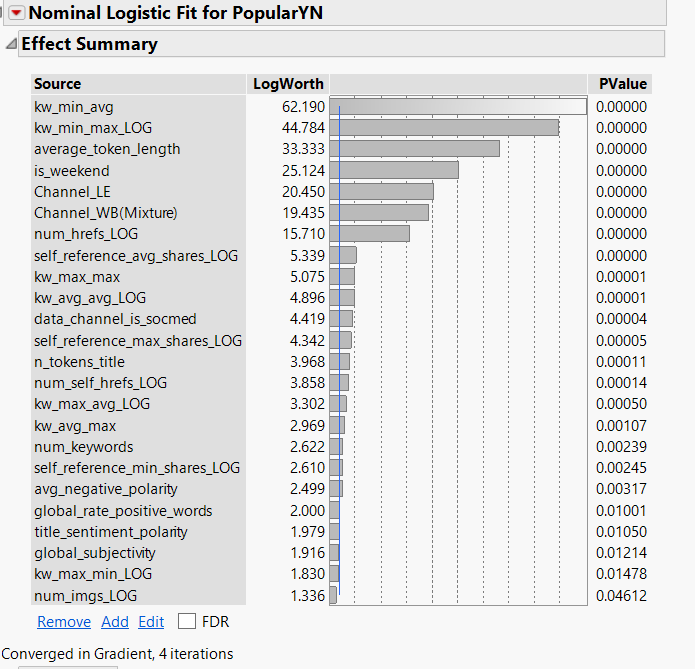
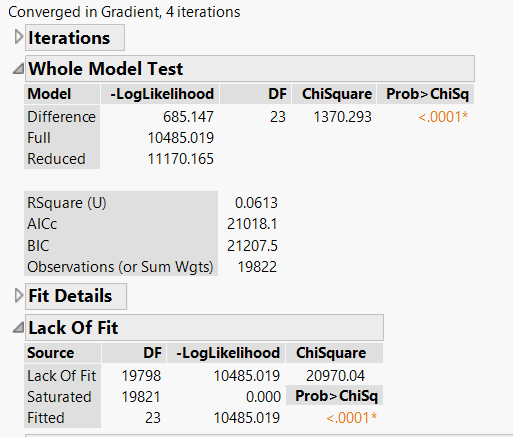
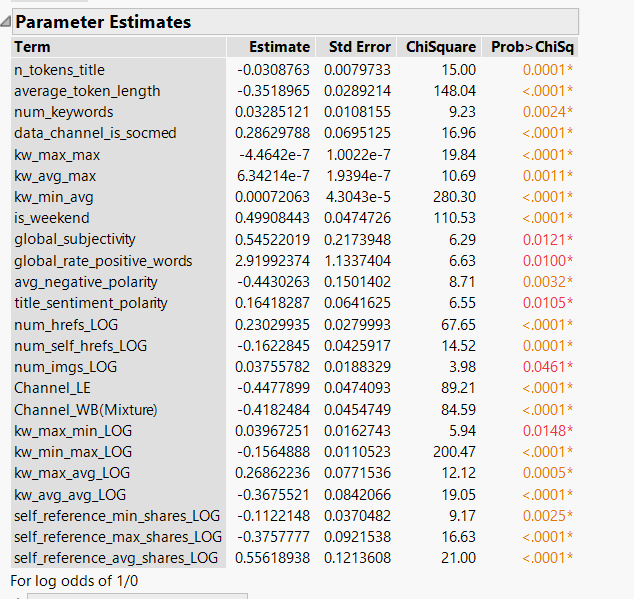
Running Logit first including n\_tokens\_content\_LOG,num\_hrefs\_LOG,num\_self\_hrefs\_LOG,num\_imgs\_LOG,num\_videos\_LOG,Channel\_LE,Channel\_WB,kw\_max\_min\_LOG,kw\_min\_max\_LOG,kw\_max\_avg\_LOG,kw\_avg\_avg\_LOG, self\_reference\_min\_shares\_LOG, self\_reference\_max\_shares\_LOG, self\_reference\_avg\_shares\_LOG and excluding likewise original variables without taking log.

