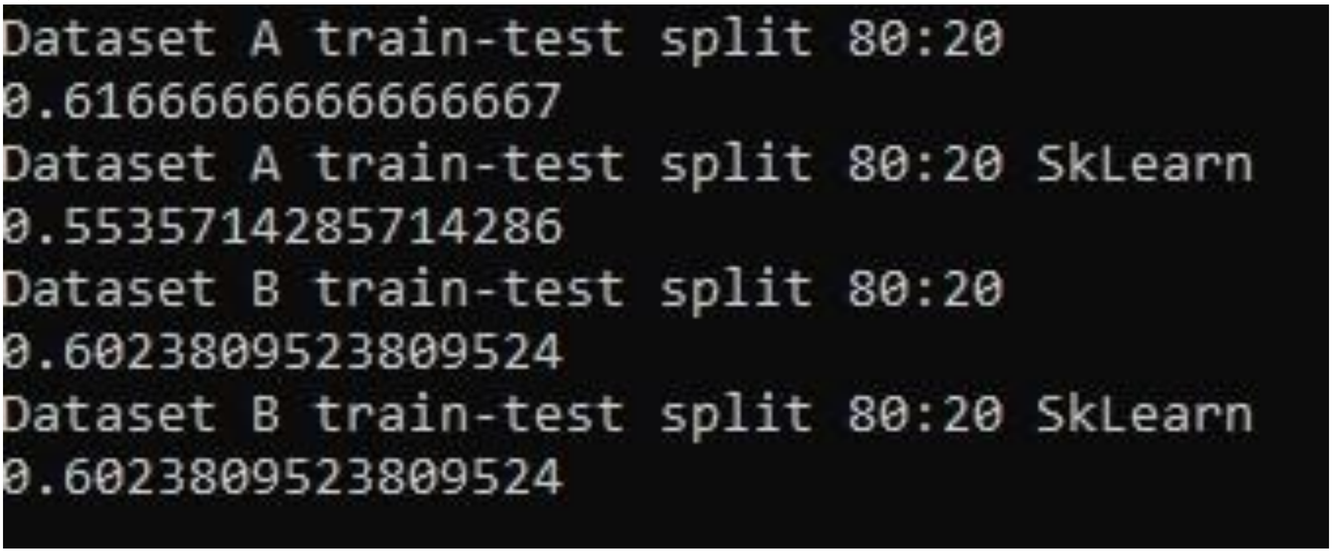
ROC-B





```
-------------------------DATASET-B----------------------
Fitting on datset-B
Max mean Validation accuracy at depth= 4
Best mean accu using GNB 0.5633928571428571
Best mean accu using DT 0.5919642857142858

Predicting on dataset-B

Predicting using best DT model
Testing Accuracy =  0.5916666666666667
Precision= 0.5722222222222222
Recall= 0.7339667458432304
F1 Score= 0.6430801248699272
```

4)



```
Dataset A train-test split 80:20
0.6166666666666667
Dataset A train-test split 80:20 SkLearn
0.5535714285714286
Dataset B train-test split 80:20
0.6023809523809524
Dataset B train-test split 80:20 SkLearn
0.6023809523809524
```

(5) a) initial entropy $= -\left[\frac{5}{14}\log_2\frac{5}{14} + \frac{9}{14}\log_2\frac{9}{14}\right] = 0.9403 = E(S)$

Info Gain $=$ ~~b~~ $E(S) - \sum_{v\in A} \frac{|S_v|}{|S|} E(S_v)$

Info Gain $(S, \text{Climate}) = 0.9403 - \frac{4}{13} - \frac{6}{14}\left[-\left(\frac{2}{6}\log_2\frac{2}{6} + \frac{4}{6}\log_2\frac{4}{6}\right)\right] - \frac{4}{14}\left[-\left(\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right.\right.$

