

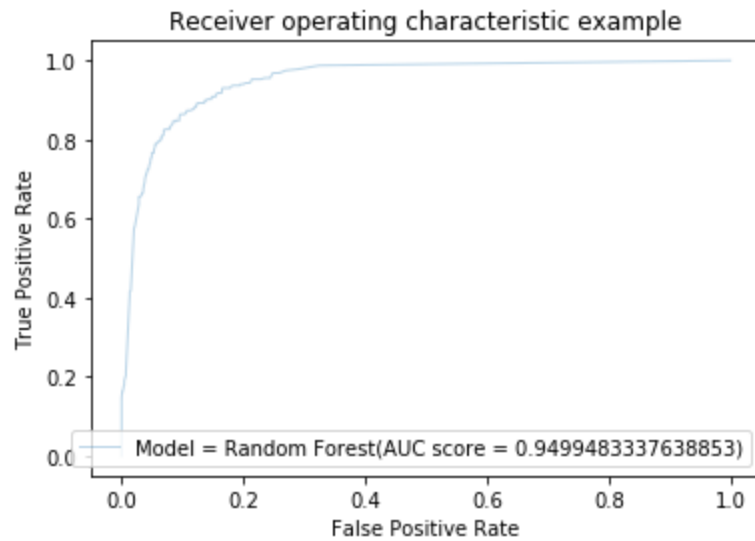
**AUC** for Model: **Random Forest** is: 0.9499483337638853

**Accuracy:** 0.9099099099099099%

**Precision:** 0.7980769230769231%

**Recall:** 0.7879746835443038%

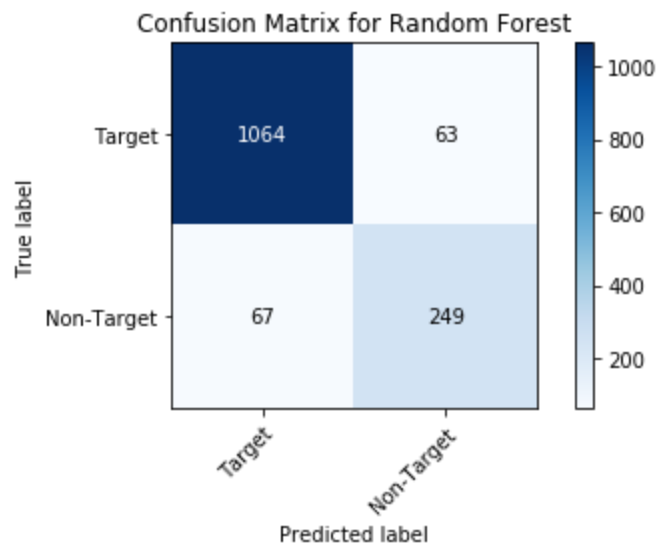
**F1 score:**0.7929936305732483%



Confusion matrix, without normalization

```
[[1064  63]
```

```
 [ 67 249]]
```



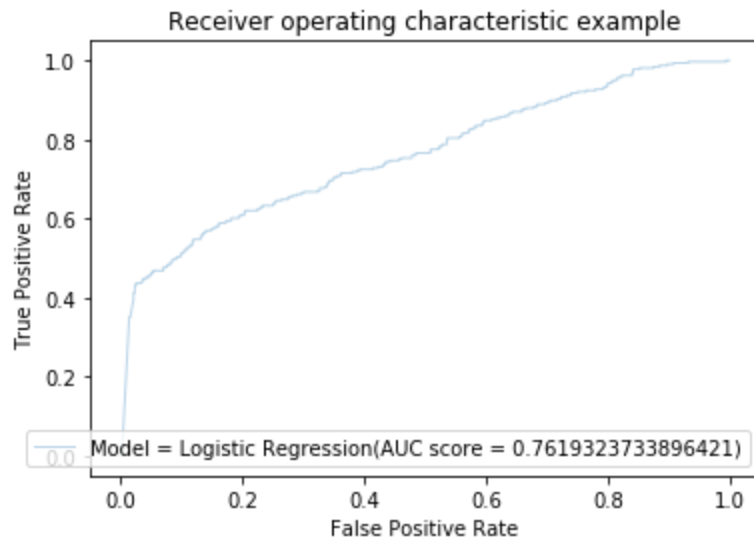
AUC for Model: **Logistic Regression** is: 0.7619323733896421

