

**FACULTY OF COMPUTER AND MATHEMATICAL SCIENCES**

**STA761 STATISTICAL DATA MINING**

**CUSTOMER’S PROFILE PREDICTION FOR PURCHASE OF ORGANIC PRODUCT FROM SUPERMARKET (ORGANIC)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PREPARED BY:** | |  | | --- | | NUR ALIA BINTI MOHAMMAD ZUBIR | | |  | | --- | | 2020947963 | |
|  |  |
| **DATE OF SUBMISSION:** | 30th NOVEMBER 2020 | | |
| **TO BE REVIEWED BY:** | PROF. YAP BEE WAH | | |

Table of Contents

[Stage 1 Business Understanding 1](#_Toc57764805)

[Stage 2 Data Understanding 1](#_Toc57764806)

[Stage 3 Data Preparation 3](#_Toc57764807)

[3.1 DATA FINDING AND ANALYSIS 3](#_Toc57764808)

[Stage 4 Data Modeling 8](#_Toc57764809)

[4.1 MODEL VISUALIZATION 8](#_Toc57764810)

[4.2 DATA MODELING 9](#_Toc57764811)

[4.2.1 Logistic Regression Model (ENTER and STEPWISE) 9](#_Toc57764812)

[4.2.2 Decision Tree Model 12](#_Toc57764813)

[4.2.3 Data Modeling for Neural Network 16](#_Toc57764814)

[4.2.4 Data Modeling for Ensemble 16](#_Toc57764815)

[Stage 5 Model Evaluation 17](#_Toc57764816)

[5.1 Logistic Regression (ENTER and STEPWISE) model comparison 17](#_Toc57764817)

[5.2 Model Comparison for Decision Tree (CHI-SQUARE, GINI and ENTROPY) 19](#_Toc57764818)

[5.3 Model Comparison between LR STEPWISE, DT GINI, Neural Network and Ensemble 21](#_Toc57764824)

[STAGE 6 MODEL DEPLOYMENT 23](#_Toc57764825)

[STAGE 7 CONCLUSION 24](#_Toc57764826)

[APPENDIX A VARIABLE LIST AND DESCRIPTION 25](#_Toc57764827)

[APPENDIX B STATEXPLORE REPORT 26](#_Toc57764828)

[APPENDIX C LOGISTIC REGRESSION ENTER MODEL SUMMARY REPORT 29](#_Toc57764829)

[APPENDIX D LOGISTIC REGRESSION STEPWISE MODEL SUMMARY REPORT 38](#_Toc57764830)

[APPENDIX e LOGISTIC REGRESSION MODEL COMPARISON 52](#_Toc57764831)

[APPENDIX G DECISION TREE CHI-SQUARE MODEL 56](#_Toc57764832)

[APPENDIX M DECISION TREE MODEL COMPARISON 79](#_Toc57764833)

[APPENDIX N NEURAL NETWORK MODEL SUMMARY REPORT 83](#_Toc57764834)

[APPENDIX O ENSEMBLE MODEL SUMMARY REPORT 92](#_Toc57764835)

[APPENDIX P MODEL COMPARISON (STEPWISE-DT GINI-NEURAL NETWORK-ENSEMBLE) 96](#_Toc57764836)

[APPENDIX Q PREDICTION SCORE RESULT 101](#_Toc57764837)

# Stage 1 Business Understanding

The aim is to classify the profiles of clients who are more likely to buy organic goods from the supermarket. The aim of data mining is to construct predictive models to predict clients who are more likely to buy organic goods. Y= 1 is the case of interest (purchase of organic products).

# Stage 2 Data Understanding

The data have 22223 sample. The target variable is ORGYN which is a binary variable with two categories (1=Purchase Organic Products, 0= Not Purchase Organic Products). There are eight (8) input variables. Table 1 provides the description of the variables.

Table 1 Description of Variables

| **VARIABLE** | **ROLE** | **MEASUREMENT TYPE** | **DESCRIPTION** |
| --- | --- | --- | --- |
| ORGYN | Target | Binary | Organics Purchased?  *1 - Purchase Organic Product*  *0 - Not Purchase Organic Product* |
| AFFL | Input | Interval | Affluence Grade |
| AGE | Input | Interval | Age |
| BILL | Input | Interval | Total Amount Spent |
| LTIME | Input | Interval | Years as Loyalty Card Member |
| CLASS | Input | Nominal | Customer Loyalty Status  *Gold*  *Platinum*  *Silver*  *Tin* |
| GENDER | Input | Nominal | Gender  *U - Unknown*  *F - Female*  *M - Male* |
| NGROUP | Input | Nominal | Neighborhood Group  *A*  *B*  *C*  *D*  *E*  *F*  *U* |
| TV\_REG | Input | Nominal | TV Region  *East*  *London*  *Border*  *Ulster*  *N Scot*  *N East*  *N West*  *S West*  *Wales & West*  *S & S East*  *C Scotland*  *Yorkshire*  *Midlands* |

There are several variables that need to reject based on the following reasons:

* Redundant data
* The variables AGE, AGEGRP1 and AGEGRP2 are all different measurement for the same information.
* LCDATE and LTIME essentially measure the same thing. Presume that LTIME is sufficient for building a predictive model.
* NGROUP contains collapsed levels of NEIGHBORHOOD. Presume that NGROUP is sufficient for building a predictive model.
* TV\_REG and REGION essentially measure the same thing. Presume that TV\_REG is sufficient for building a predictive model.
* ORGANICS was rejected because it contains information that would not be known at the time that developing a model to predict the purchase of organic products.
* The variables contain high percentage of missing value.

# Stage 3 Data Preparation

## 3.1 DATA FINDING AND ANALYSIS

The data was stored as a SAS data file. The sample consists of 22,223 cases. Table 2 illustrate the pre-analysis for categorical variables.

Table 2 Data Audit Summary, Frequency Distribution and Interpretation for Categorical Variables

| **VARIABLES** | **FREQUENCY** |
| --- | --- |
| **ORGYN** |  |
| **CLASS** |  |
| **GENDER** |  |
| **NGROUP** |  |
| **TV\_REG** |  |

Based on table 2, ORGYN variable have 75.2% of the sample population that will not purchase organic product. Only 25% will purchase organic product. This information indicates imbalance sample population in target variable. CLASS variable, it is clear that Silver have the most sample population with a percentage of 38.6% follow by Tin (29.2%), Gold (28.5%) and platinum which has the lowest sample population with only 3.8%. There is no missing value observed in this variable. Sample population of male 5829 (26.2%) and female sample population of 12149 (54.7%). Gender sample shows presence of unknown gender by 19% equivalent to a total of 4245. In NGROUP, group C (being the highest by 20.5%) having balance sample with group D, F and B. However, for E the sample population is lower at 11.7% followed by A (at 8.3%). Group U found to be the lowest (by 0.2%) sample population. Presence of missing values (total 674) equivalent to 3% from the total sample population. TV\_REG indicating a varieties of percentage of sample population across the area in the United Kingdom. High sample population reside in London category with leading percentage of 27.8% and Ulster (with 1.2%), N Scot (with 1.5%) showing lower sample population. There is 465 missing values observed in TV\_REG variable.

Summary statistics are represented in Table 3 for continuous variables, while Table 4 further explains the analysis for continuous variables using boxplot visualization.

*Table 3 Summary Statistics*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **Standard Deviation** | **Minimum** | **Median** | **Maximum** | **Skewness** | **Kurtosis** |
| BILL | 4420.59 | 7559.048 | 0.01 | 2000 | 296313.9 | 8.037186 | 184.8715 |
| AFFL | 8.711893 | 3.421125 | 0 | 8 | 34 | 0.891684 | 2.09686 |
| LTIME | 6.56467 | 4.657113 | 0 | 5 | 39 | 2.28279 | 8.077622 |
| AGE | 53.79715 | 13.20605 | 18 | 54 | 79 | -0.07983 | -0.84389 |

Table 4 Analysis and Visualization for Continuous Variables

| **VARIABLES** | **EXPLANATIONs** |
| --- | --- |
|  | Most data are dense, closer to the minimum value, backed by a skewness value of 8,037, which means that the information is skewed. In BILL variables above limit, outliers with some suggest that extreme cases were observed (Q3+1.5\*IQR where IQR is Interquartile Range). The Kurtosis value of 184.87 indicates an exceptionally large sample population over the overall sample population range that resides closer to the lower zone. |
|  | Sample population with skewness value of 0.892, signify that the data is skewed right. Outliers was also observed in AFFL variables above maximum (Q3+1.5\*IQR where IQR is Interquartile Range). There is a different in data dipersion |
|  | In LTIME variable, distributions of sample population are imbalance where more sample reside in 3rd quartile. Significant numbers of outliers observed in LTIME above maximum (Q3+15\*IQR where IQR is Interquartile Range). Missing value indicates 281 missing values reside in LTIME variable. |
|  | Boxplot in AGE variable is visible (whiskers is visible). Data almost equally distributed with slightly skewed left as supported by skewness value of -0.079. Presence of missing values were observed as indicated by histogram and outliers are presence below minimum (Q1+1.5\*IQR). Data audits indicate a total of 1508 missing values found in AGE variable. |

In order to test the relation between ORGYN and the input variables, the Chi-Square test was used. The Chi-Square outcome of the chi-square statistical test is tabulated according to Table 5.

Table 4 Chi-Square Statistic Test Results

|  |  |  |  |
| --- | --- | --- | --- |
| Input | Chi-Square | Df | Prob |
| AFFL | 2252.5333 | 5 | <.0001 |
| AGE | 2842.0746 | 5 | <.0001 |
| CLASS | 281.1281 | 3 | <.0001 |
| GENDER | 1495.259 | 2 | <.0001 |
| NGROUP | 97.7389 | 7 | <.0001 |
| LTIME | 79.673 | 5 | <.0001 |
| TV\_REG | 23.6238 | 13 | 0.0348 |
| BILL | 2.9896 | 4 | 0.5596 |

Based on the results, it indicates that AFFL (Chi-Square (5df) = 2252.5333, p<0.05), AGE (Chi-Square (5df) = 2842.0746, p<0.05), CLASS (Chi-Square (3df) = 281.1281, p<0.05), GENDER (Chi-Square (2df) = 1495.2590, p<0.05), , NGROUP (Chi-Square (7df) = 97.7389, p<0.05) and LTIME (Chi-Square (2df) = 1495.2590, p<0.05) have significant association with ORGYN. There is no significant association between ORGYN with TV\_REG since the value of Chi-Square with 13df is 0.0348 and p>0.05 and BILL (Chi-Square (4df) = 0.5596, p>0.05).

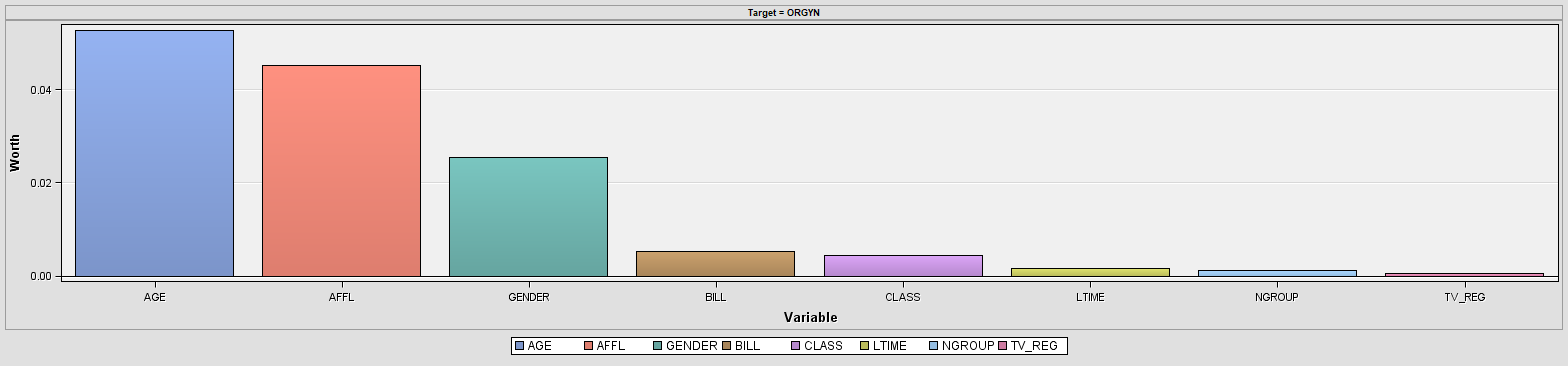


Figure 1 Plot of Variable Worth

From the visualization, AGE is the most important independent variable, followed by AFFL, GENDER and BILL.

Based on the samples in the datasets whereby there were missing values, outliers and imbalanced data for target variable, model will include conditioning the datasets, by either to

* consider complete cases only; or
* removal of outliers; or
* perform balanced data; or
* discard missing value.

For data cleaning process, IBM SPSS modeler and SPSS statistics were used to manage the process. As a result, to the data preparation, 8540 rows of data will be used for prediction model. Based on management of outliers through binning processes, it has resulted to original variables been replace with the new variables such as BILL\_TILE4 and AFFL\_TILE4. A balance node was used to reduce the imbalance target data, which is ORGYN. A cleaned data was then fed into SAS Enterprise Miner for model deployment.

# Stage 4 Data Modeling

## 4.1 MODEL VISUALIZATION

Predictive modeling using logistic regression (ENTER and STEPWISE), decision tree model (CHI-SQUARE, GINI and ENTROPY), ARTIFICIAL NEURAL NETWORK and ENSEMBLE was carried out by using SAS Enterprise Miner. Figure 3 shows the predictive modeling flow. The data source was connected to StatExplore node (visualize plot of worth and chi-qsuare), Graph Explore node (visualize descriptive statistic) and Data Partition node. The data was partitioned using 70% as training sample and 30% as validation (testing) sample.

The Data Partition node was related to the Regression nodes (ENTER and STEPWISE), Decision Tree nodes (CHI-SQUARE, ENTROPY AND GINI) and the Neural Network node. The results of the two logistic regression models, three decision tree models and the one of the neural network were then evaluated and compared using the Model Comparison node.

