Paper 1863-2014

# Internet Gambling Behavioral Markers: Using the Power of SAS® Enterprise Miner™ 12.1 to Predict High-Risk Internet Gamblers

Sai Vijay Kishore Movva, Vandana Reddy and Dr. Goutam Chakraborty; Oklahoma State University, Stillwater, OK

## ABSTRACT

Using 4,056 subscribers' data about actual gambling behavior over the Internet, we developed behavioral markers which can be used to predict the level of risk that a subscriber is prone to gambling addiction. SAS® Enterprise Miner™ 12.1 is used to build a set of models to predict which subscriber is likely to become a high-risk internet gambler. The data contains 114 variables such as “first active date” and “first active product used” on the website as well as the characteristics of the game such as fixed odds, poker, casino, games, etc. Other measures of a subscriber's data such as money put at stake and what odds are being bet are also included in the analysis. These variables provide a comprehensive view of a subscriber's behavior while gambling over the website. The target variable is modeled as a binary variable, 0 indicating a risky gambler and 1 indicating a controlled gambler. The model comparison algorithm of SAS® Enterprise Miner™ 12.1 is used to determine the best model. The stepwise regression performs the best among a set of 25 models which are run using over 100 permutations of each model. The stepwise regression model predicts a high-risk Internet gambler with an accuracy of 69.63% with variables capturing individual's behavioral patterns.

## INTRODUCTION

For many people, gambling can turn into an addiction associated with a high risk factor. Those in the younger age group (16-24) are most likely to become online gambling addicts. There are many factors which can differentiate a gambling addict from someone who is not. People gamble for many different reasons. Some people do it for fun while some people have money to burn. Some do it as a habit and some are very good at it. Which socio economic or hedonic factor pushes someone to gamble is important but that does not tell us whether that person is likely to become an addict or not. Perhaps this can best be determined by tracking a user's behavior while the person is gambling online over a period of time. In this paper, we analyze data from a gambling website spread across a 4 week period to:

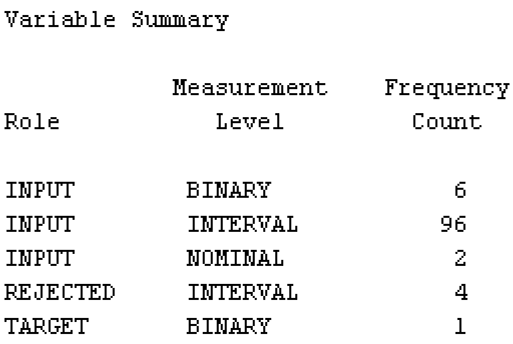
- Determine betting patterns displayed at the time of actual Internet gambling on a betting site that can serve as behavioral markers to predict the development of addiction-related problems.
- Analyze each independent variable in relation to the target variable to determine the most important attributes which influence the risk of gambling addiction.
- Define a specific set of behavioral patterns which are most likely to originate in an internet user who has a possibility of becoming an addict.
- Build a data mining model, which will predict whether a user may become an internet gambling addict.

## DATA PREPARATION

The data was subjected to a few transformations before being used for analysis or model building. The number of available variables in the data set are quite high. Because the variables involved in the data set are actual behavioral markers of a subscriber online, most of these variables need to remain intact for analysis. However there are a few variables which are removed after careful consideration (those with more than 50% missing values).

Most of the variables in the data are continuous and a few are nominal. Variables such as first game played and the most frequent game played have been set as nominal variables with 5 levels. The variables such as risk group1 and risk group2 have been combined into a single variable called risk group combined. After recoding the variables and grouping rare levels and removing variables with high percentage of missing values we have the final set of variables as shown in Display 1.

This paper used data from the Transparency Project (www.thetransparencyproject.org), Division on Addiction, Cambridge Health Alliance, a teaching affiliate of Harvard Medical School.

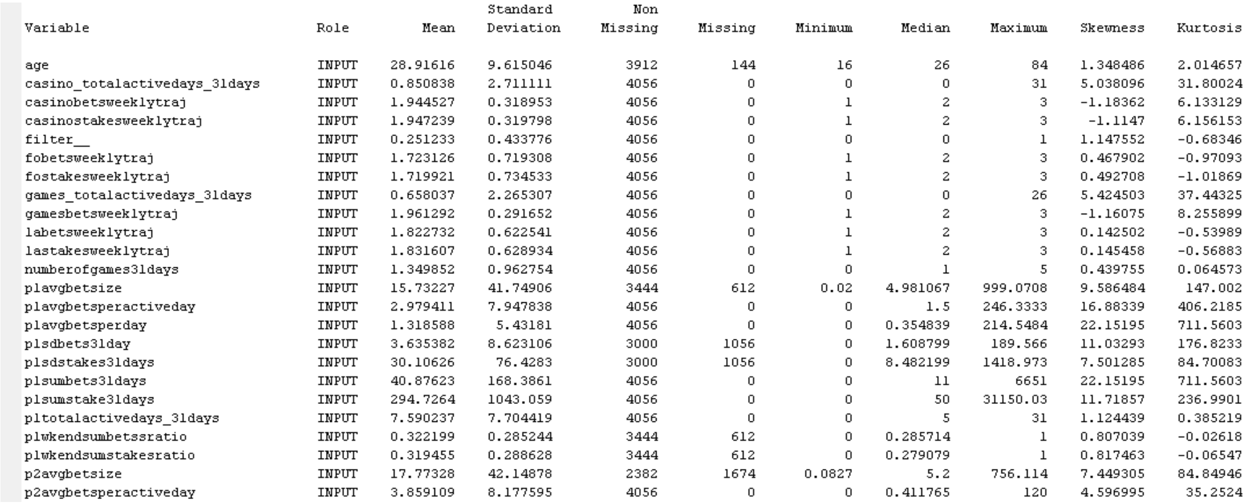


Variable Summary

| Role | Measurement Level | Frequency Count |
|---|---|---|
| INPUT | BINARY | 6 |
| INPUT | INTERVAL | 96 |
| INPUT | NOMINAL | 2 |
| REJECTED | INTERVAL | 4 |
| TARGET | BINARY | 1 |

