Q1.

a)

Timeline

Description automatically generated

Decision Tree Plot

Training Accuracy: 0.8

Testing Accuracy: 0.25

b)

Shape

Description automatically generated

Min sample value of 8,9 shows the best testing accuracy.

c)

Timeline

Description automatically generated

Decision Tree Plot

Training Accuracy: 0.8

Testing Accuracy: 0.25

d)

Text, letter

Description automatically generated

e)

Text, letter

Description automatically generated

f)The right way to solve the problem is to find a difference of gini index before and after splitting and choosing one with smallest change, but for convenience I have only calculated the gini value after splitting and chose the attribute with highest value. Since the gini value before splitting is same for every attribute, highest gini value will return lowest change of gini value.

Text, letter

Description automatically generated

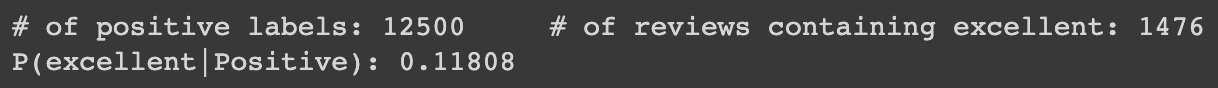
Q2.

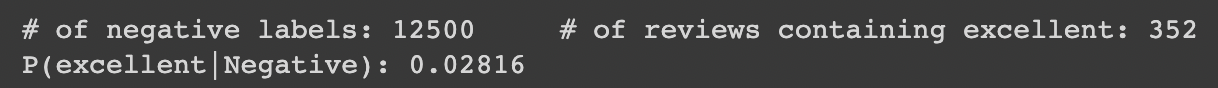
a)

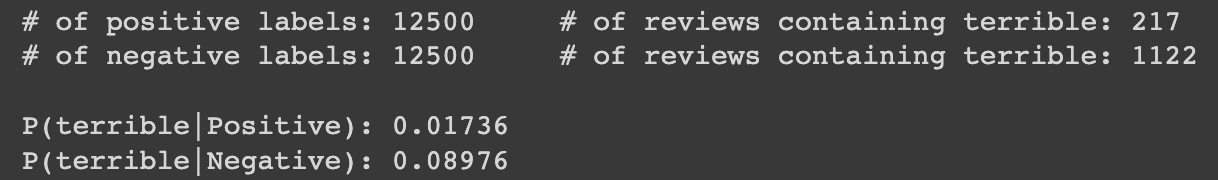
Training Accuracy: 1.0

Testing Accuracy: 0.82672

b)







It is reasonable as the positive word(excellent) associates well with the positive sentiment and the negative word(terrible) associates well with the negative sentiment.