

**1. Use the attached confusion matrix to answer the questions below:**

**a.  What is the total accuracy of the model on each partition?**



**Model 1:**

**Accuracy = (4293 + 5315) / 12000**

**= 9608/12000**

**= 80.06%**

**Model 2:**

**Accuracy = (4307 + 5286) / 12000**

**= 9608/12000**

**= 79.94%**

**b.  Assume the cost of false negative is $500 and the cost of a false positive $20, what is the cost of the model for each partition?**

**Cost of false negative - $500**

**Cost of false positive - $20**

**Model 1:**

**Total Loss Cost = 1707\*$500 + 685\*$20**

**= $853,500 + $13700**

**= $867,200**

**Model 2:**

**Total Loss Cost = 1693\*$500 + 714\*$20**

**= $846,500 + $14280**

**= $881,480**

**c.   What is the accuracy of a Yes prediction for each partition?**

**Model 1:**

**Accuracy = 4293 / 6000**

**= 71.55%**

**Model 2:**

**Accuracy = 4307 / 6000**

**= 71.78%**

**d.  Do you think this model is overfitted to the training data?  Why or why not?**

**If the model does much better on the training set than on the validation set, then we’re likely overfitting. But in our example, the accuracy on training model is 80.06% and the accuracy on validation model is 79.94%. Therefore, there is no problem of overfitting since there is no significance variance in their accuracy.**