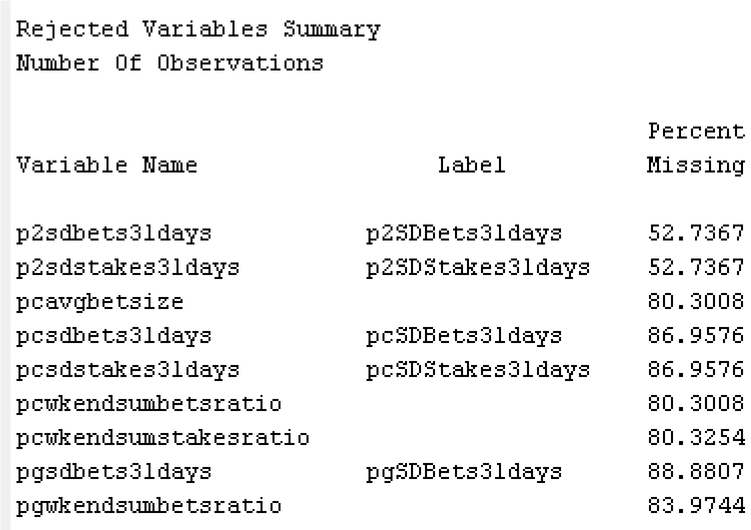
**Display 1. Results from the Metadata Node.**

Many variables have a considerable amount of missing values. These are dealt with using the impute node. First the missing cutoff is set to 50%. The most frequently occurring (mode) class or level is used to substitute for the missing value in a nominal variable. Most of the variables are continuous variables and the missing values for such variables have been imputed by the mean. The stat explore node is used to determine the missing percentages and values of skewness and kurtosis.



| Variable | Role | Mean | Standard Deviation | Non Missing | Missing | Minimum | Median | Maximum | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|
| age | INPUT | 28.91616 | 9.615046 | 3912 | 144 | 16 | 26 | 84 | 1.348486 | 2.014657 |
| casino_totalactivedays_31days | INPUT | 0.850838 | 2.711111 | 4056 | 0 | 0 | 0 | 31 | 5.038096 | 31.80024 |
| casinobetsweeklytraj | INPUT | 1.944527 | 0.318953 | 4056 | 0 | 1 | 2 | 3 | -1.18362 | 6.133129 |
| casinostakesweeklytraj | INPUT | 1.947239 | 0.319798 | 4056 | 0 | 1 | 2 | 3 | -1.1147 | 6.156153 |
| filter__ | INPUT | 0.251233 | 0.433776 | 4056 | 0 | 0 | 0 | 1 | 1.147552 | -0.68346 |
| fobetsweeklytraj | INPUT | 1.723126 | 0.719308 | 4056 | 0 | 1 | 2 | 3 | 0.467902 | -0.97093 |
| fostakesweeklytraj | INPUT | 1.719921 | 0.734533 | 4056 | 0 | 1 | 2 | 3 | 0.492708 | -1.01869 |
| games_totalactivedays_31days | INPUT | 0.658037 | 2.265307 | 4056 | 0 | 0 | 0 | 26 | 5.424503 | 37.44325 |
| gamesbetsweeklytraj | INPUT | 1.961292 | 0.291652 | 4056 | 0 | 1 | 2 | 3 | -1.16075 | 8.255899 |
| labetsweeklytraj | INPUT | 1.822732 | 0.622541 | 4056 | 0 | 1 | 2 | 3 | 0.142502 | -0.53989 |
| lastakesweeklytraj | INPUT | 1.831607 | 0.628934 | 4056 | 0 | 1 | 2 | 3 | 0.145458 | -0.56883 |
| numberofgames31days | INPUT | 1.349852 | 0.962754 | 4056 | 0 | 0 | 1 | 5 | 0.439755 | 0.064573 |
| p1avgbetsize | INPUT | 15.73227 | 41.74906 | 3444 | 612 | 0.02 | 4.981067 | 999.0708 | 9.586484 | 147.002 |
| p1avgbetsperactiveday | INPUT | 2.979411 | 7.947838 | 4056 | 0 | 0 | 1.5 | 246.3333 | 16.88339 | 406.2185 |
| p1avgbetsperday | INPUT | 1.318588 | 5.43181 | 4056 | 0 | 0 | 0.354839 | 214.5484 | 22.15195 | 711.5603 |
| p1sdbets31day | INPUT | 3.635382 | 8.623106 | 3000 | 1056 | 0 | 1.608799 | 189.566 | 11.03293 | 176.8233 |
| p1sdstakes31days | INPUT | 30.10626 | 76.4283 | 3000 | 1056 | 0 | 8.482199 | 1418.973 | 7.501285 | 84.70083 |
| p1sumbets31days | INPUT | 40.87623 | 168.3861 | 4056 | 0 | 0 | 11 | 6651 | 22.15195 | 711.5603 |
| p1sumstake31days | INPUT | 294.7264 | 1043.059 | 4056 | 0 | 0 | 50 | 31150.03 | 11.71857 | 236.9901 |
| p1totalactivedays_31days | INPUT | 7.590237 | 7.704419 | 4056 | 0 | 0 | 5 | 31 | 1.124439 | 0.385219 |
| p1wkendsumbetssratio | INPUT | 0.322199 | 0.285244 | 3444 | 612 | 0 | 0.285714 | 1 | 0.807039 | -0.02618 |
| p1wkendsumstakesratio | INPUT | 0.319455 | 0.288628 | 3444 | 612 | 0 | 0.279079 | 1 | 0.817463 | -0.06547 |
| p2avgbetsize | INPUT | 17.77328 | 42.14878 | 2382 | 1674 | 0.0827 | 5.2 | 756.114 | 7.449305 | 84.84946 |
| p2avgbetsperactiveday | INPUT | 3.859109 | 8.177595 | 4056 | 0 | 0 | 0.411765 | 120 | 4.596995 | 35.2524 |