Together with the neural network model, the best output model from logistic regression models and decision tree models is then linked to the Control Point node that is linked to the Ensemble node bypassing the Model Comparison node. This is a model of the ensemble. The best model chosen will be linked to the Score node for deployment.

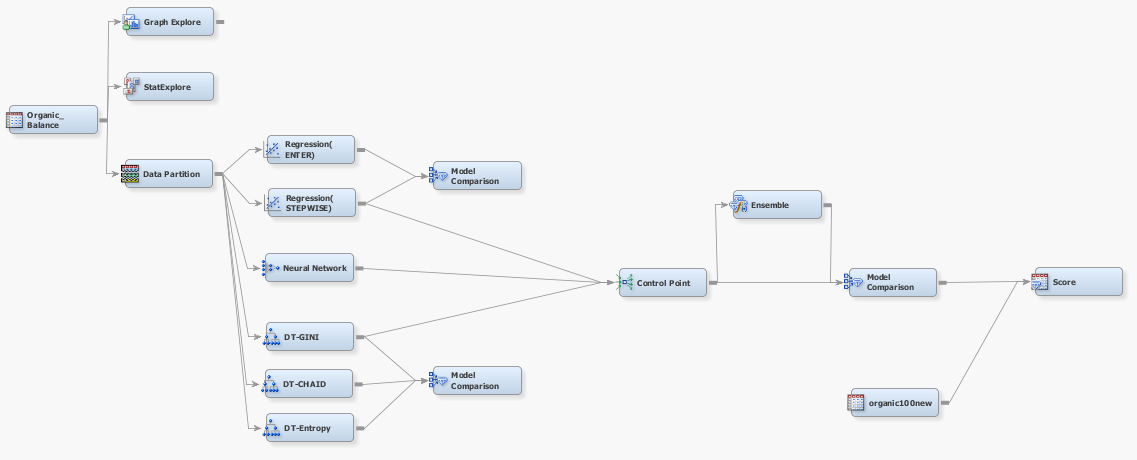


Figure 2 Predictive Modeling Flow using SAS Miner

## 4.2 DATA MODELING

### 4.2.1 Logistic Regression Model (ENTER and STEPWISE)

Logistic Regression model was developed using ENTER and STEPWISE method. Table 6 summarizes the model results.

Table 6 Estimated Logistic Regression Coefficients

|  |  |  |
| --- | --- | --- |
| Parameter | ENTER | STEPWISE |
| Intercept | 2.0873 | 2.0904 |
| AFFL\_TILE4 1 | -1.0451\*\* | -1.0375\*\* |
| AFFL\_TILE4 2 | -0.425\*\* | -0.4229\*\* |
| AFFL\_TILE4 3 | 0.2312\*\* | 0.2236\*\* |
| AGE | -0.0494\*\* | -0.0502\*\* |
| BILL\_TILE4 1 | -0.0724 |  |
| BILL\_TILE4 2 | -0.0855 |  |
| CLASS Gold | -0.1035 |  |
| CLASS Platinum | -0.135 |  |
| CLASS Silver | 0.1285 |  |
| GENDER F | 0.4671\*\* | 0.4643\*\* |
| LTIME\_TILE4 1 | 0.037 |  |
| LTIME\_TILE4 2 | -0.0891 |  |
| LTIME\_TILE4 3 | 0.0408 |  |
| NGROUP A | -0.0964 |  |
| NGROUP B | -0.1748 |  |
| NGROUP C | -0.0948 |  |
| NGROUP D | -0.0542 |  |
| NGROUP E | 0.086 |  |
| NGROUP F | 0.00442 |  |
| TV\_REG Border | 0.1926 |  |
| TV\_REG C Scotland | -0.0168 |  |
| TV\_REG East | -0.0567 |  |
| TV\_REG London | 0.0114 |  |
| TV\_REG Midlands | 0.0152 |  |
| TV\_REG N East | 0.0978 |  |
| TV\_REG N Scot | 0.2675 |  |
| TV\_REG N West | -0.1994\*\* |  |
| TV\_REG S & S East | -0.1695 |  |
| TV\_REG S West | 0.0826 |  |
| TV\_REG Wales & West | 0.046 |  |

\*p<0.05; \*\*p<0.01

**Logistic Regression Enter Model Summary**

Logistic Regression using ENTER selection method shows that the **model is significant** since the value of Chi-Square (30df) = 1654.8738, p<0.05. The predictive accuracy for testing of this model is 70.22%, sensitivity (71.59%), specificity (68.88%) and precision (69.30%). Results show that AFFL\_TILE4 (Wald=628.201 (DF=3), p<0.05), AGE (Wald=367.0642 (DF=1), p<0.05) and GENDER (Wald=191.8521 (DF=1), p<0.05) are significant predictors of ORGYN.

Table 7 Chi-Square Test Result for Logistic Regression Enter

|  |  |  |  |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| AFFL\_TILE4 | 3 | 628.201 | <.0001 |
| AGE | 1 | 367.0642 | <.0001 |
| BILL\_TILE4 | 2 | 0.7098 | 0.7012 |
| CLASS | 3 | 0.6974 | 0.8738 |
| GENDER | 1 | 191.8521 | <.0001 |
| LTIME\_TILE4 | 3 | 1.9123 | 0.5908 |
| NGROUP | 6 | 6.9517 | 0.3253 |
| TV\_REG | 11 | 14.2666 | 0.2186 |

**Logistic Regression Stepwise Model Summary**

Logistic Regression using STEPWISE selection method shows that the model is significant (Chi-Square (4df) = 1631.3033, p<0.05). The predictive accuracy for testing of this model is 70.68%, sensitivity (72.06%), specificity (69.34%) and precision (69.76%). Results show that AFFL\_TILE4 (Wald=629.7210 (DF=3), p<0.05), AGE (Wald=487.6217 (DF=1), p<0.05) and GENDER (Wald=191.1079 (DF=1), p<0.05) are significant predictors of ORGYN.

Table 8 Chi-Square Test Result for Logistic Regression Stepwise

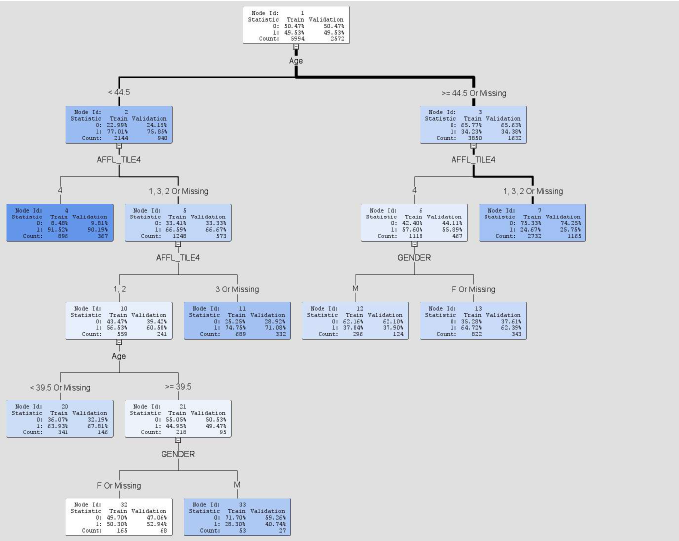
|  |  |  |  |
| --- | --- | --- | --- |
| **Effect** | **DF** | **Wald Chi-Square** | **Pr > ChiSq** |
| AFFL\_TILE4 | 3 | 629.7210 | <.0001 |
| AGE | 1 | 487.6217 | <.0001 |
| GENDER | 1 | 191.1079 | <.0001 |

Table 9 Odds Ratio Interpretation for Stepwise Model

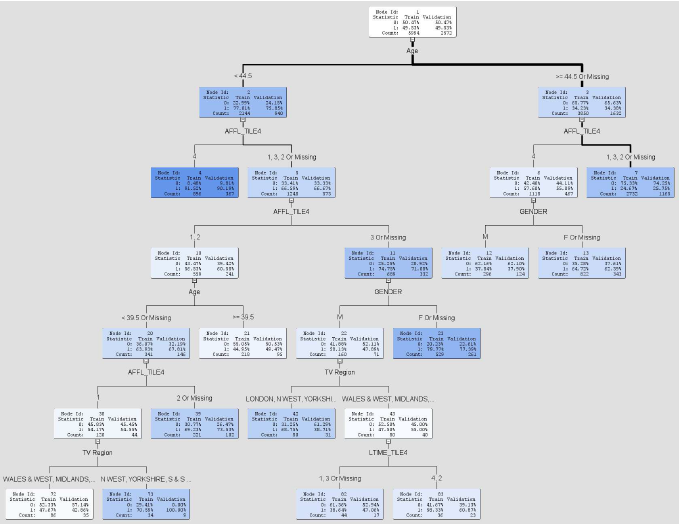
|  |  |  |  |
| --- | --- | --- | --- |
| **VARIABLES** | **SUB-VARIABLE** | **ODDS-RATIO (OR)** | **INTERPRETATION** |
| AFFL\_TILE4 | 1 vs. 4 | 0.103 | The odds-ratio for Category 1 AFFL is less than 1 (OR=0.103). This indicates that this AFFL category are less likely to buy organic products compared to Category 4 AFFL. |
| AFFL\_TILE4 | 2 vs. 4 | 0.190 | The odds ratio for Category 2 AFFL is less than 1 (OR=0.190). This indicates that this Category AFFL 2 are less likely to buy organic products compared to Category 4 AFFL. |
| AFFL\_TILE4 | 3 vs. 4 | 0.363 | The odds ratio (3 vs. 4) is 0.363. This indicates that buyers of Category 3 AFFL are less likely to purchase the organic products compare to Category 4 AFFL. |
| AGE |  | 0.951 | The odds ratio for AGE (0.951) indicates that for every increase in age, the odds ratio of purchased products decrease by 4.9% |
| GENDER | Female vs. Male | 2.531 | The odds ratio FEMALE vs. MALE is 2.531. This indicates that female buyers are three times more likely to buy organic products compared to MALE. |

### 4.2.2 Decision Tree Model

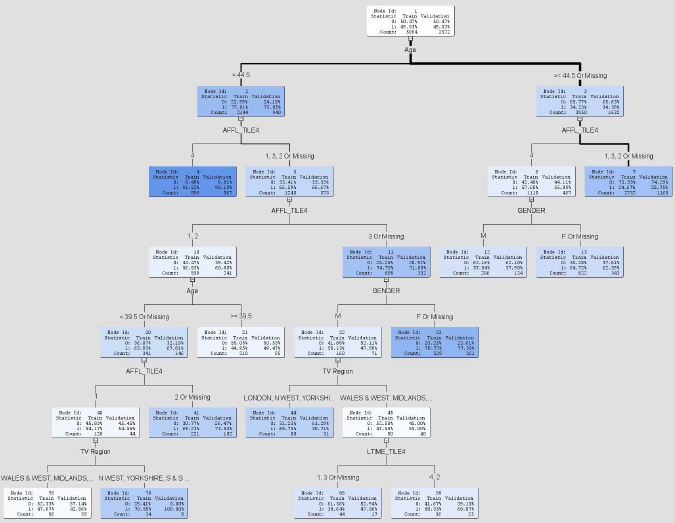
The decision tree models using CHI-SQUARE, ENTROPY and GINI splitting criteria are shown in Figure 3 to Figure 5 along with its decision rules and interpretation of it. Decision Tree using CHI-SQUARE, ENTROPY and GINI as splitting criteria produced 8, 12 and 12 decision rules, respectively.



*Figure 3 Decision tree using CHI-SQUARE splitting criteria.*



*Figure 4 Decision tree using ENTROPY splitting criteria.*



*Figure 5 Decision tree using GINI splitting criteria.*

**Some of Decision Tree GINI interpretations:**

1. If AGE less or equal to 44.5, and AFFL\_TILE4 is 4, then ORGYN is 1 (organic purchased).
2. If AGE more or equal to 44.5 or MISSING, and AFFL\_TILE4 is one of (1, 2, 3 or MISSING), then ORGYN = 0 (not purchase organic product).
3. If GENDER is F or MISSING, and AGE less than 44.5, and AFFL\_TILE4 is one of (1, 2, 3, or MISSING), then ORGYN = 0 (not purchase organic product).
4. If GENDER is M, and AGE more or equal to 44.5 or MISSING, and AFFL\_TILE4 is 4, then ORGYN = 0 (not purchase organic product).
5. If GENDER is F or MISSING, AND Age less than 44.5 AND Age >= 39.5, AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING then ORGYN = 0 (customer will not purchase organic product).

The details of Decision Tree results comprise of the decision rules are available in the following Appendices:

* Appendix G - for Decision Tree CHI-SQUARE
* Appendix I - for Decision Tree GINI
* Appendix K - for Decision Tree ENTROPY

### 4.2.3 Data Modeling for Neural Network

The selected neural network is MLP. The accuracy for validation in this model is 71.66% with sensitivity, specificity, and precision of 70.33%, 72.96% and 71.85%. The neural network model will be compared with other models to evaluate the model performance and to finalize the selection of model for deployment.

### 4.2.4 Data Modeling for Ensemble

Three (3) methods were used for the Ensemble model, which is the STEPWISE logistic regression model, the GINI decision tree model and the Neural Network model. The accuracy measured on the basis of the event classification table shows that accuracy is 72.82 percent, 69.07 percent sensitivity, 76.50 percent precision and 74.26 percent accuracy. To assess the efficiency of the experiment and to finalize the model selection for implementation, the Ensemble model will be compared with other models.

# Stage 5 Model Evaluation

## 5.1 Logistic Regression (ENTER and STEPWISE) model comparison

Table 10 displays the model evaluation results. The results of the two models are very comparable. The validation accuracy of the STEPWISE model is 70.68 %, sensitivity (72.06%), specificity (69.34%) and precision (69.76%). As they have selected the same predictor variables, the models have similar classification results.

Table 10 Model Performance (Logistic Regression)

|  |  |  |  |
| --- | --- | --- | --- |
| **MODEL** | **PERFORMANCE MEASURE** | **TRAINING (%)** | **VALIDATION (%)** |
| ENTER | Accuracy | 71.10 | 70.22 |
| Sensitivity | 70.33 | 71.59 |
| Specificity | 71.87 | 68.88 |
| Precision | 71.04 | 69.30 |
| STEPWISE | Accuracy | 70.90 | 70.68 |
| Sensitivity | 70.50 | 72.06 |
| Specificity | 71.31 | 69.34 |
| Precision | 70.69 | 69.76 |

The ROC charts in Figure 6 show that for the validation sample, both curve for STEPWISE model and ENTER model is almost identical. Area Under Curve for both Logistic ENTER and STEPWISE are like each other. This is also because both models choose the same predictor variables.