Accuracy: 0.8558558558558559%

Precision: 0.8214285714285714%

Recall: 0.43670886075949367%

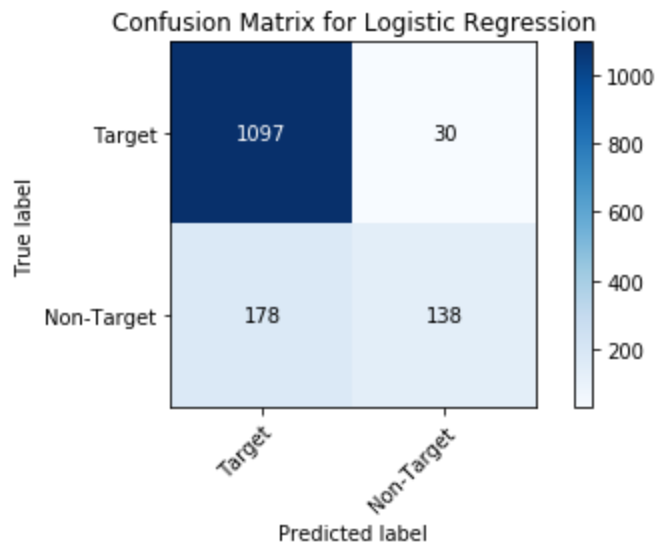
F1 score:0.5702479338842975%



Confusion matrix, without normalization

```
[[1097  30]
```

```
[ 178 138]]
```



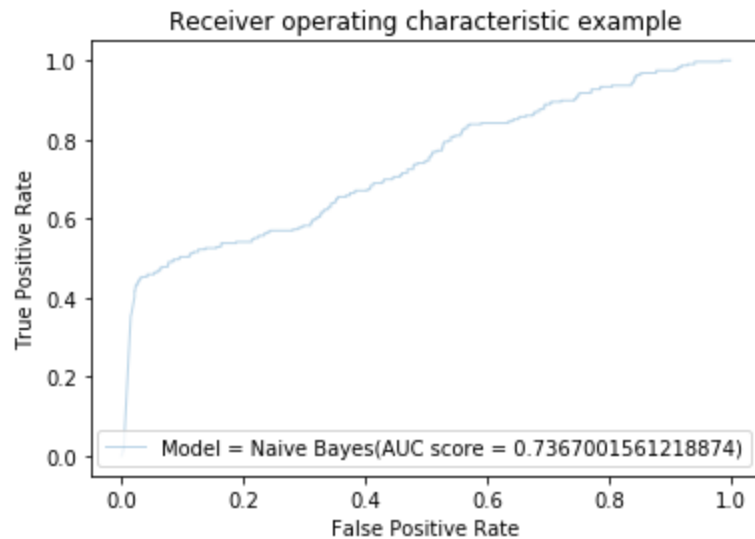
AUC for Model: **Naive Bayes** is: 0.7367001561218874

Accuracy: 0.8482328482328483%

Precision: 0.7512953367875648%

Recall: 0.4588607594936709%

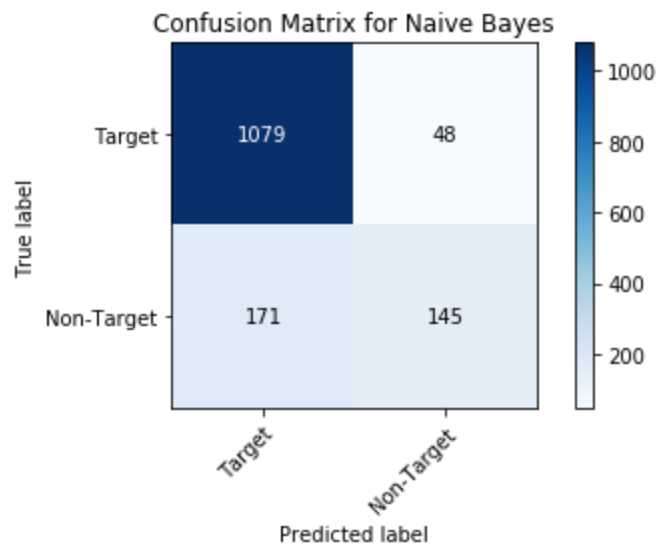
F1 score:0.5697445972495089%



Confusion matrix, without normalization

```
[[1079  48]
```

```
[ 171 145]]
```



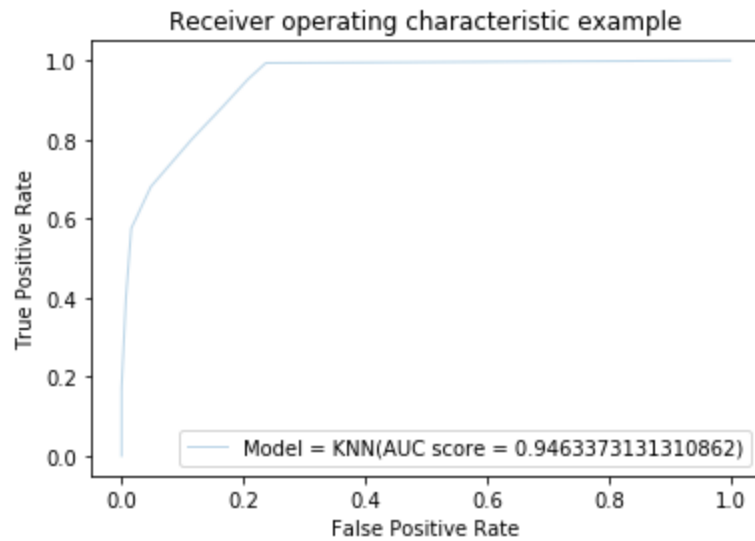
AUC for Model: **KNN** is: 0.9463373131310862

