1. What is the Total Accuracy of the model on each partition?

Table

Description automatically generated

Sol: - Training: Total Rows – 2461, Validation: Total Rows – 1236, Test: Total Rows – 1303

Total accuracy = (True positive + True Negative)/ Total Tows

Training: (156+2196)/2461 = 95.570%

Validation: (83+1105)/1236 = 96.116%

Test: (89+1162)/1303 = 96.009%

1. Assume that it costs $5 to print and send the marketing material to make the offer.  Assume that the average profit earned by the bank on each loan is $500.  Taking those costs into account, what is the Model Cost on each partition?

Sol:- Model cost: False predicted values that cost the business:

Training: 24\*5 + 85\*500 = $42620

Validation: 15\*5 + 33\*500= $16575

Test: 18\*5 + 34\*500 = $17090

1. What is the Accuracy of the Yes Prediction for each partition?

Sol: - Yes prediction accuracy = True Positive / (True Positive+ False Positive)

Training: 156/ (156+24) = 86.66%

Validation: 83/ (83+15) = 84.69%

Test: 89/ (89+18) = 83.177%

1. Do you think the model is overfitting to the Training data?  Why or why not?

Sol: - Error Rate in training (wrongly predicted values/ total): (FN + FP) / Total rows = 85+24/2461 =4.429%

Error Rate in test (wrongly predicted values/ total): (FN + FP) / Total rows = 34+18/1303 = 3.9907%

No, the model is not overfitting, since the model is said to be overfitted only when the Test error rate is > Training error rate, in our scenario, our training data has a high error rate than the test data, hence it’s not overfitting.

1. What is the Lift for each partition?

Sol: - The ratio between the results obtained with and without the model.

Training:

* Lift = Predicted yes/ Actual yes.
* Predicted yes for training partition= 156/ (156+24) = 86.66%
* Actual yes for training partition = (156+85)/2461 = 241/2461 = 9.7927%
* Lift = 86.66/ 9.7927 = 8.849

Validation:

* Lift = Predicted yes/ Actual yes.
* Predicted yes for training partition= 83/ (83+15) = 84.69%
* Actual yes for training partition = (83+33)/1236 = 116/1236 = 9.3851%
* Lift = 84.69/9.3851 = 9.051

Test:

* Lift = Predicted yes/ Actual yes.
* Predicted yes for training partition= 89/ (89+18) = 83.177%
* Actual yes for training partition = (89+34/1303) = 123/1303 = 9.439%
* Lift = 83.177/9.439 = 8.812

1. The confusion matrixes above were formed with a 0.5 cutoff.  If you were to increase the cutoff to 0.7 would your model predict more Yeses or fewer Yeses?  Would your Accuracy of the Yes Prediction increase or decrease?  Would your Lift increase or decrease?

Sol: - When we increase the cutoff to 0.7 then the model predicts fewer yeses, which reduces false positives, relatively which increases the accuracy of yes prediction and the lift also increases. Since all of them are directly proportional.

1. If you were to decrease the cutoff to 0.3 how would that impact your FP and FN counts.  Would they increase or decrease?  Would your Model Cost increase or decrease?

Sol: - Relative to the above answer, when cutoff is decreased, then false positive count increases, but false negative count decreases, Model cost decreases considering false positive count increases and false negative decreases.

Example: Training: 24(FP)\*5 + 85(FN)\*500 = $42620 (referring to 2prob)

If our FP increases and FN decreases, the loss occurring factor is higher in FN, hence we are cutting down our loss. If any value that’s <85 would save us $500, unless FP increases more than 124 there’s no loss to our model if cut off is reduced to 0.3.

“The work contained and presented here is my work and my work alone.”

Vinay Reddy Vangala

NetId: vrv22001 un