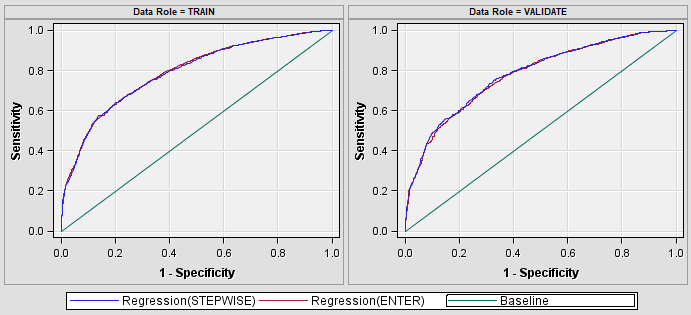


Figure 6 ROC Charts for Logistic Regression Models

Comparing the logistic regression of the ENTER method with the logistic regression of the STEPWISE method, the output of the model using the STEPWISE method was marginally higher than the logistic regression of the ENTER model, as the model is slightly capable of correctly predicting more samples.

The selection of the best model is based on the results of the validation sample and the performance evaluation, as shown in the ROC Table. For this project, we therefore propose that the STEPWISE logistics regression model be further compared with the decision tree, neural network, and ensemble model.

## 5.2 Model Comparison for Decision Tree (CHI-SQUARE, GINI and ENTROPY)

Table 11 displays the model evaluation results. The results of the three models are very comparable. The GINI and ENTROPY decision tree model has the highest validation accuracy, specificity and precision, while the CHI-SQUARE decision tree has the highest sensitivity (71.90%).

Table 5 Model Comparison Results (Decision Tree)

|  |  |  |  |
| --- | --- | --- | --- |
| **MODEL** | **PERFORMANCE MEASURE** | **TRAINING (%)** | **VALIDATION (%)** |
| CHI-SQUARE | Accuracy | 74.21 | 72.86 |
| Sensitivity | 73.02 | 71.90 |
| Specificity | 75.37 | 73.81 |
| Precision | 74.42 | 72.93 |
| GINI | Accuracy | 74.42 | 72.94 |
| Sensitivity | 68.27 | 67.27 |
| Specificity | 80.46 | 78.51 |
| Precision | 77.43 | 75.44 |
| ENTROPY | Accuracy | 74.42 | 72.94 |
| Sensitivity | 68.27 | 67.27 |
| Specificity | 80.46 | 78.51 |
| Precision | 77.43 | 75.44 |

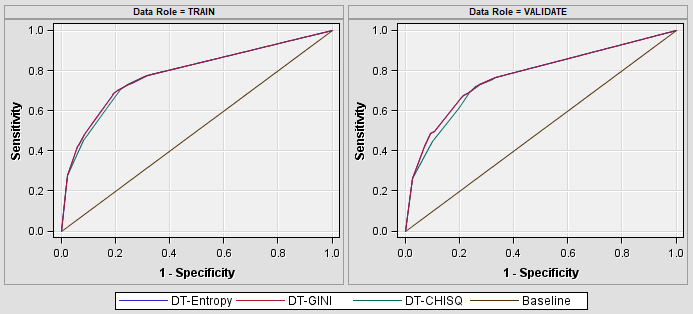


Figure 1 ROC Charts for Decision Tree Models

The ROC charts in Figure 5 shows that the decision tree GINI can predict more samples correctly compared to both decision tree CHI-SQUARE and ENTROPHY. Area Under the Curve (AUC) of decision tree GINI is larger than AUC of both decision tree CHI-SQUARE and ENTROPY. Selection of the best model is based on the validation sample results as well as evaluating the performance as depicted by ROC Chart. Thus, for this project decision tree GINI is recommended to be further compared with logistic regression STEPWISE, neural network and ensemble model.



## Model Comparison between LR STEPWISE, DT GINI, Neural Network and Ensemble

Based on the model comparison result from both logistic regression and decision tree models, logistic regression STEPWISE and decision tree of GINI model was further compared with NEURAL NETWORK and ENSEMBLE model which combined the stated models.

The results are tabulated as per Table 12.

Table 6 Model Comparison Results (STEPWISE, GINI, Neural Network and Ensemble)

|  |  |  |  |
| --- | --- | --- | --- |
| **MODEL** | **PERFORMANCE MEASURE** | **TRAINING (%)** | **VALIDATION (%)** |
| STEPWISE | Accuracy | 70.90 | 70.68 |
| Sensitivity | 70.50 | 72.06 |
| Specificity | 71.31 | 69.34 |
| Precision | 70.69 | 69.76 |
| GINI | Accuracy | 74.42 | 72.94 |
| Sensitivity | 68.27 | 67.27 |
| Specificity | 80.46 | 78.51 |
| Precision | 77.43 | 75.44 |
| Neural Network | Accuracy | 74.64 | 71.66 |
| Sensitivity | 72.41 | 70.33 |
| Specificity | 76.83 | 72.96 |
| Precision | 75.41 | 71.85 |
| Ensemble Model | Accuracy | 74.44 | 72.82 |
| Sensitivity | 70.06 | 69.07 |
| Specificity | 78.74 | 76.50 |
| Precision | 76.39 | 74.26 |

As shown in Table 12, all models performed relative the same from one to another. However, the accuracy is slightly higher in GINI with 72.94%. The next higher model performance is Ensemble (with 72.82%), Neural Network (with 70.33%) and last is Logistic Regression STEPWISE model with accuracy of 70.68%. Therefore, Decision Tree GINI is selected to next stage of model deployment based on its performance.

Table 7 Model Reporting Comparison for DT GINI, Ensemble, Neural Network and LR STEPWISE

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Selected Model | Model Description | Selection Criterion: Valid: Misclassification Rate | Train: Misclassification Rate | Valid: Misclassification Rate | Train: Roc Index | Train: Gini Coefficient | Train: Kolmogorov-Smirnov Statistic | Valid: Roc Index | Valid: Gini Coefficient |
| Y | DT-GINI | 0.270606532 | 0.255755756 | 0.270606532 | 0.785 | 0.57 | 0.487 | 0.771 | 0.543 |
|  | Ensemble | 0.271772939 | 0.255588922 | 0.271772939 | 0.814 | 0.627 | 0.489 | 0.796 | 0.593 |
|  | Neural Network | 0.283437014 | 0.25358692 | 0.283437014 | 0.818 | 0.635 | 0.496 | 0.791 | 0.581 |
|  | Regression (STEPWISE) | 0.293157076 | 0.290957624 | 0.293157076 | 0.787 | 0.574 | 0.436 | 0.779 | 0.558 |

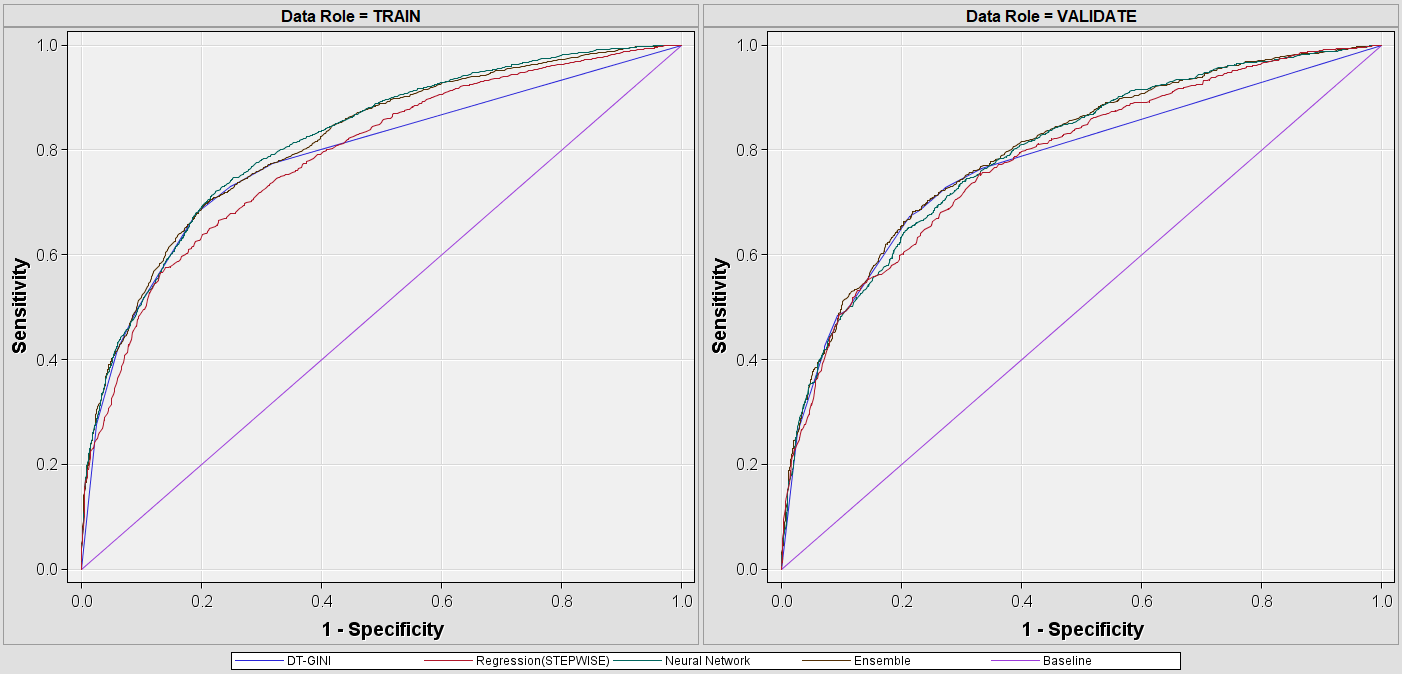


Figure 2 ROC Charts for DT GINI, Ensemble, Neural Network and LR STEPWISE Models

# STAGE 6 MODEL DEPLOYMENT

The Decision Tree GINI model was deployed to the new datasets which consist of 100 cases. A new dataset comprises of 100 samples was used for model deployment. Findings from new datasets is not a complete and clean dataset. Missing values was presence in variable:

* GENDER - total 22
* TV\_REG - total 2
* AFFL - 2
* LTIME - 1

Outliers was also found in BILL (2) and LTIME (2).

As previously, with new variables added, the original datasets organics were prepared, the same applies to this new organic dataset where it must be conditioned and matched with the variables used in the model. Missing values from the datasets have been deleted. The result is shown in Figure 7.



Figure 3 Sample of the Predicted Organic Product Purchase Probabilities for New Customers

Based on the results as shown in Figure 7, 10 customers out of 100 were predicted as more likely to purchase organic products from the Supermarket.

# STAGE 7 CONCLUSION

Based on the result obtain from the analysis, it can be concluded as follows:

* For better model development, the right data and completeness of data must be taken into consideration.
* Model performance – overall performance is considered as quite good. However, with better data quality, the model shall be retrained to achieve better model performance.

# APPENDIX A VARIABLE LIST AND DESCRIPTION

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Variable Name** | **Description** | **Measurement Type** |
|  | PrimaryLast | Indicator of each last matching case as Primary | Ordinal |
|  | CUSTID | Customer ID | Nominal |
|  | GENDER | Gender | Nominal |
|  | AGEGRP1 | Age Group 1 | Nominal |
|  | AGEGRP2 | Age Group 2 | Nominal |
|  | TV\_REG | TV Region | Nominal |
|  | NGROUP | Neighborhood Group | Nominal |
|  | NEIGHBORHOOD | Type of Residential Neighborhood | Nominal |
|  | REGION | Geographic Region | Nominal |
|  | CLASS | Customer Loyalty Status | Nominal |
|  | ORGYN | Organic Purchased | Nominal |
|  | DOB | Date of Birth | Continuous |
|  | EDATE | Date data taken from data base | Continuous |
|  | AGE | Age | Continuous |
|  | LCDATE | Loyalty card Application Date | Continuous |
|  | ORGANICS | Number of Organic Products Purchased | Continuous |
|  | BILL | Total Amount Spent | Continuous |
|  | AFFL | Affluence Grade | Continuous |
|  | LTIME | Years as Loyalty Card Member | Continuous |

# APPENDIX B STATEXPLORE REPORT

Variable Summary

Measurement Frequency

Role Level Count

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Variable Levels Summary

(maximum 500 observations printed)

Frequency

Variable Role Count

ORGYN TARGET 2

Class Variable Summary Statistics

(maximum 500 observations printed)

Data Role=TRAIN

Number

Data Variable of Mode Mode2

Role Name Role Levels Missing Mode Percentage Mode2 Percentage

TRAIN AFFL\_TILE4 INPUT 4 0 3 33.80 4 33.25

TRAIN BILL\_TILE4 INPUT 3 0 1 51.32 3 29.09

TRAIN CLASS INPUT 4 0 Silver 38.72 Tin 31.09

TRAIN GENDER INPUT 2 0 F 71.85 M 28.15

TRAIN LTIME\_TILE4 INPUT 4 0 4 34.61 3 32.31

TRAIN NGROUP INPUT 7 0 C 21.67 D 20.27

TRAIN TV\_REG INPUT 12 0 London 29.47 Midlands 14.83

TRAIN ORGYN TARGET 2 0 0 50.47 1 49.53

Distribution of Class Target and Segment Variables

(maximum 500 observations printed)

Data Role=TRAIN

Data Variable Frequency

Role Name Role Level Count Percent

TRAIN ORGYN TARGET 0 4323 50.4670

TRAIN ORGYN TARGET 1 4243 49.5330

Interval Variable Summary Statistics

(maximum 500 observations printed)

Data Role=TRAIN

Standard Non

Variable Role Mean Deviation Missing Missing Minimum Median Maximum Skewness Kurtosis

AGE INPUT 51.54168 13.86091 8566 0 19 51 79 0.10436 -0.99771

Class Variable Summary Statistics by Class Target

(maximum 500 observations printed)