$= 0.0292$

Info Gain $(S, \text{Humidity}) = 0.9403 - \frac{7}{14}\left[-\left(\frac{4}{7}\log_2\frac{4}{7} + \frac{3}{7}\log_2\frac{3}{7}\right)\right] - \frac{7}{14}\left[-\left(\frac{6}{7}\log_2\frac{6}{7} + \frac{1}{7}\log_2\frac{1}{7}\right)\right]$

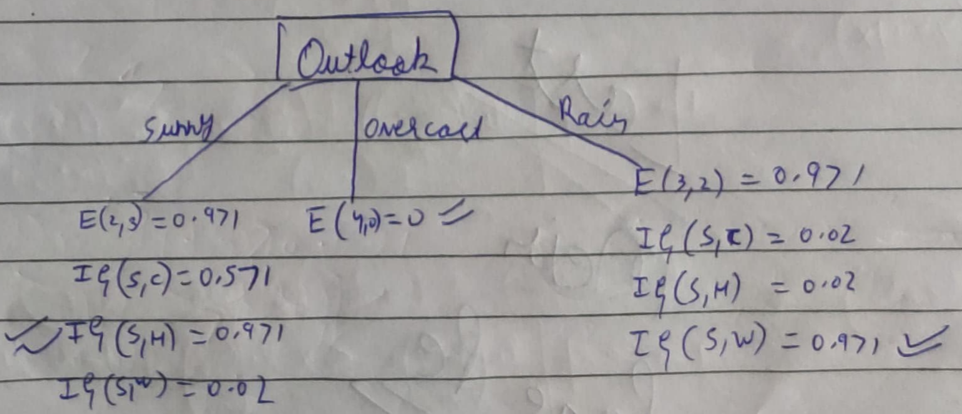
$=$ ~~0.0~~ 0.1518

Info Gain $(S, \text{Wind}) = 0.9403 - \frac{8}{14}\left[-\left(\frac{2}{8}\log_2\frac{2}{8} + \frac{6}{8}\log_2\frac{6}{8}\right)\right] - \frac{6}{14} = 0.0481$

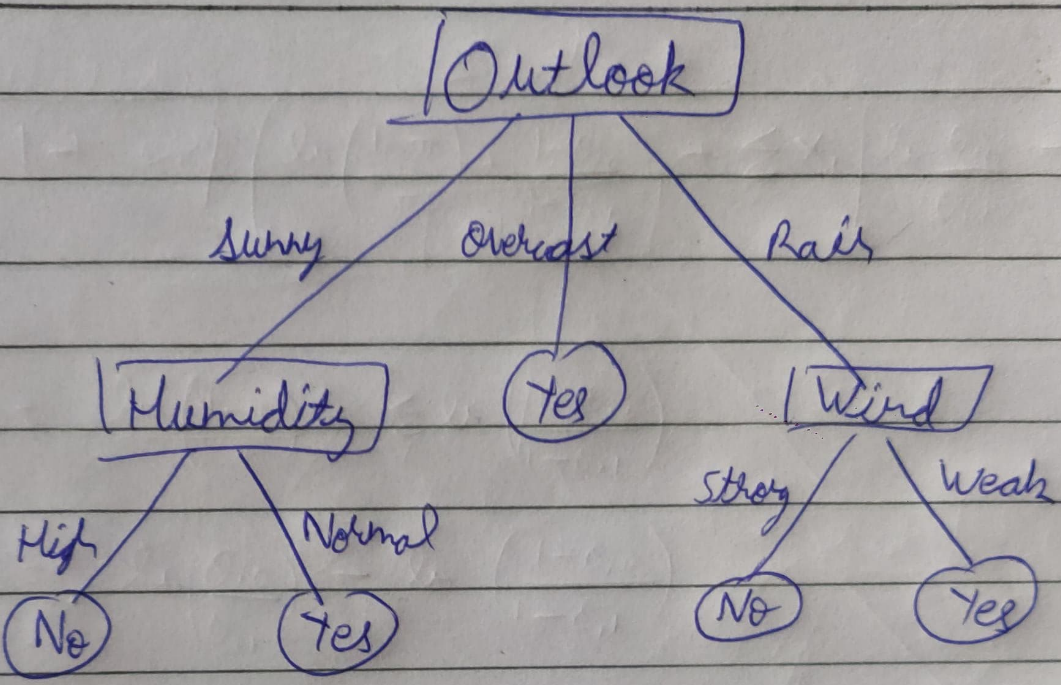
Info Gain $(S, \text{Outlook}) = 0.9403 - \frac{5}{14}\left[-\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right)\right] - \frac{5}{14}\left[-\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right) - \frac{4}{14}\right)$