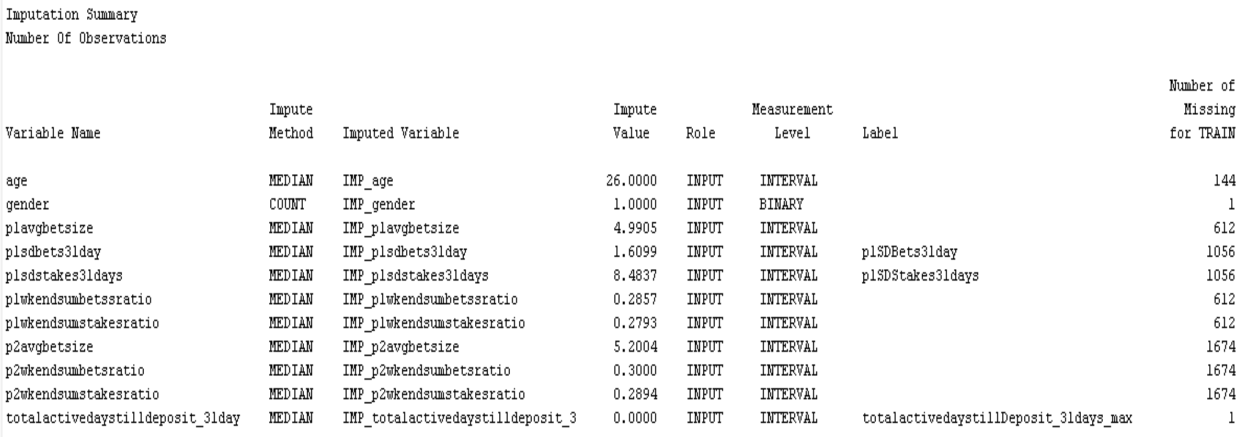
**Display 2. Results from the Stat Explore Node**



Rejected Variables Summary
Number Of Observations

| Variable Name | Label | Percent Missing |
|---|---|---|
| p2sdbets31days | p2SDBets31days | 52.7367 |
| p2sdstakes31days | p2SDStakes31days | 52.7367 |
| pcavgbetsize | | 80.3008 |
| pcsdbets31days | pcSDBets31days | 86.9576 |
| pcsdstakes31days | pcSDStakes31days | 86.9576 |
| pcwkendsumbetsratio | | 80.3008 |
| pcwkendsumstakesratio | | 80.3254 |
| pgsdbets31days | pgSDBets31days | 88.8807 |
| pgwkendsumbetsratio | | 83.9744 |

**Display 3. Results from the Stat Explore Node.**



```
Imputation Summary
Number Of Observations

                                                                                                                            Number of
                                  Impute                                     Impute           Measurement                    Missing
Variable Name                     Method    Imputed Variable                  Value    Role      Level      Label                                     for TRAIN

age                               MEDIAN    IMP_age                         26.0000    INPUT    INTERVAL                                                 144
gender                            COUNT     IMP_gender                       1.0000    INPUT    BINARY                                                     1
plavgbetsize                      MEDIAN    IMP_plavgbetsize                 4.9905    INPUT    INTERVAL                                                 612
p1sdbets31day                     MEDIAN    IMP_p1sdbets31day                1.6099    INPUT    INTERVAL    p1SDBets31day                             1056
p1sdstakes31days                  MEDIAN    IMP_p1sdstakes31days             8.4837    INPUT    INTERVAL    p1SDStakes31days                          1056
p1wkendsumbetssratio              MEDIAN    IMP_p1wkendsumbetssratio         0.2857    INPUT    INTERVAL                                                 612
p1wkendsumstakesratio             MEDIAN    IMP_p1wkendsumstakesratio        0.2793    INPUT    INTERVAL                                                 612
p2avgbetsize                      MEDIAN    IMP_p2avgbetsize                 5.2004    INPUT    INTERVAL                                                1674
p2wkendsumbetsratio               MEDIAN    IMP_p2wkendsumbetsratio          0.3000    INPUT    INTERVAL                                                1674
p2wkendsumstakesratio             MEDIAN    IMP_p2wkendsumstakesratio        0.2894    INPUT    INTERVAL                                                1674
totalactivedaystilldeposit_31day  MEDIAN    IMP_totalactivedaystilldeposit_3 0.0000    INPUT    INTERVAL    totalactivedaystillDeposit_31days_max        1
```

**Display 4. Results from the impute node.**

## TRANSFORMING VARIABLES

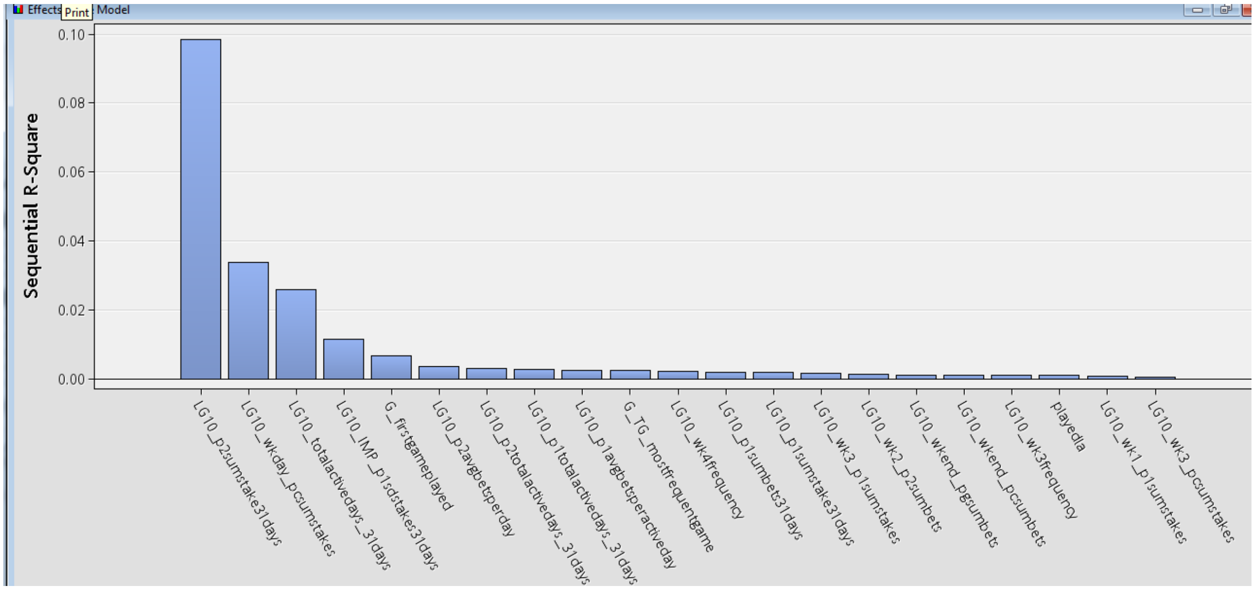
96 variables of the original set of variables are interval variables. Various transformation methods were considered and each transformation method available in SAS® Enterprise Miner™ 12.1 was checked and analyzed to determine which method is performing better in terms of reducing the values of skewness and kurtosis. After all the methods available were run using the transform node and compared, it was found that three simple methods work best during the transformation. These are:

Log — Variable is transformed by taking the natural log of the variable. This is done if there is substantially positive skewness with zero values Formula: New Variable X = Log10(X+C)

Log — Variable is transformed by taking the natural log of the variable. This is done if there is substantially positive skewness with no zero values Formula: New Variable X = Log10(X)

Square Root — Variable is transformed by taking the square root of the variable. This is done if there is moderately positive skewness. Formula: New Variable X=SQrt(x)

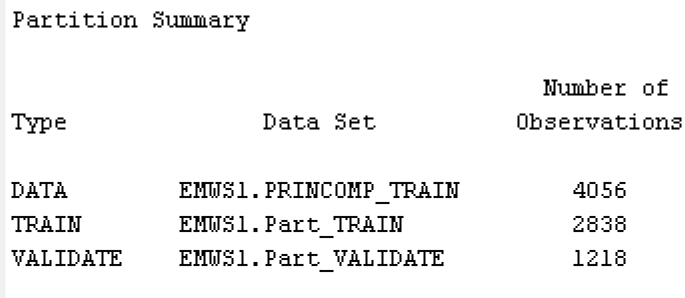
After these transformations have been applied the values of skewness and kurtosis have been considerably reduced. A variable selection node is then run to get a sense of which variables may be the most important variables. These variables are passed on to the modeling nodes as input variables.



**Display 5. Results from the Variable Importance Node.**

## DATA PARTITION

The data needs to be partitioned before it can be used for model building for an honest assessment of models. The data is split into 70% training and 30% validation data. Prior probabilities should be adjusted when the sample proportions of the classes in the training set differ substantially from the proportions in the operational data to be scored. However, in this case of analysis there is not much difference in these values.

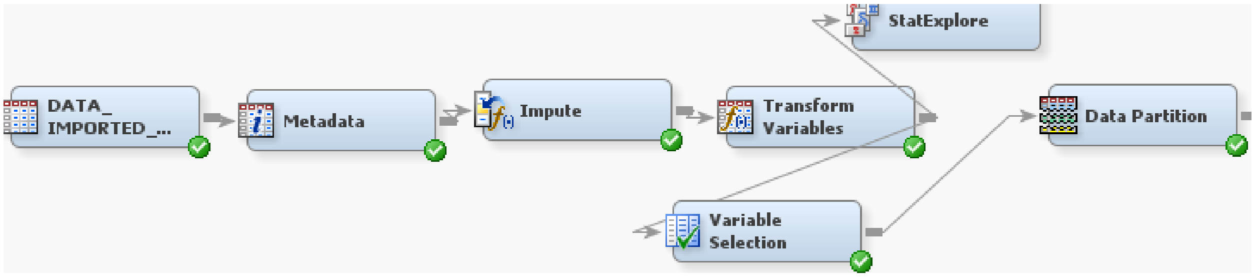


```
Partition Summary


                                      Number of
Type              Data Set           Observations


DATA        EMWS1.PRINCOMP_TRAIN         4056
TRAIN       EMWS1.Part_TRAIN             2838
VALIDATE    EMWS1.Part_VALIDATE          1218
```

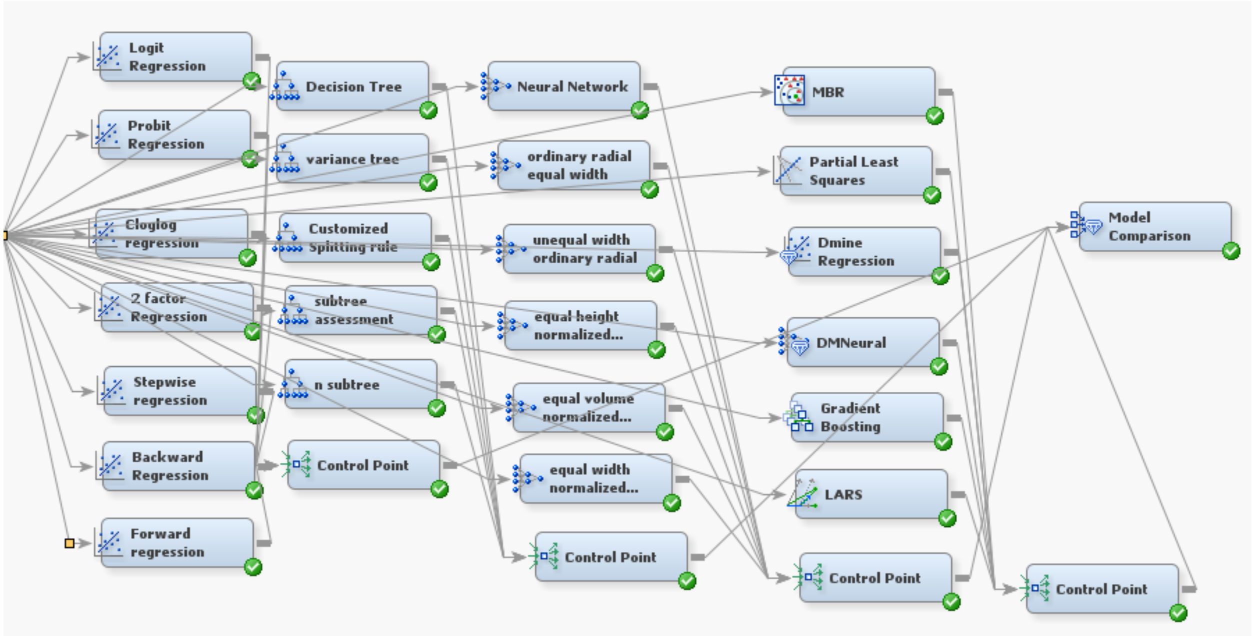
**Display 6. Partition Summary**



**Display 7. Process Flow (Partial).**

## MODEL BUILDING

The first sets of models run are the logistic regression models. The term editor property of the regression node is used to create and add specific variable interactions. The selection criterion for each regression model is set to validation misclassification. A set of neural networks models are also run using a progressively increased number of hidden units and layers with multilayer perception or radial basis with equal width and unequal width. Decision tree models are also run using various options available through SAS® Enterprise Miner™ 12.1. Decision trees have also been used for collapsing the set of categorical values into ranges that are aligned with the values of a selected target variable or value. The gradient boosting model has been used to create a series of decision trees that together form a single predictive model. Each tree in the series has been fit to the residual of the prediction of the trees that have been used in the model. A total of 24 models are selected after analyzing over a 100 models as shown in the flow diagram reported in Display 8.