Data Role=TRAIN Variable Name=AFFL\_TILE4

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 4 0 3 34.95 2 26.74

ORGYN 1 4 0 4 48.46 3 32.62

\_OVERALL\_ 4 0 3 33.80 4 33.25

Data Role=TRAIN Variable Name=BILL\_TILE4

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 3 0 1 44.32 3 34.17

ORGYN 1 3 0 1 58.45 3 23.92

\_OVERALL\_ 3 0 1 51.32 3 29.09

Data Role=TRAIN Variable Name=CLASS

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 4 0 Silver 38.75 Gold 31.04

ORGYN 1 4 0 Silver 38.70 Tin 36.51

\_OVERALL\_ 4 0 Silver 38.72 Tin 31.09

Data Role=TRAIN Variable Name=GENDER

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 2 0 F 62.73 M 37.27

ORGYN 1 2 0 F 81.15 M 18.85

\_OVERALL\_ 2 0 F 71.85 M 28.15

Data Role=TRAIN Variable Name=LTIME\_TILE4

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 4 0 4 36.73 3 30.21

ORGYN 1 4 0 3 34.46 4 32.45

\_OVERALL\_ 4 0 4 34.61 3 32.31

Data Role=TRAIN Variable Name=NGROUP

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 7 0 C 22.07 B 20.24

ORGYN 1 7 0 F 21.80 C 21.26

\_OVERALL\_ 7 0 C 21.67 D 20.27

Data Role=TRAIN Variable Name=TV\_REG

Number

Target of Mode Mode2

Target Level Levels Missing Mode Percentage Mode2 Percentage

ORGYN 0 12 0 London 28.15 Midlands 14.67

ORGYN 1 12 0 London 30.80 Midlands 14.99

\_OVERALL\_ 12 0 London 29.47 Midlands 14.83

Interval Variable Summary Statistics by Class Target

(maximum 500 observations printed)

Data Role=TRAIN Variable=AGE

Target Non Standard

Target Level Median Missing Missing Minimum Maximum Mean Deviation Skewness Kurtosis Role Label

ORGYN 0 56 0 4323 19 79 56.12607 12.13403 -0.16452 -0.61851 INPUT Age

ORGYN 1 43 0 4243 20 79 46.87085 13.95231 0.563107 -0.74653 INPUT Age

\_OVERALL\_ 51 0 8566 19 79 51.54168 13.86091 0.10436 -0.99771 INPUT Age

Chi-Square Statistics

(maximum 500 observations printed)

Data Role=TRAIN Target=ORGYN

Input Chi-Square Df Prob

AGE 1331.3286 4 <.0001

AFFL\_TILE4 1127.8206 3 <.0001

GENDER 358.9016 1 <.0001

BILL\_TILE4 177.0205 2 <.0001

CLASS 167.0832 3 <.0001

NGROUP 51.3874 6 <.0001

LTIME\_TILE4 26.9580 3 <.0001

TV\_REG 18.7447 11 0.0658

# APPENDIX C LOGISTIC REGRESSION ENTER MODEL SUMMARY REPORT

Variable Summary

Measurement Frequency

Role Level Count

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

The DMREG Procedure

Model Information

Training Data Set WORK.EM\_DMREG.VIEW

DMDB Catalog WORK.REG3\_DMDB

Target Variable ORGYN (Organics Purchased?)

Target Measurement Level Ordinal

Number of Target Categories 2

Error MBernoulli

Link Function Logit

Number of Model Parameters 31

Number of Observations 5994

Target Profile

Ordered Total

Value ORGYN Frequency

1 1 2969

2 0 3025

The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates 31

Optimization Start

Active Constraints 0 Objective Function 4154.4626015

Max Abs Gradient Element 687.82982983

Ratio

Between

Actual

Objective Max Abs and

Function Active Objective Function Gradient Predicted

Iter Restarts Calls Constraints Function Change Element Ridge Change

1 0 2 0 3355 799.0 97.5668 0 1.080

2 0 3 0 3327 28.1542 9.3433 0 1.064

3 0 4 0 3327 0.2960 0.1188 0 1.008

4 0 5 0 3327 0.000048 0.000020 0 1.000

Optimization Results

Iterations 4 Function Calls 7

Hessian Calls 5 Active Constraints 0

Objective Function 3327.025713 Max Abs Gradient Element 0.0000204522

Ridge 0 Actual Over Pred Change 1.0001134999

Convergence criterion (GCONV=1E-6) satisfied.

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Likelihood

Intercept Intercept & Ratio

Only Covariates Chi-Square DF Pr > ChiSq

8308.925 6654.051 1654.8738 30 <.0001

Type 3 Analysis of Effects

Wald

Effect DF Chi-Square Pr > ChiSq

AFFL\_TILE4 3 628.2010 <.0001

AGE 1 367.0642 <.0001

BILL\_TILE4 2 0.7098 0.7012

CLASS 3 0.6974 0.8738

GENDER 1 191.8521 <.0001

LTIME\_TILE4 3 1.9123 0.5908

NGROUP 6 6.9517 0.3253

TV\_REG 11 14.2666 0.2186

Analysis of Maximum Likelihood Estimates

Standard Wald Standardized

Parameter DF Estimate Error Chi-Square Pr > ChiSq Estimate Exp(Est)

Intercept 1 2.0873 0.1766 139.70 <.0001 8.063

AFFL\_TILE4 1 1 -1.0451 0.0753 192.90 <.0001 0.352

AFFL\_TILE4 2 1 -0.4250 0.0572 55.21 <.0001 0.654

AFFL\_TILE4 3 1 0.2312 0.0484 22.82 <.0001 1.260

AGE 1 -0.0494 0.00258 367.06 <.0001 -0.3767 0.952

BILL\_TILE4 1 1 -0.0724 0.1184 0.37 0.5411 0.930

BILL\_TILE4 2 1 -0.0855 0.1026 0.69 0.4049 0.918

CLASS Gold 1 -0.1035 0.1550 0.45 0.5043 0.902

CLASS Platinum 1 -0.1350 0.1965 0.47 0.4921 0.874

CLASS Silver 1 0.1285 0.1550 0.69 0.4070 1.137

GENDER F 1 0.4671 0.0337 191.85 <.0001 1.595

LTIME\_TILE4 1 1 0.0370 0.0617 0.36 0.5484 1.038

LTIME\_TILE4 2 1 -0.0891 0.0675 1.74 0.1873 0.915

LTIME\_TILE4 3 1 0.0408 0.0507 0.65 0.4204 1.042

NGROUP A 1 -0.0964 0.1267 0.58 0.4465 0.908

NGROUP B 1 -0.1748 0.1058 2.73 0.0986 0.840

NGROUP C 1 -0.0948 0.1025 0.85 0.3553 0.910

NGROUP D 1 -0.0542 0.1030 0.28 0.5992 0.947

NGROUP E 1 0.0860 0.1136 0.57 0.4492 1.090

NGROUP F 1 0.00442 0.1040 0.00 0.9661 1.004

TV\_REG Border 1 0.1926 0.3084 0.39 0.5323 1.212

TV\_REG C Scotland 1 -0.0168 0.1468 0.01 0.9088 0.983

TV\_REG East 1 -0.0567 0.1084 0.27 0.6013 0.945

TV\_REG London 1 0.0114 0.0702 0.03 0.8709 1.011

TV\_REG Midlands 1 0.0152 0.0855 0.03 0.8588 1.015

TV\_REG N East 1 0.0978 0.1451 0.45 0.5002 1.103

TV\_REG N Scot 1 0.2675 0.2381 1.26 0.2613 1.307

TV\_REG N West 1 -0.1994 0.1000 3.98 0.0461 0.819

TV\_REG S & S East 1 -0.1695 0.0944 3.23 0.0724 0.844

TV\_REG S West 1 0.0826 0.1586 0.27 0.6023 1.086

TV\_REG Wales & West 1 0.0460 0.1074 0.18 0.6683 1.047

Odds Ratio Estimates

Point

Effect Estimate

AFFL\_TILE4 1 vs 4 0.102

AFFL\_TILE4 2 vs 4 0.189

AFFL\_TILE4 3 vs 4 0.365

AGE 0.952

BILL\_TILE4 1 vs 3 0.794

BILL\_TILE4 2 vs 3 0.784

CLASS Gold vs Tin 0.808

CLASS Platinum vs Tin 0.783

CLASS Silver vs Tin 1.019

GENDER F vs M 2.545

LTIME\_TILE4 1 vs 4 1.026

LTIME\_TILE4 2 vs 4 0.905

LTIME\_TILE4 3 vs 4 1.030

NGROUP A vs U 0.653

NGROUP B vs U 0.604

NGROUP C vs U 0.654

NGROUP D vs U 0.681

NGROUP E vs U 0.784

NGROUP F vs U 0.722

TV\_REG Border vs Yorkshire 1.589

TV\_REG C Scotland vs Yorkshire 1.289

TV\_REG East vs Yorkshire 1.239

TV\_REG London vs Yorkshire 1.326

TV\_REG Midlands vs Yorkshire 1.331

TV\_REG N East vs Yorkshire 1.446

TV\_REG N Scot vs Yorkshire 1.713

TV\_REG N West vs Yorkshire 1.074

TV\_REG S & S East vs Yorkshire 1.107

TV\_REG S West vs Yorkshire 1.424

TV\_REG Wales & West vs Yorkshire 1.373

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_AIC\_ Akaike's Information Criterion 6716.05 .

\_ASE\_ Average Squared Error 0.19 0.19

\_AVERR\_ Average Error Function 0.56 0.57

\_DFE\_ Degrees of Freedom for Error 5963.00 .

\_DFM\_ Model Degrees of Freedom 31.00 .

\_DFT\_ Total Degrees of Freedom 5994.00 .

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_ERR\_ Error Function 6654.05 2928.66

\_FPE\_ Final Prediction Error 0.19 .

\_MAX\_ Maximum Absolute Error 0.97 0.97

\_MSE\_ Mean Square Error 0.19 0.19

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_NW\_ Number of Estimate Weights 31.00 .

\_RASE\_ Root Average Sum of Squares 0.43 0.44

\_RFPE\_ Root Final Prediction Error 0.44 .

\_RMSE\_ Root Mean Squared Error 0.43 0.44

\_SBC\_ Schwarz's Bayesian Criterion 6923.71 .

\_SSE\_ Sum of Squared Errors 2246.44 994.28

\_SUMW\_ Sum of Case Weights Times Freq 11988.00 5144.00

\_MISC\_ Misclassification Rate 0.29 0.30

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 71.1620 71.8678 2174 36.2696

1 0 28.8380 29.6733 881 14.6980

0 1 28.9554 28.1322 851 14.1975

1 1 71.0446 70.3267 2088 34.8348

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 71.1783 68.8752 894 34.7589

1 0 28.8217 28.4144 362 14.0747

0 1 30.6991 31.1248 404 15.7076

1 1 69.3009 71.5856 912 35.4588

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

881 2174 851 2088

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

362 894 404 912

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 92.4648 1.92465 1.92465 95.3333 95.3333 300 0.90495

10 89.4365 1.86408 1.89437 92.3333 93.8333 300 0.86810

15 81.0246 1.64201 1.81025 81.3333 89.6667 300 0.81956

20 73.0934 1.49220 1.73093 73.9130 85.7381 299 0.77046

25 70.3709 1.59490 1.70371 79.0000 84.3896 300 0.73458

30 66.4242 1.46704 1.66424 72.6667 82.4347 300 0.69608

35 60.7968 1.26938 1.60797 62.8763 79.6473 299 0.65272

40 54.3192 1.09019 1.54319 54.0000 76.4387 300 0.60815

45 48.2344 0.99597 1.48234 49.3333 73.4248 300 0.56057

50 42.6743 0.92503 1.42674 45.8194 70.6707 299 0.51391

55 37.7137 0.88157 1.37714 43.6667 68.2135 300 0.46859

60 33.4121 0.86138 1.33412 42.6667 66.0828 300 0.42169

65 29.2030 0.78736 1.29203 39.0000 63.9979 300 0.38237

70 25.0001 0.70221 1.25000 34.7826 61.9161 299 0.34184

75 20.4761 0.57201 1.20476 28.3333 59.6753 300 0.30043

80 16.2235 0.52490 1.16223 26.0000 57.5688 300 0.26112

85 12.1370 0.46589 1.12137 23.0769 55.5447 299 0.21784

90 7.7350 0.32975 1.07735 16.3333 53.3642 300 0.17339

95 4.0804 0.38358 1.04080 19.0000 51.5540 300 0.13150

100 0.0000 0.22282 1.00000 11.0368 49.5329 299 0.07742

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 89.3639 1.89364 1.89364 93.7984 93.7984 129 0.90005

10 87.7989 1.86234 1.87799 92.2481 93.0233 129 0.86069

15 77.8251 1.57722 1.77825 78.1250 88.0829 128 0.81467

20 70.9152 1.50239 1.70915 74.4186 84.6602 129 0.76786

25 69.8587 1.65608 1.69859 82.0313 84.1369 128 0.72681

30 63.4422 1.31459 1.63442 65.1163 80.9585 129 0.68860

35 56.1743 1.12679 1.56174 55.8140 77.3585 129 0.65118

40 49.6962 1.04096 1.49696 51.5625 74.1497 128 0.61323

45 45.2239 1.09549 1.45224 54.2636 71.9344 129 0.57056

50 40.9733 1.02519 1.40973 50.7813 69.8289 128 0.53175

55 36.5391 0.92334 1.36539 45.7364 67.6325 129 0.48900

60 32.3228 0.86075 1.32323 42.6357 65.5440 129 0.44322

65 27.9886 0.75706 1.27989 37.5000 63.3971 128 0.39695

70 23.5291 0.65730 1.23529 32.5581 61.1882 129 0.35505

75 19.6232 0.64666 1.19623 32.0313 59.2535 128 0.30966

80 15.4603 0.53210 1.15460 26.3566 57.1914 129 0.26958

85 11.4192 0.46950 1.11419 23.2558 55.1898 129 0.22633

90 7.8749 0.47317 1.07875 23.4375 53.4341 128 0.18033

95 4.2461 0.39125 1.04246 19.3798 51.6367 129 0.13318

100 0.0000 0.18927 1.00000 9.3750 49.5334 128 0.07742

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 157 6 0.91558 2.71939

0.85-0.90 386 27 0.87728 6.89022

0.80-0.85 234 46 0.82576 4.67134

0.75-0.80 262 94 0.77286 5.93927

0.70-0.75 311 89 0.72664 6.67334

0.65-0.70 253 110 0.67556 6.05606

0.60-0.65 178 154 0.62406 5.53887

0.55-0.60 154 149 0.57523 5.05506

0.50-0.55 153 176 0.52515 5.48882

0.45-0.50 151 171 0.47598 5.37204

0.40-0.45 143 205 0.42346 5.80581

0.35-0.40 141 235 0.37569 6.27294

0.30-0.35 122 245 0.32519 6.12279

0.25-0.30 108 282 0.27364 6.50651

0.20-0.25 69 256 0.22419 5.42209

0.15-0.20 64 292 0.17357 5.93927

0.10-0.15 57 269 0.12531 5.43877

0.05-0.10 23 181 0.07797 3.40340

0.00-0.05 3 38 0.03864 0.68402

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.95-1.00 0 1 0.95803 0.03888

