

1. Why is the value of the Accuracy different than the one given for NB in Step 5.4?
  - a. Since we used a smaller amount of training data for our model, our accuracy will decrease in accuracy since there will be less of a fit (80% vs 50% of data used). Since we have less of a fit in our model, our accuracy will go down, thus leading to our lower accuracy. We can also attribute this result to using a different classification type. In the website, we used SVM instead of the Naïve Bayes model in our results.
2. Why is the confusion matrix different than the one given in Step 6.2?
  - a. We can look back to the reason why our accuracy was different from question 1. Since we are using a 50 / 50 split between our training and testing data, we can see we are totaling only 75 values instead of 30 that we see online. We also see a different ratio of values within the confusion matrix. The confusion matrix shows the number of errors that occur in our classification report. Our confusion matrix is showing 2 more misses than the one online. This is because we are using less data and a different classification model to train our model
3. Why are the P, R, and F1 scores different than those given in Step 6.2?
  - a. We can look at two main reasons for our P, R, and F1 scores being different. Inside of step 6.2, we used the SVM model as it was the most accurate. Instead, we used a simpler Naive Bayes model for ours. We can also attribute the difference into the amount of data used inside our model. We only used 75 total to train the model, making it less accurate. Because the model was less accurate in some categories, we can see a slight difference between precession, recall, and f1-score from the website and our results.
4. How many samples were in the training set?
  - a. 75 samples
5. How many samples were in the test set?
  - a. 75 samples