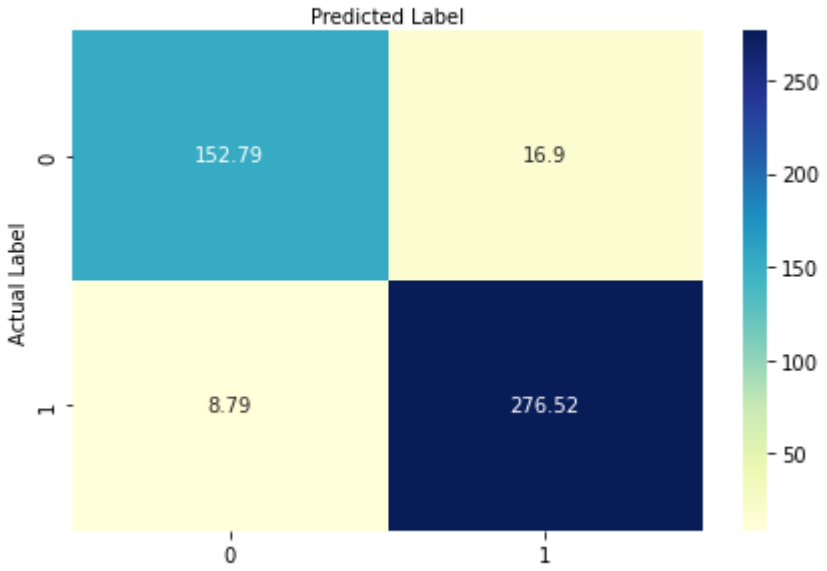
https://github.com/Christian-Martens-UNCC/ECGR-4105/tree/main/Homework_3-Naive_Bayes

Problem 1:

A) The results from this test were very similar to the results from Homework 2. Proportionally, the Naïve Bayes model had few true negatives, but this could be due to the same size or the fact that this is an averaged matrix whereas Homework 2 was using a random training class, so there is bound to be some slight variation.

```
The Average Recall, Precision, and Accuracy over 100 Iterations:
Recall = 0.96919
Precision = 0.90041
Accuracy = 0.94354
```

Figure 1 - Average Confusion Matrix for Problem 1

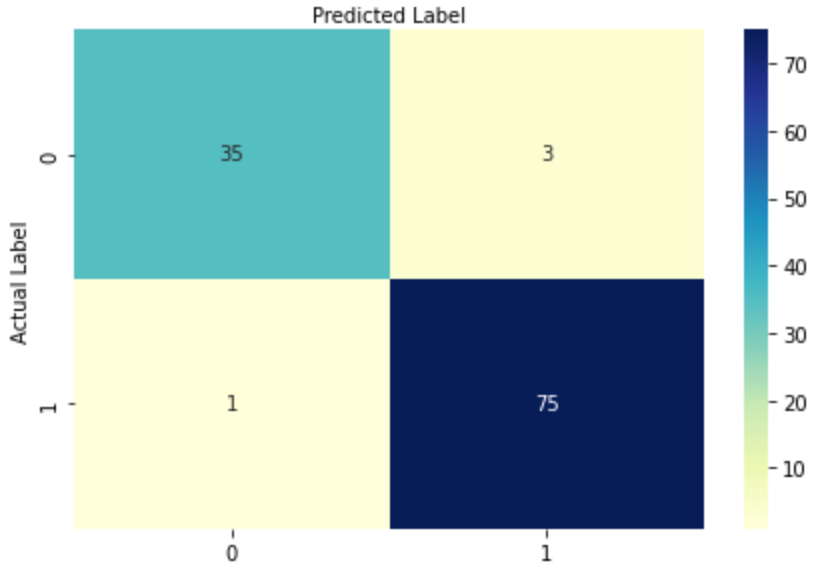


Problem 2:

A) Using a set seed, I found that the maximum accuracy occurred when C = 2 and K = 4. The accuracy was 0.964912

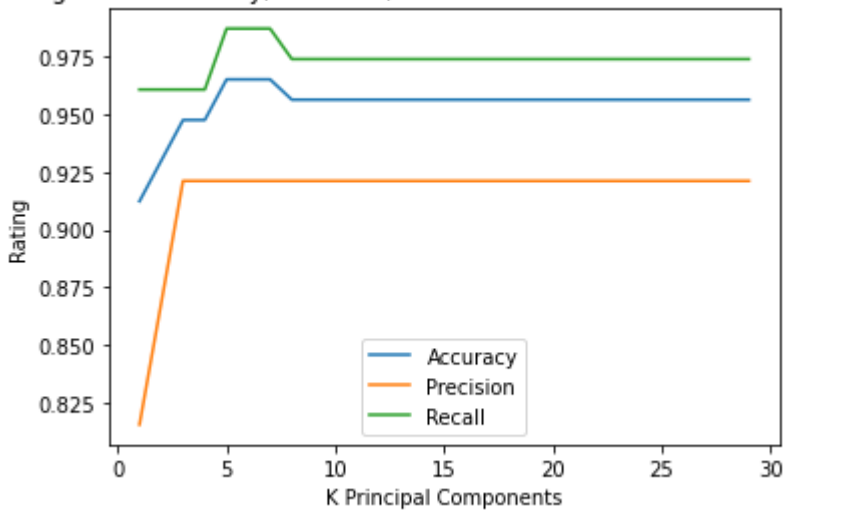
```
Max Test Accuracy: 0.9649122807017544
C: 2
N: 4
```

Figure 2 - Optimal Confusion Matrix for Problem 2, C = 2 & N = 4



B) It appears that, as the number of principal components increases, the accuracy, precision, and recall increase until around K = 8, at which point the accuracy, precision, and recall all trend slightly downward and plateau. This goes to show that having more features doesn't guarantee a higher accuracy and will only add to training time of the model. The maximum occurs when K = 5 principal components.

Figure 3 - Accuracy, Precision, and Recall for Problem 2 for Various K



Problem 3:

A) Using the Bayes model, it seems like adding more than about 10 principal components really caused the model to struggle. The Logarithmic Regression model was more accurate, precise, and recollective over the whole, but the precision was higher for the Naïve Bayes model where 5 < K < 9.

Figure 4 - Accuracy, Precision, and Recall for Problem 3 for Various K