Final significant variables:

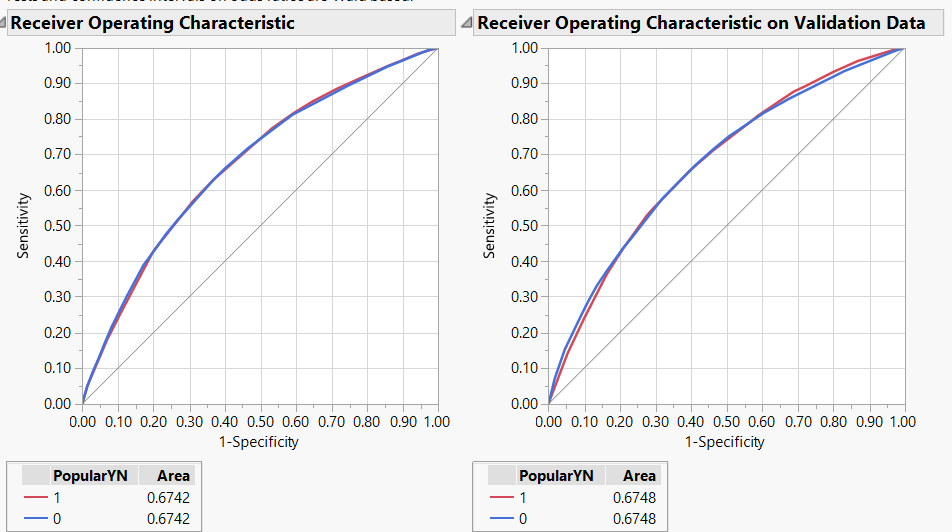




Coefficients:



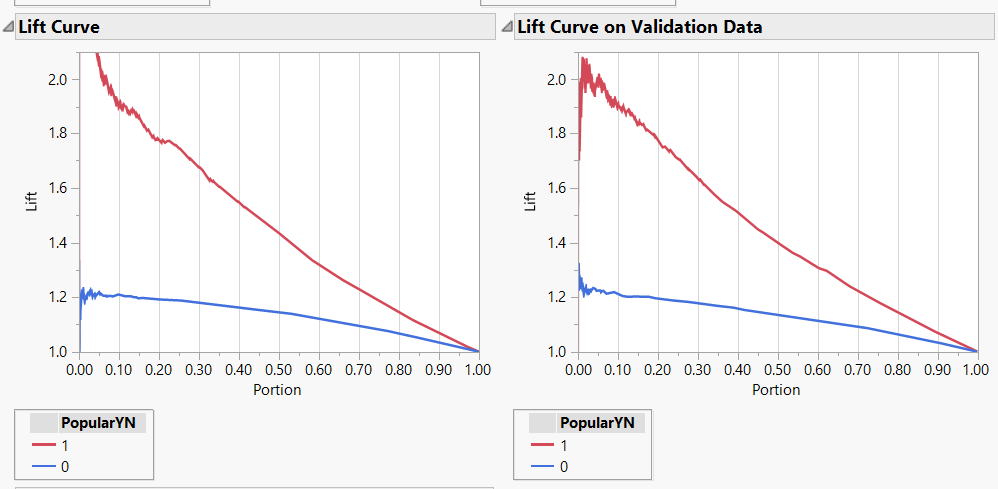
AUC and others:



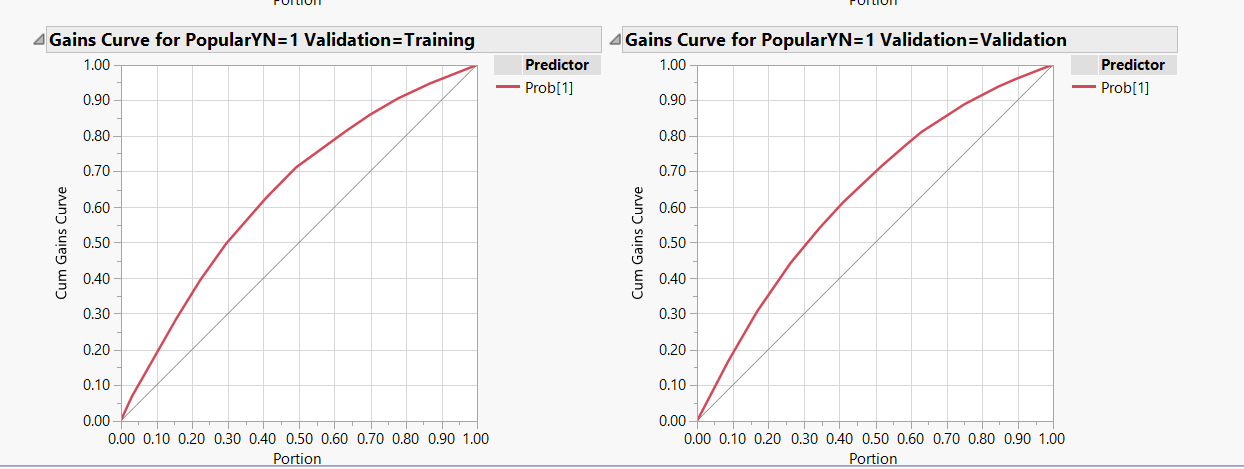


In case of Logit done without any feature engineering our AUC was 0.6842. In this case it is 0.6748. Also Brier Score (0.4181) is lower in case of without log variables compared to with log variables(0.4196) as per Validation data is concerned. Also comparing the Lift curve for validation data sets for both the above two cases, at top two deciles we have a higher lift(more than 1.7) in case of non-log variables compared to 1.7 lift in case of log variables.

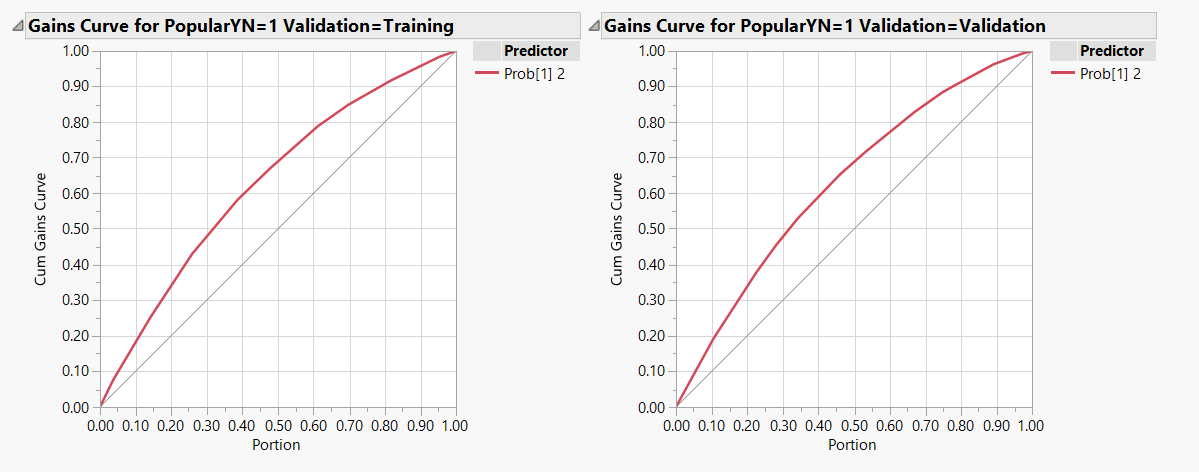
Lift Curve for Non-Log variables:



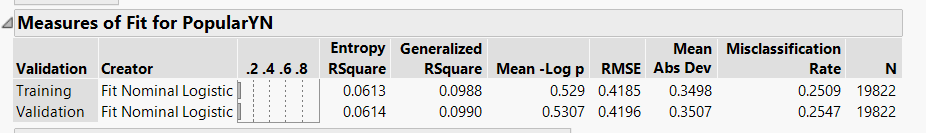
Gains Curve for variables without log:



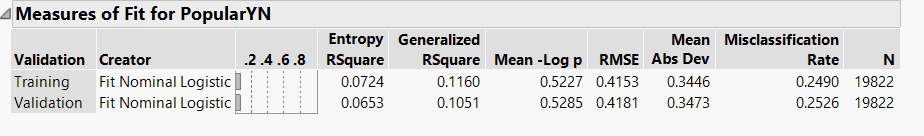
Gains Curve for variables with log:



Measures of fit for variables with log:



Measures of fit for variables without log:



Conclusion: Model with non-log variables seems to be strong-learner.