**Display 8. Model building Process Flow.**

## MODEL COMPARISON:

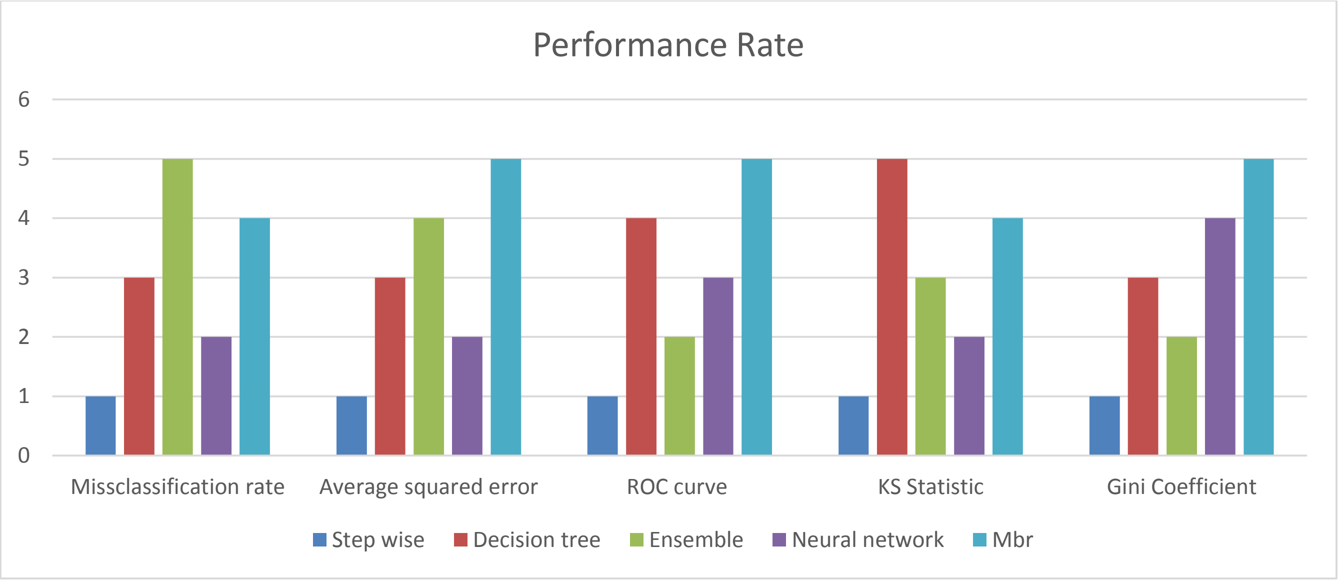
Each of the models were compared using the model comparison algorithm of SAS® Enterprise Miner™ 12.1. The validation misclassification rate is considered the most important criteria for selecting the best model because our goal is to create the best model for predicting who will be a gambling addict. The lower the misclassification rate, the better the model. Misclassification rate is the percent of the outcomes predicted incorrectly. As shown in Table 1, the regression model performs the best with lowest misclassification as well as a low averaged squared error with values of 0.30 and 0.20 respectively.

| Model | Validation Misclassification | Validation ASE |
|---|---|---|
| Stepwise regression | 0.30378 | 0.20114 |
| Backward regression | 0.30378 | 0.20114 |
| Forward regression | 0.30378 | 0.20114 |
| LARS | 0.30624 | 0.20013 |
| Probit regression | 0.30624 | 0.20032 |
| Logit regression | 0.30706 | 0.20009 |
| 2 factor regression | 0.30706 | 0.20009 |
| Partial Least Squares | 0.3087 | 0.20161 |
| Equal volume normalized radial | 0.30952 | 0.20161 |
| Equal height normalized radial | 0.31034 | 0.20185 |
| Neural Network | 0.31117 | 0.20284 |

| | | |
|---|---|---|
| Equal width normalized radial | 0.31117 | 0.20284 |
| Unequal width ordinary radial | 0.31117 | 0.20291 |
| Cloglog regression | 0.31199 | 0.20188 |
| Gradient Boosting | 0.31363 | 0.21355 |
| Customized Splitting rule | 0.31773 | 0.21307 |
| DMNeural | 0.31856 | 0.20873 |
| MBR | 0.32594 | 0.21455 |
| Dmine Regression | 0.32923 | 0.21584 |
| Decision tree | 0.33087 | 0.21623 |
| Variance tree | 0.33087 | 0.21623 |
| Interactive decision tree | 0.34072 | 0.22231 |
| Ordinary radial equal width | 0.49589 | 0.24998 |
| Interactive decision tree 2(n-subtree) | 0.49589 | 0.24998 |