✓ $= 0.2468$

1st ~~a~~ split by Outlook.



P.T.O



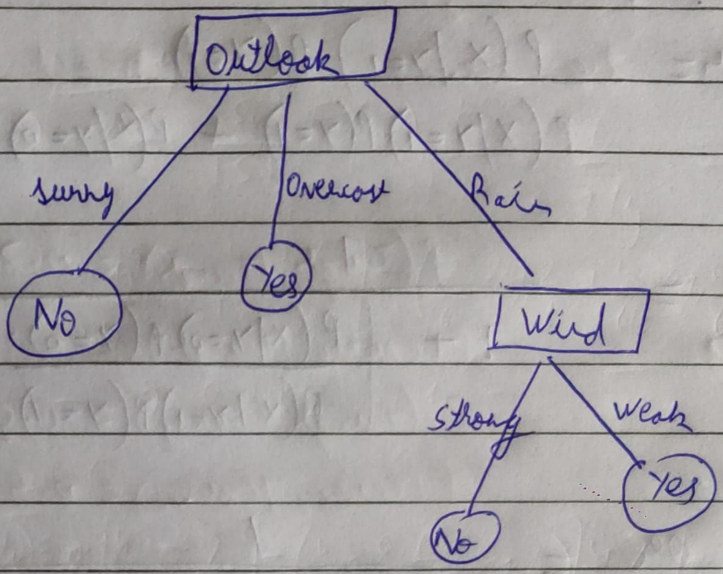
b) Yes it is possible, Take set as $\{D1, D2, D4, D11, D10, D12\}$

For this set climate attribute has entropy 0.

now we know that ① + ② = 1

$\therefore$ $P(w_4) = 0.4$

(c) Tree using (D1 – D7) as training set



Test accuracy = $\frac{5}{7}$

Train accuracy = 1

(d) We can add constraints on the minimal number of training examples in a leaf node, since a leaf node that includes only 1 example is likely to be over specific.

⑥

$w1 = \text{Tough}$

$w2 = \text{Course}$

$w3 = ?$

$w4 = \text{Course}$

$P(\text{Tough} / \text{Tough}) = 0.7$

$P(\text{course} / \text{Tough}) = 0.3$

$P(\text{tough} / \text{course}) = 0.5$

$P(\text{course} / \text{course}) = 0.5$

To find $P\left(\frac{w_3}{w_1, w_2 w_4}\right)$

By markov's assumption

$$P\left(\frac{w_3}{w_1 w_2 w_4}\right) = P\left(\frac{w_3}{w_2 w_4}\right)$$

$$= \frac{P\left(\frac{w_4}{w_3 w_2}\right) \cdot P\left(\frac{w_3}{w_4}\right)}{P\left(\frac{w_4}{w_2}\right)}$$

$$= \frac{P\left(\frac{w_4}{w_3}\right) \cdot P\left(\frac{w_3}{w_2}\right)}{P\left(\frac{w_4}{w_2}\right)}$$

$$= \frac{P\left(\frac{w_4}{w_3}\right) \cdot P\left(\frac{w_3}{w_2}\right)}{P(w_4)} \quad \{\text{markov's}\}$$

$$P\left(\frac{w_3 = \text{tough}}{w_2, w_4}\right) = \frac{P\left(\frac{\text{tough}}{\text{course}}\right) \cdot P\left(\frac{\text{course}}{\text{tough}}\right)}{P(w_4)} = \frac{0.15}{P(w_4)} \rightarrow \textcircled{1}$$

