Different classification algorithms were applied and compared based on recall and F1 score values to select the best model. Models are not 100% accurate in classifying the data and there is a cost associated with misclassification. It is usually determined empirically based on user expertise. For simplicity I am making some assumptions to estimate cost associated with misclassification.

Losing a single customer means you are losing money but what is the real cost of losing customer ? Let us assume *Cost associated to losing each customer is the same and equal to* ***L.*** In this case the modelincorrectly predict that a churning customer will stay (false negative).

*Assume cost of retention associated to each customer is the same and equal to* ***R.*** *By assuming that* each customer will accept the incentive and remain with the company, this includes costs associated with retaining a customer identified as true positive and false positive.

**Cost of retention is less than cost of losing a customer. Therefore, let’s assume L = 4\*R**

Based on all these assumptions cost of false positive, false negative and estimated total saving using the model was calculated based on the output of confusion matrix(check result section).