0.90-0.95 51 3 0.91368 2.09953

0.85-0.90 158 13 0.87583 6.64852

0.80-0.85 105 22 0.82812 4.93779

0.75-0.80 114 39 0.77451 5.94868

0.70-0.75 127 32 0.72487 6.18196

0.65-0.70 112 63 0.67552 6.80404

0.60-0.65 85 82 0.62575 6.49300

0.55-0.60 84 73 0.57341 6.10420

0.50-0.55 76 76 0.52575 5.90980

0.45-0.50 71 79 0.47537 5.83204

0.40-0.45 47 88 0.42312 5.24883

0.35-0.40 61 94 0.37355 6.02644

0.30-0.35 39 93 0.32547 5.13219

0.25-0.30 47 124 0.27535 6.64852

0.20-0.25 31 108 0.22424 5.40435

0.15-0.20 36 101 0.17468 5.32659

0.10-0.15 20 109 0.12544 5.01555

0.05-0.10 8 82 0.07848 3.49922

0.00-0.05 2 16 0.04283 0.69984

# APPENDIX D LOGISTIC REGRESSION STEPWISE MODEL SUMMARY REPORT

Variable Summary

Measurement Frequency

Role Level Count

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

The DMREG Procedure

Model Information

Training Data Set WORK.EM\_DMREG.VIEW

DMDB Catalog WORK.REG4\_DMDB

Target Variable ORGYN (Organics Purchased?)

Target Measurement Level Ordinal

Number of Target Categories 2

Error MBernoulli

Link Function Logit

Number of Model Parameters 31

Number of Observations 5994

Target Profile

Ordered Total

Value ORGYN Frequency

1 1 2969

2 0 3025

Stepwise Selection Procedure

Step 0: Intercept entered.

The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates 1

Optimization Start

Active Constraints 0 Objective Function 4154.4626015

Max Abs Gradient Element 8.058332E-12

Optimization Results

Iterations 0 Function Calls 3

Hessian Calls 1 Active Constraints 0

Objective Function 4154.4626015 Max Abs Gradient Element 8.058332E-12

Ridge 0 Actual Over Pred Change 0

Convergence criterion (ABSGCONV=0.00001) satisfied.

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Likelihood

Intercept Intercept & Ratio

Only Covariates Chi-Square DF Pr > ChiSq

8308.925 8308.925 0.0000 0 .

Analysis of Maximum Likelihood Estimates

Standard Wald Standardized

Parameter DF Estimate Error Chi-Square Pr > ChiSq Estimate Exp(Est)

Intercept 1 -0.0187 0.0258 0.52 0.4695 0.981

Step 1: Effect AFFL\_TILE4 entered.

The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates 4

Optimization Start

Active Constraints 0 Objective Function 4154.4626015

Max Abs Gradient Element 687.82982983

Ratio

Between

Actual

Objective Max Abs and

Function Active Objective Function Gradient Predicted

Iter Restarts Calls Constraints Function Change Element Ridge Change

1 0 2 0 3723 431.1 51.6895 0 1.036

2 0 3 0 3720 3.0765 1.4313 0 1.022

3 0 4 0 3720 0.00417 0.00240 0 1.001

4 0 5 0 3720 2.093E-8 1.233E-8 0 0.997

Optimization Results

Iterations 4 Function Calls 7

Hessian Calls 5 Active Constraints 0

Objective Function 3720.2859975 Max Abs Gradient Element 1.2330707E-8

Ridge 0 Actual Over Pred Change 0.9968012791

Convergence criterion (GCONV=1E-6) satisfied.

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Likelihood

Intercept Intercept & Ratio

Only Covariates Chi-Square DF Pr > ChiSq

8308.925 7440.572 868.3532 3 <.0001

Type 3 Analysis of Effects

Wald

Effect DF Chi-Square Pr > ChiSq

AFFL\_TILE4 3 757.4463 <.0001

Analysis of Maximum Likelihood Estimates

Standard Wald Standardized

Parameter DF Estimate Error Chi-Square Pr > ChiSq Estimate Exp(Est)

Intercept 1 -0.2969 0.0318 87.09 <.0001 0.743

AFFL\_TILE4 1 1 -1.0507 0.0707 220.82 <.0001 0.350

AFFL\_TILE4 2 1 -0.4293 0.0535 64.40 <.0001 0.651

AFFL\_TILE4 3 1 0.2041 0.0450 20.60 <.0001 1.226

Odds Ratio Estimates

Point

Effect Estimate

AFFL\_TILE4 1 vs 4 0.098

AFFL\_TILE4 2 vs 4 0.182

AFFL\_TILE4 3 vs 4 0.342

Step 2: Effect AGE entered.

The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates 5

Optimization Start

Active Constraints 0 Objective Function 4154.4626015

Max Abs Gradient Element 687.82982983

Ratio

Between

Actual

Objective Max Abs and

Function Active Objective Function Gradient Predicted

Iter Restarts Calls Constraints Function Change Element Ridge Change

1 0 2 0 3456 698.6 86.9451 0 1.069

2 0 3 0 3438 17.7030 6.4419 0 1.049

3 0 4 0 3438 0.1064 0.0470 0 1.005

4 0 5 0 3438 5.821E-6 2.781E-6 0 1.000

Optimization Results

Iterations 4 Function Calls 7

Hessian Calls 5 Active Constraints 0

Objective Function 3438.0727647 Max Abs Gradient Element 2.7812482E-6

Ridge 0 Actual Over Pred Change 1.0000414095

Convergence criterion (GCONV=1E-6) satisfied.

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Likelihood

Intercept Intercept & Ratio

Only Covariates Chi-Square DF Pr > ChiSq

8308.925 6876.146 1432.7797 4 <.0001

Type 3 Analysis of Effects

Wald

Effect DF Chi-Square Pr > ChiSq

AFFL\_TILE4 3 653.4069 <.0001

AGE 1 509.6831 <.0001

Analysis of Maximum Likelihood Estimates

Standard Wald Standardized

Parameter DF Estimate Error Chi-Square Pr > ChiSq Estimate Exp(Est)

Intercept 1 2.3116 0.1190 377.26 <.0001 10.090

AFFL\_TILE4 1 1 -1.0376 0.0736 198.51 <.0001 0.354

AFFL\_TILE4 2 1 -0.4150 0.0560 54.87 <.0001 0.660

AFFL\_TILE4 3 1 0.2134 0.0473 20.36 <.0001 1.238

AGE 1 -0.0505 0.00224 509.68 <.0001 -0.3855 0.951

Odds Ratio Estimates

Point

Effect Estimate

AFFL\_TILE4 1 vs 4 0.103

AFFL\_TILE4 2 vs 4 0.191

AFFL\_TILE4 3 vs 4 0.359

AGE 0.951

Step 3: Effect GENDER entered.

The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates 6

Optimization Start

Active Constraints 0 Objective Function 4154.4626015

Max Abs Gradient Element 687.82982983

Ratio

Between

Actual

Objective Max Abs and

Function Active Objective Function Gradient Predicted

Iter Restarts Calls Constraints Function Change Element Ridge Change

1 0 2 0 3366 788.7 96.4338 0 1.079

2 0 3 0 3339 26.7125 8.9472 0 1.062

3 0 4 0 3339 0.2610 0.1058 0 1.008

4 0 5 0 3339 0.000036 0.000016 0 1.000

Optimization Results

Iterations 4 Function Calls 7

Hessian Calls 5 Active Constraints 0

Objective Function 3338.8109638 Max Abs Gradient Element 0.0000157612

Ridge 0 Actual Over Pred Change 1.0001012939

Convergence criterion (GCONV=1E-6) satisfied.

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Likelihood

Intercept Intercept & Ratio

Only Covariates Chi-Square DF Pr > ChiSq

8308.925 6677.622 1631.3033 5 <.0001

Type 3 Analysis of Effects

Wald

Effect DF Chi-Square Pr > ChiSq

AFFL\_TILE4 3 629.7210 <.0001

AGE 1 487.6217 <.0001

GENDER 1 191.1079 <.0001

Analysis of Maximum Likelihood Estimates

Standard Wald Standardized

Parameter DF Estimate Error Chi-Square Pr > ChiSq Estimate Exp(Est)

Intercept 1 2.0904 0.1215 296.17 <.0001 8.088

AFFL\_TILE4 1 1 -1.0375 0.0749 192.07 <.0001 0.354

AFFL\_TILE4 2 1 -0.4229 0.0570 55.13 <.0001 0.655

AFFL\_TILE4 3 1 0.2236 0.0482 21.55 <.0001 1.251

AGE 1 -0.0502 0.00227 487.62 <.0001 -0.3830 0.951

GENDER F 1 0.4643 0.0336 191.11 <.0001 1.591

Odds Ratio Estimates

Point

Effect Estimate

AFFL\_TILE4 1 vs 4 0.103

AFFL\_TILE4 2 vs 4 0.190

AFFL\_TILE4 3 vs 4 0.363

AGE 0.951

GENDER F vs M 2.531

NOTE: No (additional) effects met the 0.05 significance level for entry into the model.

Summary of Stepwise Selection

Effect Number Score Wald

Step Entered DF In Chi-Square Chi-Square Pr > ChiSq

1 AFFL\_TILE4 3 1 831.9004 <.0001

2 AGE 1 2 550.1361 <.0001

3 GENDER 1 3 196.6378 <.0001

The selected model is the model trained in the last step (Step 3). It consists of the following effects:

Intercept AFFL\_TILE4 AGE GENDER

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood Likelihood

Intercept Intercept & Ratio

Only Covariates Chi-Square DF Pr > ChiSq

8308.925 6677.622 1631.3033 5 <.0001

Type 3 Analysis of Effects

Wald

Effect DF Chi-Square Pr > ChiSq

AFFL\_TILE4 3 629.7210 <.0001

AGE 1 487.6217 <.0001

GENDER 1 191.1079 <.0001

Analysis of Maximum Likelihood Estimates

Standard Wald Standardized

Parameter DF Estimate Error Chi-Square Pr > ChiSq Estimate Exp(Est)

Intercept 1 2.0904 0.1215 296.17 <.0001 8.088

AFFL\_TILE4 1 1 -1.0375 0.0749 192.07 <.0001 0.354

AFFL\_TILE4 2 1 -0.4229 0.0570 55.13 <.0001 0.655

AFFL\_TILE4 3 1 0.2236 0.0482 21.55 <.0001 1.251

AGE 1 -0.0502 0.00227 487.62 <.0001 -0.3830 0.951

GENDER F 1 0.4643 0.0336 191.11 <.0001 1.591

Odds Ratio Estimates

Point

Effect Estimate

AFFL\_TILE4 1 vs 4 0.103

AFFL\_TILE4 2 vs 4 0.190

AFFL\_TILE4 3 vs 4 0.363

AGE 0.951

GENDER F vs M 2.531

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_AIC\_ Akaike's Information Criterion 6689.62 .

\_ASE\_ Average Squared Error 0.19 0.19

\_AVERR\_ Average Error Function 0.56 0.57

\_DFE\_ Degrees of Freedom for Error 5988.00 .

\_DFM\_ Model Degrees of Freedom 6.00 .

\_DFT\_ Total Degrees of Freedom 5994.00 .

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_ERR\_ Error Function 6677.62 2910.99

\_FPE\_ Final Prediction Error 0.19 .

\_MAX\_ Maximum Absolute Error 0.96 0.96

\_MSE\_ Mean Square Error 0.19 0.19

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_NW\_ Number of Estimate Weights 6.00 .

\_RASE\_ Root Average Sum of Squares 0.43 0.44

\_RFPE\_ Root Final Prediction Error 0.43 .

\_RMSE\_ Root Mean Squared Error 0.43 0.44

\_SBC\_ Schwarz's Bayesian Criterion 6729.81 .