Accuracy: 0.8925848925848926%

Precision: 0.7992565055762082%

Recall: 0.680379746835443%

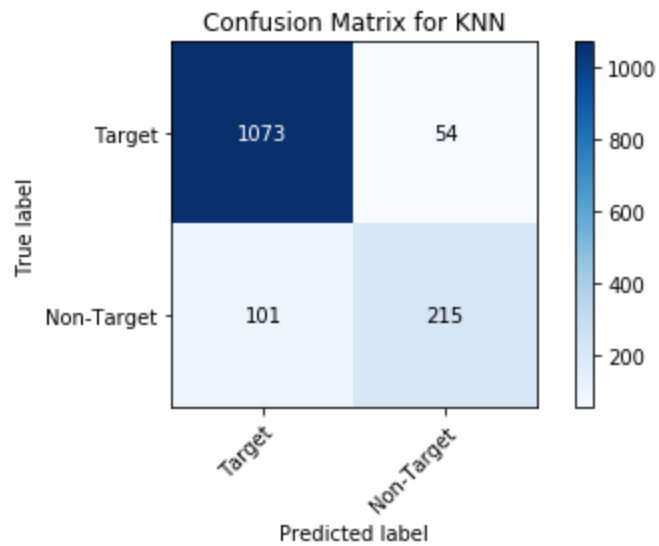
F1 score:0.735042735042735%



Confusion matrix, without normalization

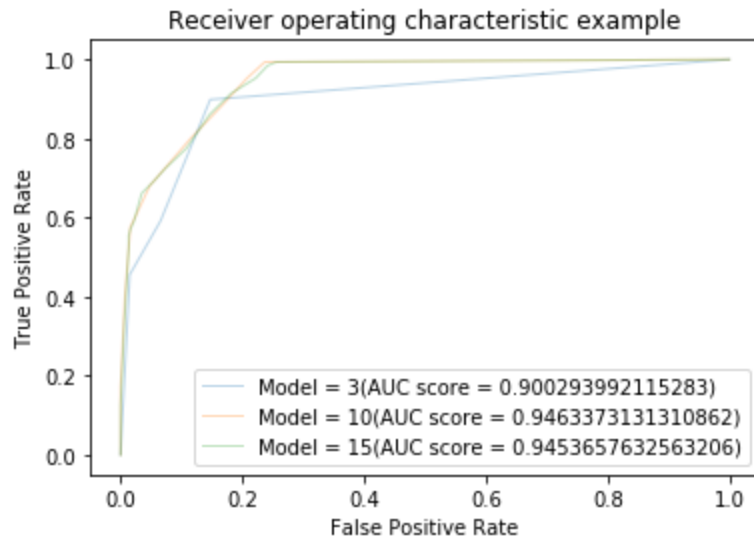
```
[[1073  54]
```

```
 [ 101 215]]
```



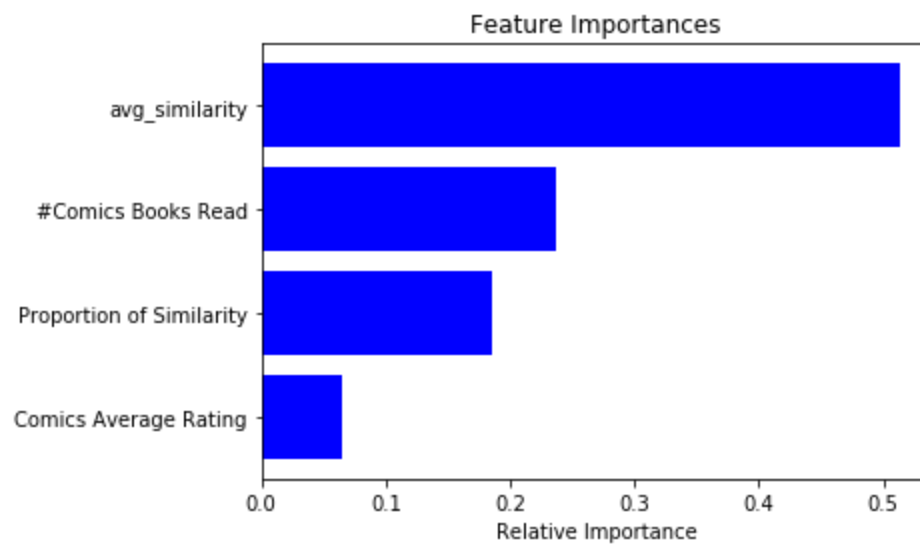
## Hyperparameter Testing

With knn, we tried different values of k, and here are the ROC curves for those



With this, we see that we get the best results when **k=10**, with **AUC = 0.946**

**The use of forests of trees to evaluate the importance of features on our classification task**



### **Current dataset statistics (which we are using)**

**Books: 1274130**

**User i teractions: 655101**

**Users: 144360**