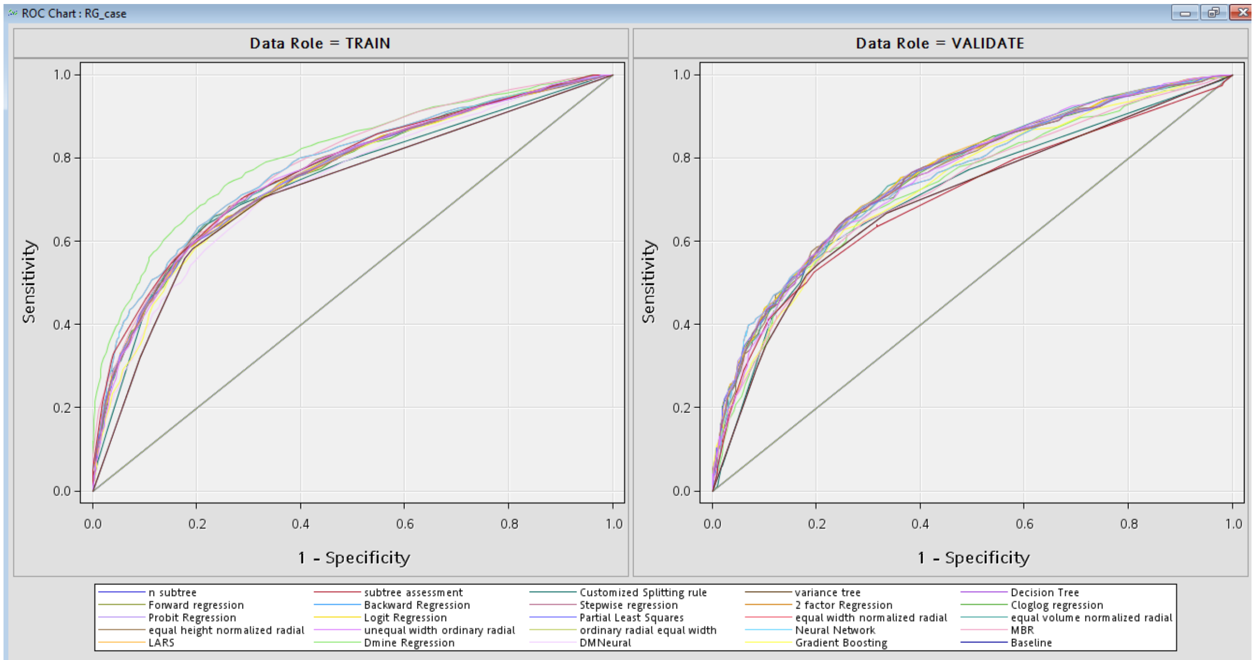
**Table 1. Summary of Validation Misclassification and ASE of Models Compared.**

All models tend to predict well at higher lift percentiles. But the stepwise regression model seems to be doing well across all the percentiles. Given below is a comparison of the best regression, the best neural network, and the best decision tree built for various performance metrics.



**Display 9: Model Performance Metrics**

In each case the stepwise regression model performs the best. From all the models built, it can be observed from Table 1 that many of the top models are all regression models. From display 10 of the ROC curve it can be seen that the stepwise regression model has the highest area under the curve, meaning it is performing better than the rest of the models.

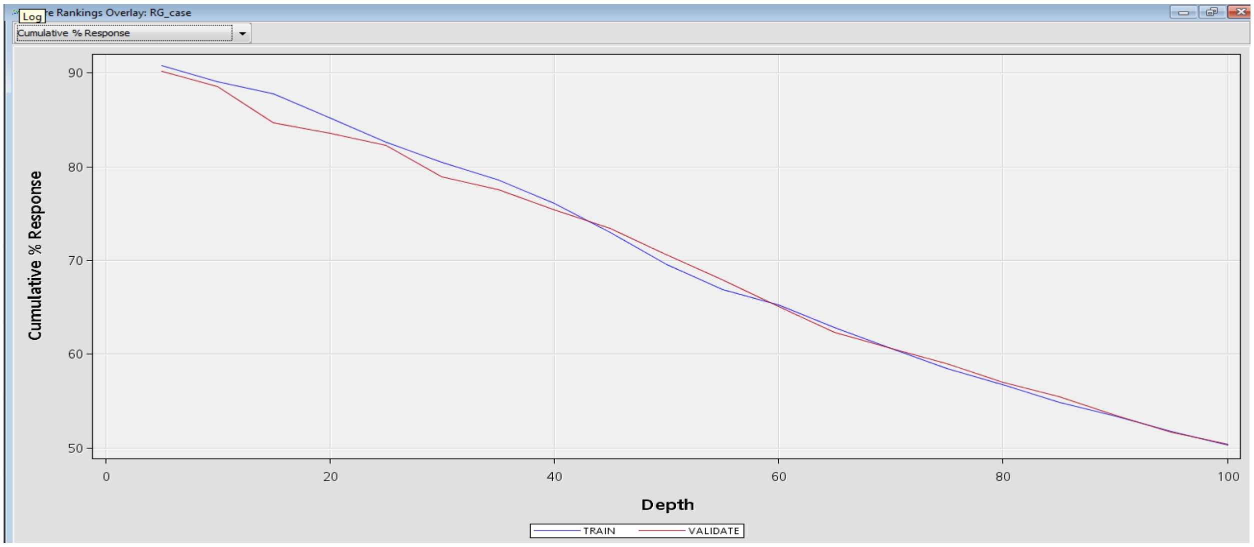


**Display 10. ROC Curve**

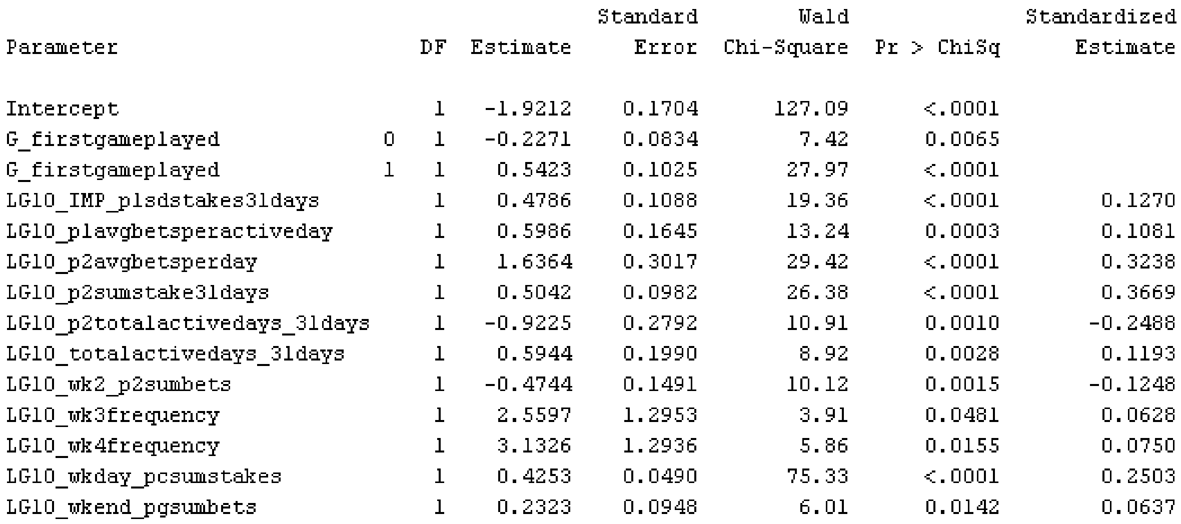
## EXPLANATION OF THE BEST MODEL

The best model identified via stepwise regression consists of the following effects: Intercept, G_firstgameplayed ,LG10_IMP_p1sdstakes31days, LG10_p1avgbetsperactiveday, LG10_p2avgbetsperday, LG10_p2sumstake31days, LG10_p2totalactivedays_31days, LG10_totalactivedays_31days, LG10_wk2_p2sumbets, LG10_wk3frequency, LG10_wk4frequency, LG10_wkday_pcsumstakes and LG10_wkend_pgsumbets.