\_SSE\_ Sum of Squared Errors 2255.87 986.81

\_SUMW\_ Sum of Case Weights Times Freq 11988.00 5144.00

\_MISC\_ Misclassification Rate 0.29 0.29

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 71.1177 71.3058 2157 35.9860

1 0 28.8823 29.5049 876 14.6146

0 1 29.3144 28.6942 868 14.4811

1 1 70.6856 70.4951 2093 34.9183

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 71.6561 69.3374 900 34.9922

1 0 28.3439 27.9435 356 13.8414

0 1 30.2432 30.6626 398 15.4743

1 1 69.7568 72.0565 918 35.6921

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

876 2157 868 2093

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

356 900 398 918

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 91.5422 1.91542 1.91542 94.8763 94.8763 300 0.90100

10 89.5711 1.87600 1.89571 92.9237 93.9000 300 0.86679

15 78.4574 1.56230 1.78457 77.3852 88.3951 300 0.81784

20 73.1852 1.57316 1.73185 77.9230 85.7836 299 0.76872

25 69.5359 1.54951 1.69536 76.7515 83.9760 300 0.73045

30 65.8519 1.47444 1.65852 73.0333 82.1512 300 0.69474

35 62.1921 1.40172 1.62192 69.4314 80.3384 299 0.65131

40 54.2490 0.98700 1.54249 48.8889 76.4039 300 0.60821

45 48.3558 1.01249 1.48356 50.1517 73.4849 300 0.56095

50 42.4156 0.88815 1.42416 43.9928 70.5425 299 0.51665

55 37.2521 0.85668 1.37252 42.4338 67.9849 300 0.46960

60 32.9070 0.85155 1.32907 42.1795 65.8326 300 0.42234

65 28.1021 0.70492 1.28102 34.9167 63.4527 300 0.38132

70 24.5009 0.77564 1.24501 38.4197 61.6689 299 0.34405

75 20.8022 0.69071 1.20802 34.2127 59.8368 300 0.30153

80 16.3532 0.49678 1.16353 24.6067 57.6331 300 0.26148

85 11.9342 0.41052 1.11934 20.3341 55.4442 299 0.22048

90 7.6601 0.35073 1.07660 17.3725 53.3272 300 0.17418

95 4.0095 0.38358 1.04009 19.0000 51.5189 300 0.13362

100 0.0000 0.23632 1.00000 11.7057 49.5329 299 0.08001

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 92.1809 1.92181 1.92181 95.1938 95.1938 129 0.89722

10 85.4514 1.78722 1.85451 88.5271 91.8605 129 0.86068

15 77.9995 1.62979 1.77999 80.7292 88.1693 128 0.81300

20 73.0178 1.58112 1.73018 78.3181 85.7017 129 0.76511

25 69.7541 1.56622 1.69754 77.5805 84.0850 128 0.72583

30 64.7498 1.39806 1.64750 69.2506 81.6062 129 0.68911

35 58.9148 1.23995 1.58915 61.4192 78.7160 129 0.64780

40 50.0232 0.87434 1.50023 43.3093 74.3116 128 0.60762

45 45.3110 1.07724 1.45311 53.3592 71.9775 129 0.56933

50 41.6013 1.08039 1.41601 53.5156 70.1400 128 0.52838

55 37.6156 0.97883 1.37616 48.4848 68.1658 129 0.48842

60 32.9053 0.81237 1.32905 40.2396 65.8325 129 0.44468

65 27.6129 0.63774 1.27613 31.5893 63.2111 128 0.39450

70 23.7284 0.73381 1.23728 36.3480 61.2869 129 0.35283

75 18.7860 0.49244 1.18786 24.3924 58.8388 128 0.31038

80 15.1940 0.61482 1.15194 30.4540 57.0596 129 0.27191

85 11.7331 0.56519 1.11733 27.9956 55.3452 129 0.23100

90 7.9122 0.42630 1.07912 21.1161 53.4526 128 0.17756

95 4.3287 0.40019 1.04329 19.8228 51.6776 129 0.13603

100 0.0000 0.17349 1.00000 8.5937 49.5334 128 0.08108

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 126 5 0.91420 2.18552

0.85-0.90 395 29 0.87820 7.07374

0.80-0.85 221 47 0.82958 4.47114

0.75-0.80 282 88 0.77462 6.17284

0.70-0.75 325 96 0.72404 7.02369

0.65-0.70 266 106 0.67317 6.20621

0.60-0.65 169 156 0.62322 5.42209

0.55-0.60 164 159 0.57267 5.38872

0.50-0.55 145 182 0.52424 5.45546

0.45-0.50 143 171 0.47425 5.23857

0.40-0.45 140 201 0.42259 5.68902

0.35-0.40 138 244 0.37365 6.37304

0.30-0.35 123 234 0.32641 5.95596

0.25-0.30 115 278 0.27351 6.55656

0.20-0.25 72 260 0.22513 5.53887

0.15-0.20 61 282 0.17341 5.72239

0.10-0.15 56 275 0.12748 5.52219

0.05-0.10 26 173 0.07996 3.31999

0.00-0.05 2 39 0.04040 0.68402

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 40 3 0.91331 1.67185

0.85-0.90 167 16 0.87637 7.11509

0.80-0.85 98 17 0.83043 4.47123

0.75-0.80 130 37 0.77341 6.49300

0.70-0.75 129 33 0.72322 6.29860

0.65-0.70 112 60 0.67405 6.68740

0.60-0.65 67 81 0.62249 5.75428

0.55-0.60 90 77 0.57507 6.49300

0.50-0.55 85 74 0.52397 6.18196

0.45-0.50 70 85 0.47486 6.02644

0.40-0.45 46 73 0.42300 4.62675

0.35-0.40 46 106 0.37622 5.90980

0.30-0.35 53 103 0.32537 6.06532

0.25-0.30 42 128 0.27464 6.60964

0.20-0.25 36 84 0.22904 4.66563

0.15-0.20 31 116 0.17414 5.71540

0.10-0.15 23 118 0.12693 5.48212

0.05-0.10 7 72 0.07871 3.07154

0.00-0.05 2 15 0.04354 0.66096

# APPENDIX e LOGISTIC REGRESSION MODEL COMPARISON

Variable Summary

Measurement Frequency

Role Level Count

TARGET BINARY 1

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.789 0.577 0.433 0.578 0.431 0.608

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.776 0.552 0.408 0.513 0.406 0.532

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.787 0.574 0.436 0.542 0.431 0.651

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.779 0.558 0.423 0.476 0.412 0.527

Fit Statistics

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Train: Valid:

Valid: Average Train: Average

Selected Model Misclassification Squared Misclassification Squared

Model Node Model Description Rate Error Rate Error

Y Reg4 Regression(STEPWISE) 0.29316 0.18818 0.29096 0.19184

Reg3 Regression(ENTER) 0.29782 0.18739 0.28896 0.19329

Fit Statistics Table

Target: ORGYN

Data Role=Train

Statistics Reg4 Reg3

Train: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff 0.65 0.61

Train: Kolmogorov-Smirnov Statistic 0.44 0.43

Train: Akaike's Information Criterion 6689.62 6716.05

Train: Average Squared Error 0.19 0.19

Train: Roc Index 0.79 0.79

Train: Average Error Function 0.56 0.56

Train: Cumulative Percent Captured Response 18.98 18.96

Train: Percent Captured Response 9.39 9.33

Selection Criterion: Valid: Misclassification Rate 0.29 0.30

Train: Degrees of Freedom for Error 5988.00 5963.00

Train: Model Degrees of Freedom 6.00 31.00

Train: Total Degrees of Freedom 5994.00 5994.00

Train: Divisor for ASE 11988.00 11988.00

Train: Error Function 6677.62 6654.05

Train: Final Prediction Error 0.19 0.19

Train: Gain 89.57 89.44

Train: Gini Coefficient 0.57 0.58

Train: Bin-Based Two-Way Kolmogorov-Smirnov Statistic 0.43 0.43

Train: Kolmogorov-Smirnov Probability Cutoff 0.54 0.58

Train: Cumulative Lift 1.90 1.89

Train: Lift 1.88 1.86

Train: Maximum Absolute Error 0.96 0.97

Train: Misclassification Rate 0.29 0.29

Train: Mean Square Error 0.19 0.19

Train: Sum of Frequencies 5994.00 5994.00

Train: Number of Estimate Weights 6.00 31.00

Train: Root Average Sum of Squares 0.43 0.43

Train: Cumulative Percent Response 93.90 93.83

Train: Percent Response 92.92 92.33

Train: Root Final Prediction Error 0.43 0.44

Train: Root Mean Squared Error 0.43 0.43

Train: Schwarz's Bayesian Criterion 6729.81 6923.71

Train: Sum of Squared Errors 2255.87 2246.44

Train: Sum of Case Weights Times Freq 11988.00 11988.00

Data Role=Valid

Statistics Reg4 Reg3

Valid: Kolmogorov-Smirnov Statistic 0.42 0.41

Valid: Average Squared Error 0.19 0.19

Valid: Roc Index 0.78 0.78

Valid: Average Error Function 0.57 0.57

Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff 0.53 0.53

Valid: Cumulative Percent Captured Response 18.60 18.84

Valid: Percent Captured Response 8.96 9.34

Valid: Divisor for VASE 5144.00 5144.00

Valid: Error Function 2910.99 2928.66

Valid: Gain 85.45 87.80

Valid: Gini Coefficient 0.56 0.55

Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic 0.41 0.41

Valid: Kolmogorov-Smirnov Probability Cutoff 0.48 0.51

Valid: Cumulative Lift 1.85 1.88

Valid: Lift 1.79 1.86

Valid: Maximum Absolute Error 0.96 0.97

Valid: Misclassification Rate 0.29 0.30

Valid: Mean Square Error 0.19 0.19

Valid: Sum of Frequencies 2572.00 2572.00

Valid: Root Average Squared Error 0.44 0.44

Valid: Cumulative Percent Response 91.86 93.02

Valid: Percent Response 88.53 92.25

Valid: Root Mean Square Error 0.44 0.44

Valid: Sum of Square Errors 986.81 994.28

Valid: Sum of Case Weights Times Freq 5144.00 5144.00

Event Classification Table

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Model Data False True False True

Node Model Description Role Target Target Label Negative Negative Positive Positive

Reg3 Regression(ENTER) TRAIN ORGYN Organics Purchased? 881 2174 851 2088

Reg3 Regression(ENTER) VALIDATE ORGYN Organics Purchased? 362 894 404 912

Reg4 Regression(STEPWISE) TRAIN ORGYN Organics Purchased? 876 2157 868 2093

Reg4 Regression(STEPWISE) VALIDATE ORGYN Organics Purchased? 356 900 398 918

# APPENDIX G DECISION TREE CHI-SQUARE MODEL

Variable Summary

Measurement Frequency

Role Level Count

ID INTERVAL 1

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Variable Importance

Ratio of

Number of Validation

Variable Splitting Validation to Training

Name Label Rules Importance Importance Importance

AGE Age 2 1.0000 1.0000 1.0000

AFFL\_TILE4 3 0.7080 0.6429 0.9080

GENDER 2 0.2623 0.2280 0.8692

Tree Leaf Report

Training

Node Training Percent Validation Validation

Id Depth Observations 1 Observations Percent 1

7 2 2732 0.25 1165 0.26

4 2 896 0.92 367 0.90

13 3 822 0.65 343 0.62

11 3 689 0.75 332 0.71

20 4 341 0.64 146 0.68

12 3 296 0.38 124 0.38

32 5 165 0.50 68 0.53

33 5 53 0.28 27 0.41

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_MISC\_ Misclassification Rate 0.26 0.27

\_MAX\_ Maximum Absolute Error 0.92 0.92

\_SSE\_ Sum of Squared Errors 2190.56 979.94

\_ASE\_ Average Squared Error 0.18 0.19

\_RASE\_ Root Average Squared Error 0.43 0.44

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_DFT\_ Total Degrees of Freedom 5994.00 .

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 74.0019 75.3719 2280 38.0380

1 0 25.9981 26.9788 801 13.3634

0 1 25.5750 24.6281 745 12.4291

1 1 74.4250 73.0212 2168 36.1695

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 72.7964 73.8059 958 37.2473

1 0 27.2036 28.1005 358 13.9191

0 1 27.0701 26.1941 340 13.2193

1 1 72.9299 71.8995 916 35.6143

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

801 2280 745 2168

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

358 958 340 916

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 84.7619 1.84762 1.84762 91.5179 91.5179 300 0.91518

10 84.7619 1.84762 1.84762 91.5179 91.5179 300 0.91518

15 84.6114 1.84310 1.84611 91.2942 91.4433 300 0.91294

20 76.2051 1.50902 1.76205 74.7460 87.2794 299 0.74746

25 71.1411 1.50902 1.71141 74.7460 84.7711 300 0.74746

30 65.3583 1.36463 1.65358 67.5943 81.9067 300 0.67594

35 60.4133 1.30661 1.60413 64.7202 79.4573 299 0.64720

40 56.6912 1.30661 1.56691 64.7202 77.6136 300 0.64720

45 53.6247 1.29113 1.53625 63.9533 76.0947 300 0.63953

50 48.1835 0.99085 1.48183 49.0798 73.3995 299 0.49080

55 41.0591 0.69887 1.41059 34.6170 69.8706 300 0.34617

60 33.4484 0.49806 1.33448 24.6706 66.1008 300 0.24671

65 27.0095 0.49806 1.27009 24.6706 62.9114 300 0.24671

70 21.5081 0.49806 1.21508 24.6706 60.1864 299 0.24671

75 16.7237 0.49806 1.16724 24.6706 57.8166 300 0.24671

80 12.5379 0.49806 1.12538 24.6706 55.7433 300 0.24671

85 8.8565 0.49806 1.08857 24.6706 53.9198 299 0.24671

90 5.5729 0.49806 1.05573 24.6706 52.2933 300 0.24671

95 2.6353 0.49806 1.02635 24.6706 50.8382 300 0.24671

100 0.0000 0.49806 1.00000 24.6706 49.5329 299 0.24671

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 82.0805 1.82081 1.82081 90.1907 90.1907 129 0.91518

