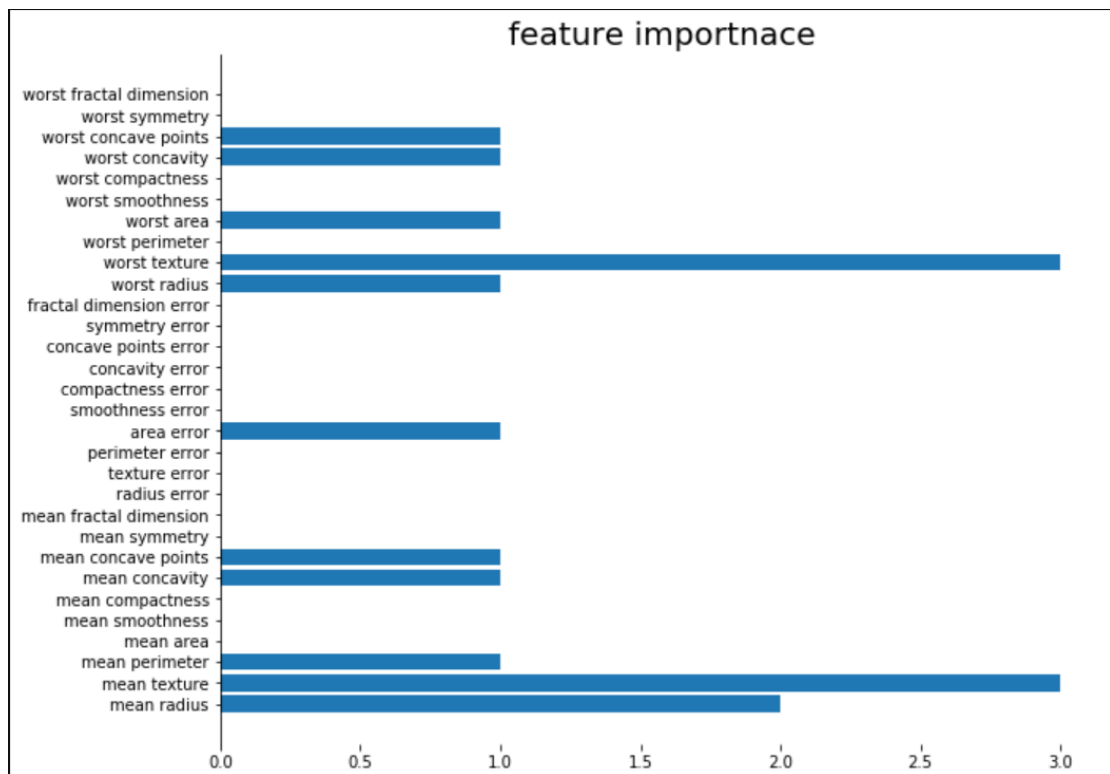


Part 1

1. Gini of data is 0.4628099173553719
Entropy of data is 0.9456603046006402
2. Accuracy % of clf_depth3 is 0.9230769230769231
Accuracy % of clf_depth10 is 0.916083916083916
Accuracy % of clf_gini is 0.9230769230769231
Accuracy % of clf_entropy is 0.9370629370629371



3. Accuracy % of clf_10tree is 0.9440559440559441
Accuracy % of clf_100tree is 0.951048951048951
Accuracy % of clf_all_features is 0.9230769230769231
Accuracy % of clf_random_features is 0.9370629370629371

Part 2

1. Model A: $\frac{66}{100} \times \frac{66}{100} = 200$, misclassification rate: $\frac{200}{800} = 25\%$

Model B: $\frac{66}{300} \times \frac{66}{300} = 200$, misclassification rate: $\frac{200}{800} = 25\%$

\Rightarrow - $\frac{1}{4}$

Model A : entropy : first node: $-\left(\frac{3}{4} \times \log_2\left(\frac{3}{4}\right) + \frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right)$
second node: $-\left(\frac{3}{4} \times \log_2\left(\frac{3}{4}\right) + \frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right)$
entropy: $(\text{first node} + \text{second node}) \times \frac{1}{2} = 0.81$

gini : first node: $1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2$
second node: $1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2$
gini: $(\text{first node} + \text{second node}) \times \frac{1}{2} = 0.375$

Model B : entropy : first node: $-\left(\frac{1}{3} \times \log_2\left(\frac{1}{3}\right) + \frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right)$
second node: 0

entropy: first node $\times \frac{3}{4} = 0.69$

gini : first node: $1 - \left(\frac{1}{3}\right)^2 - \left(\frac{2}{3}\right)^2$
second node: 0

gini: first node $\times \frac{3}{4} = 0.3$