The most important variables in the stepwise regression based on the odds ratio are LG10_wk3frequency and LG10_wk4frequency. These two variables indicate log transformed values of the number of active days of a subscriber during his 3rd and 4th week online. The cumulative % response chart shown below indicates that if the top twenty percent of subscribers are selected based on the regression model, then 83% of them will be correctly predicted as the high risk internet gamblers.



**Display 11. Percent Captured Response**

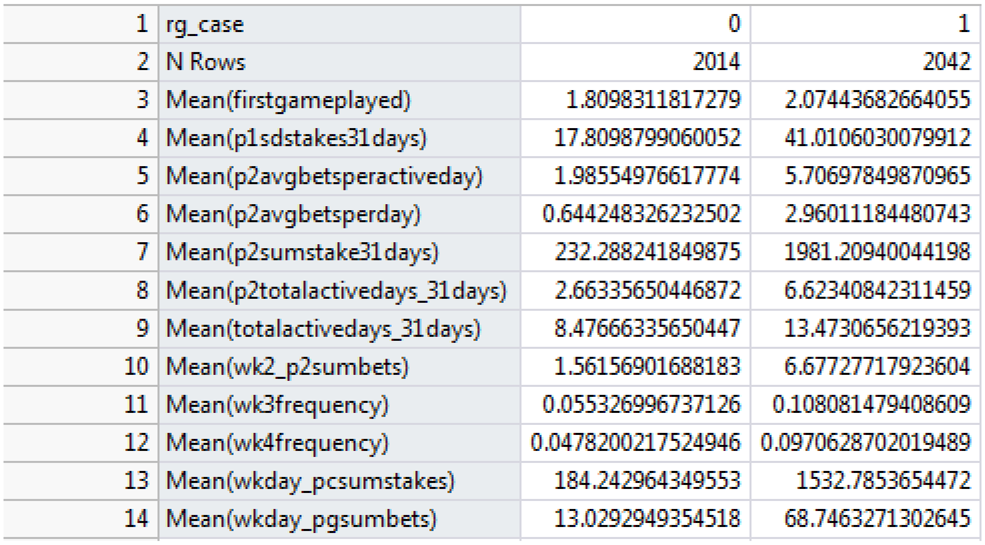


```
                                                    Standard        Wald              Standardized
Parameter                          DF   Estimate       Error  Chi-Square  Pr > ChiSq      Estimate

Intercept                           1    -1.9212      0.1704      127.09      <.0001
G_firstgameplayed              0    1    -0.2271      0.0834        7.42      0.0065
G_firstgameplayed              1    1     0.5423      0.1025       27.97      <.0001
LG10_IMP_p1sdstakes31days           1     0.4786      0.1088       19.36      <.0001        0.1270
LG10_p1avgbetsperactiveday          1     0.5986      0.1645       13.24      0.0003        0.1081
LG10_p2avgbetsperday                1     1.6364      0.3017       29.42      <.0001        0.3238
LG10_p2sumstake31days               1     0.5042      0.0982       26.38      <.0001        0.3669
LG10_p2totalactivedays_31days       1    -0.9225      0.2792       10.91      0.0010       -0.2488
LG10_totalactivedays_31days         1     0.5944      0.1990        8.92      0.0028        0.1193
LG10_wk2_p2sumbets                  1    -0.4744      0.1491       10.12      0.0015       -0.1248
LG10_wk3frequency                   1     2.5597      1.2953        3.91      0.0481        0.0628
LG10_wk4frequency                   1     3.1326      1.2936        5.86      0.0155        0.0750
LG10_wkday_pcsumstakes              1     0.4253      0.0490       75.33      <.0001        0.2503
LG10_wkend_pgsumbets                1     0.2323      0.0948        6.01      0.0142        0.0637
```

**Display 10. Parameter Estimates**

From Display 10 we find that all the variables involved in the model are statistically significant at the traditional 5% level. Total active days in 31 days since the first deposit date for live action which is, the Lg10_p2totalactivedays_31days gives the number of active days during the month, the negative coefficients in these values indicates that some variables actually reduce the chance of a subscriber becoming an gambling addict.



| | | | |
|---|---|---|---|
| 1 | rg_case | 0 | 1 |
| 2 | N Rows | 2014 | 2042 |
| 3 | Mean(firstgameplayed) | 1.8098311817279 | 2.07443682664055 |
| 4 | Mean(p1sdstakes31days) | 17.8098799060052 | 41.0106030079912 |
| 5 | Mean(p2avgbetsperactiveday) | 1.98554976617774 | 5.70697849870965 |
| 6 | Mean(p2avgbetsperday) | 0.644248326232502 | 2.96011184480743 |
| 7 | Mean(p2sumstake31days) | 232.288241849875 | 1981.20940044198 |
| 8 | Mean(p2totalactivedays_31days) | 2.66335650446872 | 6.62340842311459 |
| 9 | Mean(totalactivedays_31days) | 8.47666335650447 | 13.4730656219393 |
| 10 | Mean(wk2_p2sumbets) | 1.56156901688183 | 6.67727717923604 |
| 11 | Mean(wk3frequency) | 0.055326996737126 | 0.108081479408609 |
| 12 | Mean(wk4frequency) | 0.0478200217524946 | 0.0970628702019489 |
| 13 | Mean(wkday_pcsumstakes) | 184.242964349553 | 1532.7853654472 |
| 14 | Mean(wkday_pgsumbets) | 13.0292949354518 | 68.7463271302645 |

**Table 2. Average Values of Gambling Addicts vs Non-Addicts for the Important Variables.**

According to the stepwise model and its important variables, the first game played for a controlled customer is either fixed odds or live action. The variable first game played has five levels. 1= Fixed Odds; 2 = Live Action; 3 = Poker; 4 = Casino Type Games; 5= Other Games. For an addict it is likely to be to live action. The variable ‘P1SDStakes31days’ is the variability of wagers in fixed odds in 31 days since the first deposit date. This is very high for subscribers who are likely to develop a gambling addiction. The values vary from 17 for a controlled subscriber to 41 to not controlled subscribers likely to develop a condition.