$$P\left(\frac{w_3 = \text{course}}{w_2, w_4}\right) = \frac{P\left(\frac{\text{course}}{\text{course}}\right) \cdot P\left(\frac{\cancel{\text{tough}}\ \text{course}}{\text{course}}\right)}{P(w_4)} = \frac{0.25}{P(w_4)} \rightarrow \textcircled{2}$$

now we know that $\textcircled{1} + \textcircled{2} = 1$

$$\therefore \quad P(w_4) = 0.4$$

$$\left\{ \begin{array}{l} \therefore \quad P(w_3 = \text{course}) = \dfrac{0.25}{0.4} = \dfrac{5}{8} \\[2em] P(w_3 = \text{tough}) = \dfrac{0.15}{0.4} = \dfrac{3}{8} \end{array} \right.$$

(2) a) Logistic regression treats each feature independently whereas decision trees do not assume each input feature to be independent and can thus encode complicated formulas related to relationship b/w these variables

b) Decision trees generally overfit on the data since they can split on different-combination of features, whereas logistic regression associates only 1 parameter with each feature

c) Yes, since data is linearly separable, for each x_1 value there is a cutoff on x_2.
Splittig data based on (x_1) can be done with a tree of depth $\{\log(n)\}$.

d) Yes, Here we may need to consider diff. values for (x_2).
This can be done in at most log(n) depths.

(8) we know

$$P(Y=1/X) = \frac{P(X/Y=1)\,P(Y=1)}{P(X/Y=1)\,P(Y=1) + P(X/Y=0)\,P(Y=0)}$$

$$= \frac{1}{1 + \left[\frac{P(X|Y=0)\,P(Y=0)}{P(X|Y=1)\,P(Y=1)}\right]}$$

or $$= \frac{1}{1 + \exp\left[\ln\left(\frac{P(X|Y=0)\,P(Y=0)}{P(X|Y=1)\,P(Y=1)}\right)\right]}$$

$$= \frac{1}{1 + \exp\left[\ln\left(\frac{P(Y=0)}{P(Y=1)}\right) + \sum_i \ln\left(\frac{P(X_i|Y=0)}{P(X_i|Y=1)}\right)\right]}$$

$P(Y=1) = \pi$ $\therefore P(Y=0) = 1-\pi$

each (X_i) has binomial distribution

$P(X_i|Y=0) = \theta_{i0}^{X_i}(1-\theta_{i0})^{(1-X_i)}$

$P(X_i|Y=1) = \theta_{i1}^{X_i}(1-\theta_{i1})^{(1-X_i)}$

$$\therefore P(Y=1|X) = \frac{1}{1 + \exp\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_i \ln\left[\frac{\theta_{i0}^{X_i}(1-\theta_{i0})^{(1-X_i)}}{\theta_{i1}^{X_i}(1-\theta_{i1})^{(1-X_i)}}\right]\right]}$$

$$= \frac{1}{1 + \exp\left[\ln\left(\frac{1-\pi}{\pi}\right) + \sum_i X_i \ln\left(\frac{\theta_{i0}}{\theta_{i1}}\right) + (1-X_i)\ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)\right]}$$

$$= \frac{1}{1 + \exp\left(\ln\left(\frac{1-\pi}{\pi}\right) + \left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right) + \sum_i x_i \left[\ln \frac{\theta_{i0}}{\theta_{i1}} - \ln \frac{(1-\theta_{i0})}{(1-\theta_{i1})}\right]\right)}$$

let $w_0 = \ln\left(\frac{1-\pi}{\pi}\right) + \sum_i \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)$

$w_i = \ln\left(\frac{\theta_{i0}}{\theta_{i1}}\right) - \ln\left(\frac{1-\theta_{i0}}{1-\theta_{i1}}\right)$

$$\therefore \quad P(Y=1/X) = \frac{1}{1+\exp(\Sigma_i w_i x_i)}$$

Hence proved.