Conclusion

In the 299 LoanIds in the test set are predicted as “Y”, which is more than 81%, whereas in the training set only about 69% had this status, and as i noticed earlier, the distributions of the variables in both sets are similar. The distribution of the Loan status is also similar in both sets, this must mean that both models are predicting as “Y” a lot of cases that should be predicted as “N”. There must be some other feature, that hasn’t been taken into account which could help to predict more of these cases correctly. Class error reported by the model is more than 49%.