Great Lakes Institute of Management

PGPM 2019-2020

Predictive Analytics for Marketing(PAM)

**Case Problem-BlueOrRed**

Campaign organizers for both the Republican and Democrat parties are interested in identifying individual undecided voters who would consider voting for their party in an upcoming election. The file *BlueOrRed* contains data on a sample of voters with tracked variables including: whether or not they are undecided regarding their candidate preference, age, whether they own a home, gender, marital status, household size, income, years of education, and whether they attend church.

**Description of the Variables**

|  |
| --- |
| Undecided=0 if remains Undecided, =1 if Decided |
| Age=Age in Years |
| HomeOwner=1 if Owner , =0 Non-Owner |
| Female=1 if Female, = 0 if Male |
| Married=1 if Married, =0 if Unmarried |
| Householdsize=Number of Members |
| Income= $1000 |
| Education=Number of Years spent in School/College |
| Church =1 if attends Church, =0 not attends Church |

Partition the data into training (70 percent), validation (30 percent).

Questions

1. Develop the Predictive Model Logistic Regression, obtain the output and interpret the results
2. Compare the predictive accuracy in terms of Type I error, Type II error, and overall error both for the training and test data. What are the implications to the decision maker?
3. Quantify the relative importance of the predictor variables
4. Compare your results with Discriminant Analysis in terms of Accuracy and relative importance of predictor variables.
5. Write a managerial report giving your final recommendations based on your major findings.