Nahush D. Tambe - 801060297

https://github.com/Ntambe25

Problem 1:

```
              precision    recall  f1-score   support

           0       0.91      0.91      0.91        47
           1       0.94      0.94      0.94        67

    accuracy                           0.93       114
   macro avg       0.93      0.93      0.93       114
weighted avg       0.93      0.93      0.93       114
```
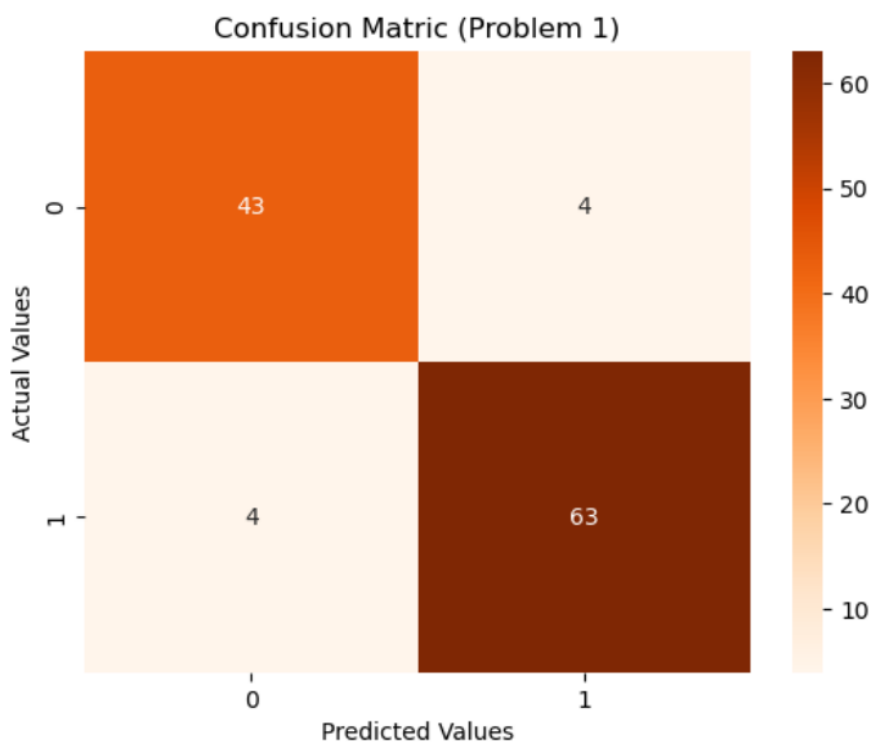
Figure 1: Classification Report for Problem 1

Figure 2: Confusion Matrix for Problem 1

After printing the classification report and plotting the confusion matrix, the precision and recall have values that are between 91 to 94 %, which is a fairly respectable number. The accuracy is 93 %, which is also respectable number. The shape of the Precision recall curve points to the fact that the model is good. When compared to HW # 2, Figure below, it can be concluded that the Naïve Bayes model proves to be less useful when working with this dataset. The numerical comparison shows that the Accuracy, Precision, and Recall all have a greater percentage for the Logistic Regression model at around 97 % than the Naïve Bayes model which is around 94%.

```
print("Accuracy = ", Accuracy)
print("Precision = ", Precision)
print("Recall = ", Recall)

Accuracy =  0.9649122807017544
Precision =  0.9701492537313433
Recall =  0.9701492537313433
```

Problem 2:

```
Optimal value of K: 11
Accuracy: 0.9649122807017544
Precision: 0.9701492537313433
Recall: 0.9701492537313433
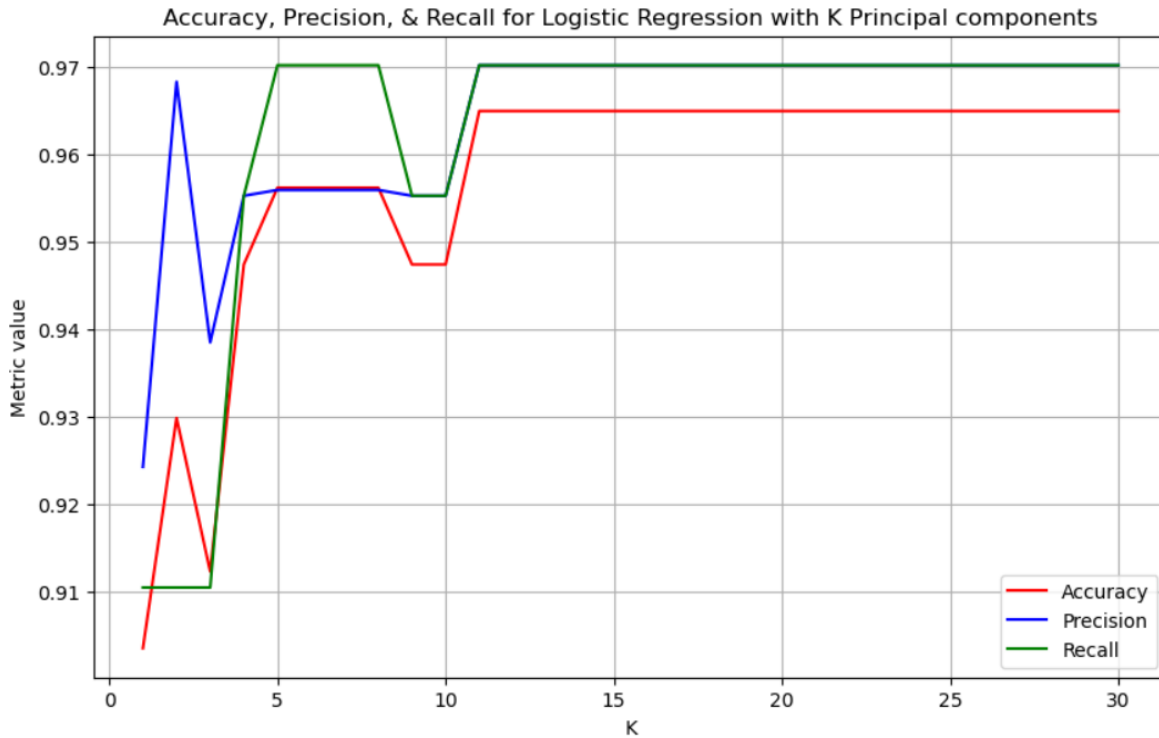```

**Figure 3: Results for Problem 2**

**Figure 4: Plot for Problem 2**

For problem 2, Logistic Regression was used along with the PCA Feature Extraction. Figure 3 and Figure 4 both show the results and the plot for the Accuracy, Precision and Recall. It can be seen that the optimal value of K was found to be 11 that gives the best results. Accuracy, Precision and Recall have a value that is around 97%.

Problem 3:

```
Optimal value of K: 5
Accuracy: 0.9210526315789473
Precision: 0.9142857142857143
Recall: 0.9552238805970149
```
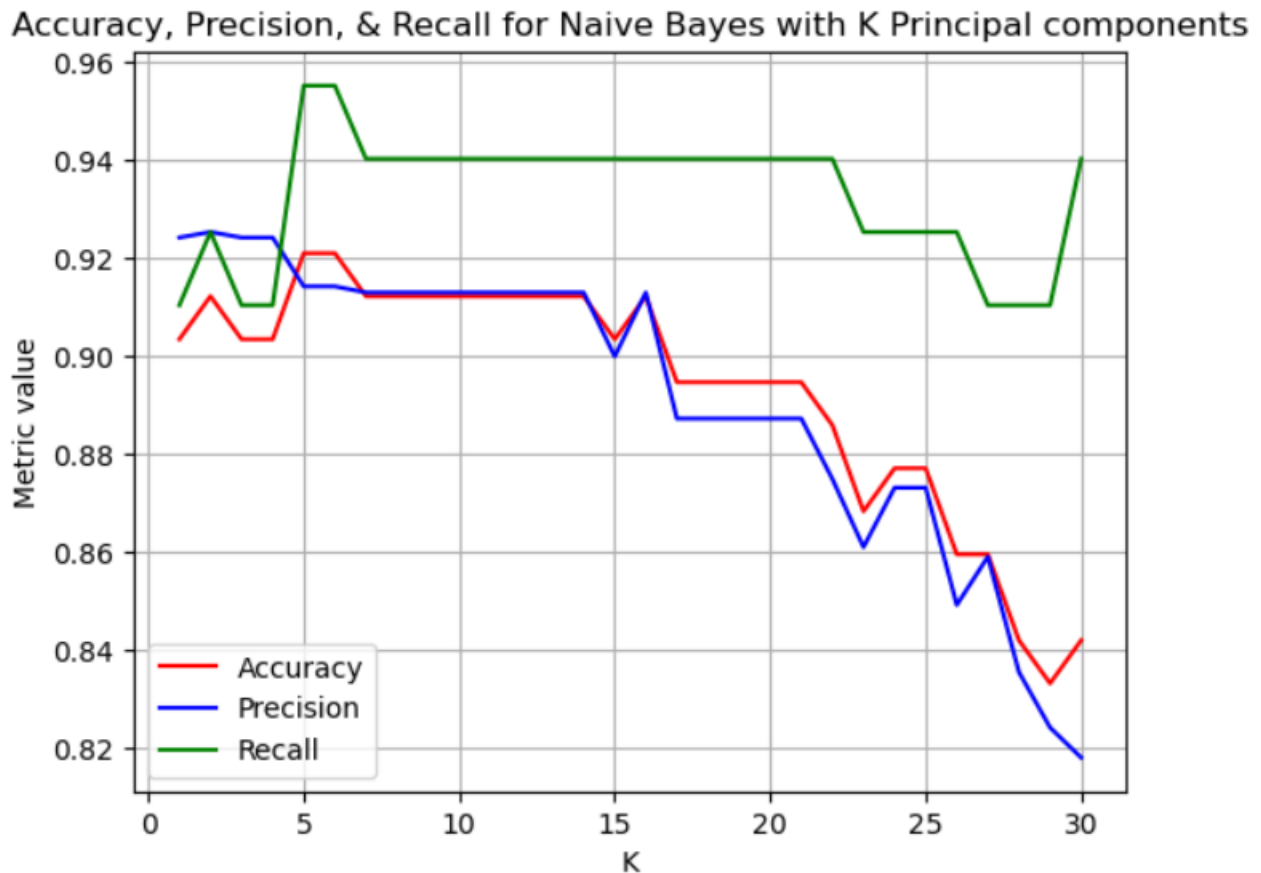
**Figure 5: Results for Problem 3**



**Figure 6: Plot for Problem 3**

Similar to earlier problem, for problem 3, Naïve Bayes was used instead of Logistic Regression along with the PCA Feature Extraction. Figure 5 and Figure 6 both show the results and the plot for the Accuracy, Precision and Recall. It can be seen that the optimal value of K was found to be 5 that gives the best results. Accuracy and Precision have a value that is around 92%, while Recall has a value that is around 95%. In comparison to problem 2, all 3, Accuracy, Precision, and Recall have a value that is less than the values for the Logistic Regression model.

Thus, it can be safely concluded that the Logistic Regression works best for this dataset to provide the best results.