10 82.0805 1.82081 1.82081 90.1907 90.1907 129 0.91518

15 80.1819 1.76355 1.80182 87.3546 89.2503 128 0.89028

20 70.9955 1.43508 1.70996 71.0843 84.7000 129 0.74746

25 65.5236 1.43508 1.65524 71.0843 81.9895 128 0.74746

30 60.1852 1.33576 1.60185 66.1647 79.3452 129 0.69072

35 55.2846 1.25957 1.55285 62.3907 76.9178 129 0.64720

40 51.6364 1.25957 1.51636 62.3907 75.1107 128 0.64720

45 49.8713 1.35792 1.49871 67.2623 74.2364 129 0.64009

50 45.5841 1.06799 1.45584 52.9011 72.1128 128 0.50575

55 39.2585 0.76198 1.39258 37.7435 68.9795 129 0.35025

60 31.9670 0.51987 1.31967 25.7511 65.3678 129 0.24671

65 25.8442 0.51987 1.25844 25.7511 62.3349 128 0.24671

70 20.5540 0.51987 1.20554 25.7511 59.7146 129 0.24671

75 16.0042 0.51987 1.16004 25.7511 57.4609 128 0.24671

80 11.9915 0.51987 1.11992 25.7511 55.4732 129 0.24671

85 8.4522 0.51987 1.08452 25.7511 53.7201 129 0.24671

90 5.3301 0.51987 1.05330 25.7511 52.1736 128 0.24671

95 2.5146 0.51987 1.02515 25.7511 50.7790 129 0.24671

100 0.0000 0.51987 1.00000 25.7511 49.5334 128 0.24671

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 820 76 0.91518 14.9483

0.70-0.75 515 174 0.74746 11.4948

0.60-0.65 750 413 0.64488 19.4027

0.50-0.55 83 82 0.50303 2.7528

0.35-0.40 112 184 0.37838 4.9383

0.25-0.30 15 38 0.28302 0.8842

0.20-0.25 674 2058 0.24671 45.5789

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 331 36 0.91518 14.2691

0.70-0.75 236 96 0.74746 12.9082

0.60-0.65 313 176 0.64484 19.0124

0.50-0.55 36 32 0.50303 2.6439

0.35-0.40 47 77 0.37838 4.8212

0.25-0.30 11 16 0.28302 1.0498

0.20-0.25 300 865 0.24671 45.2955

\*------------------------------------------------------------\*

Node = 3

\*------------------------------------------------------------\*

if Age >= 44.5 or MISSING

then

Tree Node Identifier = 3

Number of Observations = 7558

Predicted: ORGYN=1 = 0.17

Predicted: ORGYN=0 = 0.83

\*------------------------------------------------------------\*

Node = 4

\*------------------------------------------------------------\*

if Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 4

Number of Observations = 1034

Predicted: ORGYN=1 = 0.78

Predicted: ORGYN=0 = 0.22

\*------------------------------------------------------------\*

Node = 10

\*------------------------------------------------------------\*

if GENDER IS ONE OF: M

AND Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 10

Number of Observations = 533

Predicted: ORGYN=1 = 0.29

Predicted: ORGYN=0 = 0.71

\*------------------------------------------------------------\*

Node = 22

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 3 or MISSING

then

Tree Node Identifier = 22

Number of Observations = 715

Predicted: ORGYN=1 = 0.61

Predicted: ORGYN=0 = 0.39

\*------------------------------------------------------------\*

Node = 23

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 2, 1

then

Tree Node Identifier = 23

Number of Observations = 623

Predicted: ORGYN=1 = 0.42

Predicted: ORGYN=0 = 0.58

**APPENDIX I DECISION TREE GINI MODEL**

Variable Summary

Measurement Frequency

Role Level Count

ID INTERVAL 1

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Variable Importance

Ratio of

Number of Validation

Variable Splitting Validation to Training

Name Label Rules Importance Importance Importance

AGE Age 2 1.0000 1.0000 1.0000

AFFL\_TILE4 4 0.7129 0.6513 0.9137

GENDER 2 0.2893 0.3068 1.0605

TV\_REG TV Region 2 0.1096 0.1190 1.0855

LTIME\_TILE4 1 0.0547 0.0514 0.9401

Tree Leaf Report

Training

Node Training Percent Validation Validation

Id Depth Observations 1 Observations Percent 1

7 2 2732 0.25 1165 0.26

4 2 896 0.92 367 0.90

13 3 822 0.65 343 0.62

23 4 529 0.80 261 0.77

12 3 296 0.38 124 0.38

41 5 221 0.69 102 0.74

21 4 218 0.45 95 0.49

78 6 86 0.48 35 0.43

44 5 80 0.69 31 0.39

88 6 44 0.39 17 0.47

89 6 36 0.58 23 0.61

79 6 34 0.71 9 1.00

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_MISC\_ Misclassification Rate 0.26 0.27

\_MAX\_ Maximum Absolute Error 0.92 0.92

\_SSE\_ Sum of Squared Errors 2171.69 969.00

\_ASE\_ Average Squared Error 0.18 0.19

\_RASE\_ Root Average Squared Error 0.43 0.43

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_DFT\_ Total Degrees of Freedom 5994.00 .

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 72.0972 80.4628 2434 40.6073

1 0 27.9028 31.7279 942 15.7157

0 1 22.5745 19.5372 591 9.8599

1 1 77.4255 68.2721 2027 33.8172

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 70.9610 78.5054 1019 39.6190

1 0 29.0390 32.7316 417 16.2131

0 1 24.5599 21.4946 279 10.8476

1 1 75.4401 67.2684 857 33.3204

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

942 2434 591 2027

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

417 1019 279 857

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 84.7619 1.84762 1.84762 91.5179 91.5179 300 0.91518

10 84.7619 1.84762 1.84762 91.5179 91.5179 300 0.91518

15 84.6565 1.84446 1.84657 91.3613 91.4657 300 0.91361

20 78.7699 1.61051 1.78770 79.7732 88.5498 299 0.79773

25 74.2352 1.56112 1.74235 77.3265 86.3037 300 0.77327

30 68.2468 1.38325 1.68247 68.5162 83.3375 300 0.68516

35 62.8902 1.30661 1.62890 64.7202 80.6842 299 0.64720

40 58.8582 1.30661 1.58858 64.7202 78.6870 300 0.64720

45 54.5304 1.19937 1.54530 59.4082 76.5433 300 0.59408

50 47.8430 0.87500 1.47843 43.3412 73.2309 299 0.43341

55 41.0591 0.73288 1.41059 36.3017 69.8706 300 0.36302

60 33.4484 0.49806 1.33448 24.6706 66.1008 300 0.24671

65 27.0095 0.49806 1.27009 24.6706 62.9114 300 0.24671

70 21.5081 0.49806 1.21508 24.6706 60.1864 299 0.24671

75 16.7237 0.49806 1.16724 24.6706 57.8166 300 0.24671

80 12.5379 0.49806 1.12538 24.6706 55.7433 300 0.24671

85 8.8565 0.49806 1.08857 24.6706 53.9198 299 0.24671

90 5.5729 0.49806 1.05573 24.6706 52.2933 300 0.24671

95 2.6353 0.49806 1.02635 24.6706 50.8382 300 0.24671

100 0.0000 0.49806 1.00000 24.6706 49.5329 299 0.24671

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 82.0805 1.82081 1.82081 90.1907 90.1907 129 0.91518

10 82.0805 1.82081 1.82081 90.1907 90.1907 129 0.91518

15 80.8089 1.78246 1.80809 88.2913 89.5609 128 0.89775

20 74.6566 1.56247 1.74657 77.3946 86.5134 129 0.79773

25 71.5579 1.59090 1.71558 78.8029 84.9785 128 0.78633

30 64.8146 1.31203 1.64815 64.9892 81.6383 129 0.69045

35 59.2511 1.25957 1.59251 62.3907 78.8826 129 0.64720

40 55.1095 1.25957 1.55110 62.3907 76.8311 128 0.64720

45 51.0517 1.18684 1.51052 58.7882 74.8211 129 0.60674

50 45.7044 0.97328 1.45704 48.2098 72.1724 128 0.44225

55 39.2585 0.74999 1.39258 37.1496 68.9795 129 0.37021

60 31.9670 0.51987 1.31967 25.7511 65.3678 129 0.24671

65 25.8442 0.51987 1.25844 25.7511 62.3349 128 0.24671

70 20.5540 0.51987 1.20554 25.7511 59.7146 129 0.24671

75 16.0042 0.51987 1.16004 25.7511 57.4609 128 0.24671

80 11.9915 0.51987 1.11992 25.7511 55.4732 129 0.24671

85 8.4522 0.51987 1.08452 25.7511 53.7201 129 0.24671

90 5.3301 0.51987 1.05330 25.7511 52.1736 128 0.24671

95 2.5146 0.51987 1.02515 25.7511 50.7790 129 0.24671

100 0.0000 0.51987 1.00000 25.7511 49.5334 128 0.24671

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 820 76 0.91518 14.9483

0.75-0.80 422 107 0.79773 8.8255

0.70-0.75 24 10 0.70588 0.5672

0.65-0.70 208 93 0.69103 5.0217

0.60-0.65 532 290 0.64720 13.7137

0.55-0.60 21 15 0.58333 0.6006

0.45-0.50 41 45 0.47674 1.4348

0.40-0.45 98 120 0.44954 3.6370

0.35-0.40 129 211 0.37941 5.6723

0.20-0.25 674 2058 0.24671 45.5789

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 331 36 0.91518 14.2691

0.75-0.80 202 59 0.79773 10.1477

0.70-0.75 9 0 0.70588 0.3499

0.65-0.70 87 46 0.69119 5.1711

0.60-0.65 214 129 0.64720 13.3359

0.55-0.60 14 9 0.58333 0.8942

0.45-0.50 15 20 0.47674 1.3608

0.40-0.45 47 48 0.44954 3.6936

0.35-0.40 55 86 0.37934 5.4821

0.20-0.25 300 865 0.24671 45.2955

**APPENDIX J DECISION TREE GINI MODEL RULES**

\*------------------------------------------------------------\*

Node = 4

\*------------------------------------------------------------\*

if Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 4

Number of Observations = 1034

Predicted: ORGYN=1 = 0.78

Predicted: ORGYN=0 = 0.22

\*------------------------------------------------------------\*

Node = 7

\*------------------------------------------------------------\*

if Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 7

Number of Observations = 5761

Predicted: ORGYN=1 = 0.12

Predicted: ORGYN=0 = 0.88

\*------------------------------------------------------------\*

Node = 10

\*------------------------------------------------------------\*

if GENDER IS ONE OF: M

AND Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 10

Number of Observations = 533

Predicted: ORGYN=1 = 0.29

Predicted: ORGYN=0 = 0.71

\*------------------------------------------------------------\*

Node = 13

\*------------------------------------------------------------\*

if GENDER IS ONE OF: M

AND Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 13

Number of Observations = 569

Predicted: ORGYN=1 = 0.22

Predicted: ORGYN=0 = 0.78

\*------------------------------------------------------------\*

Node = 23

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 44.5 AND Age >= 39.5

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 23

Number of Observations = 509

Predicted: ORGYN=1 = 0.39

Predicted: ORGYN=0 = 0.61

\*------------------------------------------------------------\*

Node = 25

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age >= 59.5

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 25

Number of Observations = 568

Predicted: ORGYN=1 = 0.36

Predicted: ORGYN=0 = 0.64

\*------------------------------------------------------------\*

Node = 44

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 39.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 2, 1

then

Tree Node Identifier = 44

Number of Observations = 357

Predicted: ORGYN=1 = 0.49

Predicted: ORGYN=0 = 0.51

\*------------------------------------------------------------\*

Node = 45

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 39.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 3 or MISSING

then

Tree Node Identifier = 45

Number of Observations = 472

Predicted: ORGYN=1 = 0.67

Predicted: ORGYN=0 = 0.33

\*------------------------------------------------------------\*

Node = 48

\*------------------------------------------------------------\*

if TV Region IS ONE OF: N EAST

AND GENDER IS ONE OF: F or MISSING

AND Age < 59.5 AND Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 48

Number of Observations = 19

Predicted: ORGYN=1 = 0.74

Predicted: ORGYN=0 = 0.26

\*------------------------------------------------------------\*

Node = 49

\*------------------------------------------------------------\*

if TV Region IS ONE OF: S & S EAST, LONDON, WALES & WEST, MIDLANDS, N WEST, C SCOTLAND, EAST, N SCOT, YORKSHIRE, S WEST or MISSING

AND GENDER IS ONE OF: F or MISSING

AND Age < 59.5 AND Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 49

Number of Observations = 641

Predicted: ORGYN=1 = 0.45

Predicted: ORGYN=0 = 0.55

**APPENDIX K DECISION TREE ENTROPY Model SUMMARY REPORT**

Variable Summary

Measurement Frequency

Role Level Count

ID INTERVAL 1

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Variable Importance

Ratio of

Number of Validation

Variable Splitting Validation to Training

Name Label Rules Importance Importance Importance

AGE Age 2 1.0000 1.0000 1.0000

AFFL\_TILE4 4 0.7129 0.6513 0.9137

GENDER 2 0.2893 0.3068 1.0605

TV\_REG TV Region 2 0.1096 0.1190 1.0855

LTIME\_TILE4 1 0.0547 0.0514 0.9401

Tree Leaf Report

Training

Node Training Percent Validation Validation

Id Depth Observations 1 Observations Percent 1

7 2 2732 0.25 1165 0.26

4 2 896 0.92 367 0.90

13 3 822 0.65 343 0.62

23 4 529 0.80 261 0.77

12 3 296 0.38 124 0.38

39 5 221 0.69 102 0.74

21 4 218 0.45 95 0.49

72 6 86 0.48 35 0.43

42 5 80 0.69 31 0.39

82 6 44 0.39 17 0.47

83 6 36 0.58 23 0.61

73 6 34 0.71 9 1.00

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_MISC\_ Misclassification Rate 0.26 0.27

\_MAX\_ Maximum Absolute Error 0.92 0.92

\_SSE\_ Sum of Squared Errors 2171.69 969.00

\_ASE\_ Average Squared Error 0.18 0.19

\_RASE\_ Root Average Squared Error 0.43 0.43

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_DFT\_ Total Degrees of Freedom 5994.00 .