The variable ‘p2avgbetsperactiveday’ is the average number of bets per active day of live action. Again this is very high for the subscribers who are likely to turn into gambling addicts compared to the ones who are controlled. Summarizing the patterns in Table 1 indicate that a subscriber who is likely to develop a gambling addiction and become a gambling addict will have a high average bet per day on live action and will put more at stake on live action compared to controlled gamblers. The number of active days playing live action will be considerably more for the likely addict and so will be the total number of active days during the month. As the likely addict reaches the second week he will start putting more on stake on live action bets. And the number of active days during week 3 and week 4 increase in comparison to that of a non-gambling addict. The gambling addict will also have more staked on the casino game online as indicated by the values in Table 1. Apart from these variables, the combination of the first game played and the most frequent game played has significant effect on the target. If the first game played is live action and it has been played at least three times then it is a contributing factor.

## CONCLUSION:

From the stepwise regression and the all the other models built, the following can be considered behavioral markers for subscribers who are likely to develop an online gambling addiction.

- Internet gambling addict's first game is likely to be fixed odds or live action.
- The average time spent by the people online on the website during the first 4 weeks is considerably higher for Internet gambling addicts.
- Majority of this time of Internet gambling addicts is spent betting on live action
- The average wager and money put at stake in every case during the course of the 4 weeks is at least 2 times higher for Internet gambling addicts.
- The total active days spent online is more for Internet gambling addicts

Controlled users and gambling addicts are defined by a specific set of characteristics. The behavior analyzed throughout the data is over a period of 4 weeks on a single website. The purpose of the analysis was to explore whether the initial behavior or the early behavior during the first month will give some indicators which help us pre-label the early behavior markers for gambling addicts. While this data provided some of the desired insights, caution is needed to read too much into these numbers because these are form one particular gambling web site. These indicators may vary across multiple websites because they might offer different assortment of games, terms or conditions. More research is needed to delineate these factors using data from multiple online gambling web sites.

## REFERENCES

- https://support..com/edu/schedules.html?ctry=us&id=76
- http://support..com/publishing/pubcat/chaps/57587.pdf
- http://www..com/technologies/analytics/datamining/miner/neuralnet/index.html
- http://www.ats.ucla.edu/stat//webbooks/reg/
- http://analytics-in-writing.blogspot.com/2012/12/stochastic-gradient-boosting-modeling.html
- http://www.thetransparencyproject.org/codebooks/CodeBook_for%20Braverman_LaPlante_PAB_2013.pdf
- ABC World News with Charles Gibson. Gambling Takes Hold of College Students. March 10, 2006. http://abcnews.go.com/WNT/story?id=1710705&page=1.
- Ashwood RECOVERY. Addicted.com. 2004-2009. http://www.addicted.com/addiction-resources/self-tests/gambling-addiction-quiz
- Chang, M. K. Factor Structure for Young's Internet Addiction Test: A confirmatory Study. 24(6), 2597-2619. 2008.
- Collier, R. (July 15, 2008). Gambling treatment options: A roll of the dice. CMAJ, 179(2), retrieved from http://www.cmaj.ca/cgi/content/full/179/2/127.
- Cronce, J.M., Larimer, M.E., Lostutter, T.W., & Neighbors, C. Exploring College Students Gambling Motivation. 18(4), 361-370. 2002.
- Cunningham-Williams, R.M., Cottler, L.B., Compton III, W.M., & Spitznagel, E.L. (1998). Taking chances: Probl gamblers and mental health disorders- results from the St. Lous epidiologic Catchment Area study. American Journal of Public Health, 88(7), 1093-1096.

- Donaldson-Evans, C. (2006, May 17). Junior jackpot: teen gambling on the rise. Retrieved from http://www.foxnews.com/story/0,2933,195751,00.html.
- Hamilton, M, & Rogers, K. (2008). Internet gambling: community flop or the texas hold'em poker rules. International Review of Law Computers and Technology, 22(3).

## ACKNOWLEDGMENTS

This paper utilized data from the Transparency Project (www.thetransparencyproject.org), Division on Addiction, Cambridge Health Alliance, A teaching affiliate of Harvard Medical School.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sai Vijay Kishore Movva
Oklahoma State University
Stillwater, OK, 74075
Work Phone: 405-612-8898
Email: saivm@okstate.edu

Sai Vijay Kishore Movva is a graduate student in Management Information Systems at Spears School of Business, Oklahoma State University. He works as a Research and a Teaching assistant for the department of Marketing at OSU. Before joining the graduate program he worked as a Service Support Representative and a Content Manager. He has an undergraduate degree in Information Technology. He is a BASE SAS® 9 certified and a certified SAS® predictive modeler using Enterprise Miner 7. He holds the SAS and OSU Data Mining Certificate. He also holds the JMP Data Exploration certificate and has 2 given poster presentations at the SAS Analytics Conference 2013.

Vandana Reddy
Oklahoma State University
Stillwater, OK, 74075
Work Phone: 219-296-9394
Email: vandar@okstate.edu

Vandana Reddy is a Graduate student in Management Information Systems at Spears School of Business Oklahoma State University. Before joining the graduate program she was working as a Bioinformatics research intern and doing her under graduation in Biotechnology. She is a Base SAS® 9 certified professional, a certified SAS ® 7 predictive modeler, JMP Software data exploration certified and holds the SAS and OSU Data Mining certification. She won the SAS Global Forum 2014 student scholarship award and presented one poster 'The Soccer Oracle: Predicting soccer game outcomes using SAS Enterprise Miner during SAS Analytics conference, Orlando 2013 and three posters 'Text Analytics: Predicting the success of newly released free android apps using SAS EM and SAS Sentiment Analysis Studio' ,' The Soccer Oracle: Predicting soccer game outcomes using SAS Enterprise Miner,' 'Using SAS EM to predict the injury risk involved in car accidents in the United States' in SAS Global Forum 2014.

Dr. Goutam Chakraborty
Oklahoma State University
Stillwater, OK, 74078
goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is a professor of marketing and founder of SAS and OSU data mining certificate and SAS and OSU marketing analytics certificate at Oklahoma State University. He has published in many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired M2007 data mining conference. He has over 25 Years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.