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 72.0972 80.4628 2434 40.6073

1 0 27.9028 31.7279 942 15.7157

0 1 22.5745 19.5372 591 9.8599

1 1 77.4255 68.2721 2027 33.8172

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 70.9610 78.5054 1019 39.6190

1 0 29.0390 32.7316 417 16.2131

0 1 24.5599 21.4946 279 10.8476

1 1 75.4401 67.2684 857 33.3204

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

942 2434 591 2027

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

417 1019 279 857

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 84.7619 1.84762 1.84762 91.5179 91.5179 300 0.91518

10 84.7619 1.84762 1.84762 91.5179 91.5179 300 0.91518

15 84.6565 1.84446 1.84657 91.3613 91.4657 300 0.91361

20 78.7699 1.61051 1.78770 79.7732 88.5498 299 0.79773

25 74.2352 1.56112 1.74235 77.3265 86.3037 300 0.77327

30 68.2468 1.38325 1.68247 68.5162 83.3375 300 0.68516

35 62.8902 1.30661 1.62890 64.7202 80.6842 299 0.64720

40 58.8582 1.30661 1.58858 64.7202 78.6870 300 0.64720

45 54.5304 1.19937 1.54530 59.4082 76.5433 300 0.59408

50 47.8430 0.87500 1.47843 43.3412 73.2309 299 0.43341

55 41.0591 0.73288 1.41059 36.3017 69.8706 300 0.36302

60 33.4484 0.49806 1.33448 24.6706 66.1008 300 0.24671

65 27.0095 0.49806 1.27009 24.6706 62.9114 300 0.24671

70 21.5081 0.49806 1.21508 24.6706 60.1864 299 0.24671

75 16.7237 0.49806 1.16724 24.6706 57.8166 300 0.24671

80 12.5379 0.49806 1.12538 24.6706 55.7433 300 0.24671

85 8.8565 0.49806 1.08857 24.6706 53.9198 299 0.24671

90 5.5729 0.49806 1.05573 24.6706 52.2933 300 0.24671

95 2.6353 0.49806 1.02635 24.6706 50.8382 300 0.24671

100 0.0000 0.49806 1.00000 24.6706 49.5329 299 0.24671

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 82.0805 1.82081 1.82081 90.1907 90.1907 129 0.91518

10 82.0805 1.82081 1.82081 90.1907 90.1907 129 0.91518

15 80.8089 1.78246 1.80809 88.2913 89.5609 128 0.89775

20 74.6566 1.56247 1.74657 77.3946 86.5134 129 0.79773

25 71.5579 1.59090 1.71558 78.8029 84.9785 128 0.78633

30 64.8146 1.31203 1.64815 64.9892 81.6383 129 0.69045

35 59.2511 1.25957 1.59251 62.3907 78.8826 129 0.64720

40 55.1095 1.25957 1.55110 62.3907 76.8311 128 0.64720

45 51.0517 1.18684 1.51052 58.7882 74.8211 129 0.60674

50 45.7044 0.97328 1.45704 48.2098 72.1724 128 0.44225

55 39.2585 0.74999 1.39258 37.1496 68.9795 129 0.37021

60 31.9670 0.51987 1.31967 25.7511 65.3678 129 0.24671

65 25.8442 0.51987 1.25844 25.7511 62.3349 128 0.24671

70 20.5540 0.51987 1.20554 25.7511 59.7146 129 0.24671

75 16.0042 0.51987 1.16004 25.7511 57.4609 128 0.24671

80 11.9915 0.51987 1.11992 25.7511 55.4732 129 0.24671

85 8.4522 0.51987 1.08452 25.7511 53.7201 129 0.24671

90 5.3301 0.51987 1.05330 25.7511 52.1736 128 0.24671

95 2.5146 0.51987 1.02515 25.7511 50.7790 129 0.24671

100 0.0000 0.51987 1.00000 25.7511 49.5334 128 0.24671

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 820 76 0.91518 14.9483

0.75-0.80 422 107 0.79773 8.8255

0.70-0.75 24 10 0.70588 0.5672

0.65-0.70 208 93 0.69103 5.0217

0.60-0.65 532 290 0.64720 13.7137

0.55-0.60 21 15 0.58333 0.6006

0.45-0.50 41 45 0.47674 1.4348

0.40-0.45 98 120 0.44954 3.6370

0.35-0.40 129 211 0.37941 5.6723

0.20-0.25 674 2058 0.24671 45.5789

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 331 36 0.91518 14.2691

0.75-0.80 202 59 0.79773 10.1477

0.70-0.75 9 0 0.70588 0.3499

0.65-0.70 87 46 0.69119 5.1711

0.60-0.65 214 129 0.64720 13.3359

0.55-0.60 14 9 0.58333 0.8942

0.45-0.50 15 20 0.47674 1.3608

0.40-0.45 47 48 0.44954 3.6936

0.35-0.40 55 86 0.37934 5.4821

0.20-0.25 300 865 0.24671 45.2955

**APPENDIX L DECISION TREE ENTROPY Model RULES**

\*------------------------------------------------------------\*

Node = 4

\*------------------------------------------------------------\*

if Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 4

Number of Observations = 1034

Predicted: ORGYN=1 = 0.78

Predicted: ORGYN=0 = 0.22

\*------------------------------------------------------------\*

Node = 7

\*------------------------------------------------------------\*

if Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 7

Number of Observations = 5761

Predicted: ORGYN=1 = 0.12

Predicted: ORGYN=0 = 0.88

\*------------------------------------------------------------\*

Node = 10

\*------------------------------------------------------------\*

if GENDER IS ONE OF: M

AND Age < 44.5

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 10

Number of Observations = 533

Predicted: ORGYN=1 = 0.29

Predicted: ORGYN=0 = 0.71

\*------------------------------------------------------------\*

Node = 13

\*------------------------------------------------------------\*

if GENDER IS ONE OF: M

AND Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 13

Number of Observations = 569

Predicted: ORGYN=1 = 0.22

Predicted: ORGYN=0 = 0.78

\*------------------------------------------------------------\*

Node = 23

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 44.5 AND Age >= 39.5

AND AFFL\_TILE4 IS ONE OF: 2, 3, 1 or MISSING

then

Tree Node Identifier = 23

Number of Observations = 509

Predicted: ORGYN=1 = 0.39

Predicted: ORGYN=0 = 0.61

\*------------------------------------------------------------\*

Node = 25

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age >= 59.5

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 25

Number of Observations = 568

Predicted: ORGYN=1 = 0.36

Predicted: ORGYN=0 = 0.64

\*------------------------------------------------------------\*

Node = 44

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 39.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 2, 1

then

Tree Node Identifier = 44

Number of Observations = 357

Predicted: ORGYN=1 = 0.49

Predicted: ORGYN=0 = 0.51

\*------------------------------------------------------------\*

Node = 45

\*------------------------------------------------------------\*

if GENDER IS ONE OF: F or MISSING

AND Age < 39.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 3 or MISSING

then

Tree Node Identifier = 45

Number of Observations = 472

Predicted: ORGYN=1 = 0.67

Predicted: ORGYN=0 = 0.33

\*------------------------------------------------------------\*

Node = 48

\*------------------------------------------------------------\*

if TV Region IS ONE OF: N EAST

AND GENDER IS ONE OF: F or MISSING

AND Age < 59.5 AND Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 48

Number of Observations = 19

Predicted: ORGYN=1 = 0.74

Predicted: ORGYN=0 = 0.26

\*------------------------------------------------------------\*

Node = 49

\*------------------------------------------------------------\*

if TV Region IS ONE OF: S & S EAST, LONDON, WALES & WEST, MIDLANDS, N WEST, C SCOTLAND, EAST, N SCOT, YORKSHIRE, S WEST or MISSING

AND GENDER IS ONE OF: F or MISSING

AND Age < 59.5 AND Age >= 44.5 or MISSING

AND AFFL\_TILE4 IS ONE OF: 4

then

Tree Node Identifier = 49

Number of Observations = 641

Predicted: ORGYN=1 = 0.45

Predicted: ORGYN=0 = 0.55

# APPENDIX M DECISION TREE MODEL COMPARISON

Variable Summary

Measurement Frequency

Role Level Count

TARGET BINARY 1

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.781 0.562 0.484 0.379 0.478 0.643

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.764 0.529 0.457 0.379 0.452 0.509

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.785 0.57 0.487 0.477 0.486 0.562

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.771 0.543 0.458 0.477 0.455 0.562

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.785 0.57 0.487 0.477 0.486 0.562

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.771 0.543 0.458 0.477 0.455 0.562

Fit Statistics

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Train: Valid:

Valid: Average Train: Average

Selected Model Model Misclassification Squared Misclassification Squared

Model Node Description Rate Error Rate Error

Y Tree5 DT-GINI 0.27061 0.18116 0.25576 0.18837

Tree6 DT-Entropy 0.27061 0.18116 0.25576 0.18837

Tree4 DT-CHISQ 0.27138 0.18273 0.25792 0.19050

Fit Statistics Table

Target: ORGYN

Data Role=Train

Statistics Tree5 Tree6 Tree4

Train: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff 0.56 0.56 0.64

Train: Kolmogorov-Smirnov Statistic 0.49 0.49 0.48

Train: Average Squared Error 0.18 0.18 0.18

Train: Roc Index 0.79 0.79 0.78

Train: Cumulative Percent Captured Response 18.49 18.49 18.49

Train: Percent Captured Response 9.25 9.25 9.25

Selection Criterion: Valid: Misclassification Rate 0.27 0.27 0.27

Train: Total Degrees of Freedom 5994.00 5994.00 5994.00

Train: Divisor for ASE 11988.00 11988.00 11988.00

Train: Gain 84.76 84.76 84.76

Train: Gini Coefficient 0.57 0.57 0.56

Train: Bin-Based Two-Way Kolmogorov-Smirnov Statistic 0.49 0.49 0.48

Train: Kolmogorov-Smirnov Probability Cutoff 0.48 0.48 0.38

Train: Cumulative Lift 1.85 1.85 1.85

Train: Lift 1.85 1.85 1.85

Train: Maximum Absolute Error 0.92 0.92 0.92

Train: Misclassification Rate 0.26 0.26 0.26

Train: Sum of Frequencies 5994.00 5994.00 5994.00

Train: Root Average Squared Error 0.43 0.43 0.43

Train: Cumulative Percent Response 91.52 91.52 91.52

Train: Percent Response 91.52 91.52 91.52

Train: Sum of Squared Errors 2171.69 2171.69 2190.56

Data Role=Valid

Statistics Tree5 Tree6 Tree4

Valid: Kolmogorov-Smirnov Statistic 0.46 0.46 0.46

Valid: Average Squared Error 0.19 0.19 0.19

Valid: Roc Index 0.77 0.77 0.76

Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff 0.56 0.56 0.51

Valid: Cumulative Percent Captured Response 18.26 18.26 18.26

Valid: Percent Captured Response 9.13 9.13 9.13

Valid: Divisor for VASE 5144.00 5144.00 5144.00

Valid: Gain 82.08 82.08 82.08

Valid: Gini Coefficient 0.54 0.54 0.53

Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic 0.46 0.46 0.45

Valid: Kolmogorov-Smirnov Probability Cutoff 0.48 0.48 0.38

Valid: Cumulative Lift 1.82 1.82 1.82

Valid: Lift 1.82 1.82 1.82

Valid: Maximum Absolute Error 0.92 0.92 0.92

Valid: Misclassification Rate 0.27 0.27 0.27

Valid: Sum of Frequencies 2572.00 2572.00 2572.00

Valid: Root Average Squared Error 0.43 0.43 0.44

Valid: Cumulative Percent Response 90.19 90.19 90.19

Valid: Percent Response 90.19 90.19 90.19

Valid: Sum of Squared Errors 969.00 969.00 979.94

Event Classification Table

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Model Model Data False True False True

Node Description Role Target Target Label Negative Negative Positive Positive

Tree4 DT-CHISQ TRAIN ORGYN Organics Purchased? 801 2280 745 2168

Tree4 DT-CHISQ VALIDATE ORGYN Organics Purchased? 358 958 340 916

Tree5 DT-GINI TRAIN ORGYN Organics Purchased? 942 2434 591 2027

Tree5 DT-GINI VALIDATE ORGYN Organics Purchased? 417 1019 279 857

Tree6 DT-Entropy TRAIN ORGYN Organics Purchased? 942 2434 591 2027

Tree6 DT-Entropy VALIDATE ORGYN Organics Purchased? 417 1019 279 857

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

# APPENDIX N NEURAL NETWORK MODEL SUMMARY REPORT

Variable Summary

Measurement Frequency

Role Level Count

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 6

REJECTED INTERVAL 5

REJECTED NOMINAL 6

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

The NEURAL Procedure

Preliminary Starting Objective Number

Training Random Function of Terminating

Run Seed Value Iterations Criteria

1 12345 0.521714040584 10

2 1204123714 0.522842314007 10

3 1981206845 0.523011268012 10

4 785699814 0.523254915127 10

5 1517691079 0.520663991213 10

The NEURAL Procedure

Optimization Start

Parameter Estimates

Gradient

Objective

N Parameter Estimate Function

1 AGE\_H11 -0.168911 0.003983

2 AGE\_H12 3.053470 -0.000146

3 AGE\_H13 0.033667 0.004079

4 GENDERF\_H11 -0.174498 -0.002279

5 GENDERF\_H12 -0.000738 -0.000151

6 GENDERF\_H13 -0.216938 -0.003981

7 AFFL\_TILE41\_H11 0.378462 -0.001087

8 AFFL\_TILE42\_H11 0.270057 -0.001783

9 AFFL\_TILE43\_H11 -0.169633 -0.001868

10 BILL\_TILE41\_H11 -0.082297 -0.002778

11 BILL\_TILE42\_H11 0.083873 -0.000780

12 CLASSGold\_H11 -0.063234 0.003419

13 CLASSPlatinum\_H11 0.217538 0.002003

14 CLASSSilver\_H11 0.059833 0.003406

15 LTIME\_TILE41\_H11 -0.227265 -0.001414

16 LTIME\_TILE42\_H11 0.230627 -0.001084

17 LTIME\_TILE43\_H11 0.012775 -0.001671

18 NGROUPA\_H11 -0.103216 -0.000207

19 NGROUPB\_H11 -0.029423 0.000154

20 NGROUPC\_H11 -0.075512 -0.000140

21 NGROUPD\_H11 -0.078397 0.000686

22 NGROUPE\_H11 -0.112043 0.000073551

23 NGROUPF\_H11 0.127763 0.000508

24 TV\_REGBorder\_H11 -0.147950 0.002232

25 TV\_REGCScotland\_H11 0.070145 0.002109

26 TV\_REGEast\_H11 0.173725 0.002073

27 TV\_REGLondon\_H11 0.159988 0.002978

28 TV\_REGMidlands\_H11 -0.045771 0.002648

29 TV\_REGNEast\_H11 -0.067416 0.002316

30 TV\_REGNScot\_H11 -0.261108 0.002162

31 TV\_REGNWest\_H11 0.320873 0.002726

32 TV\_REGS\_SEast\_H11 0.200121 0.002353

33 TV\_REGSWest\_H11 -0.296375 0.002544

34 TV\_REGWales\_West\_H11 0.052017 0.002242

35 AFFL\_TILE41\_H12 -0.092098 0.000525

36 AFFL\_TILE42\_H12 0.181941 0.000447

37 AFFL\_TILE43\_H12 0.006818 0.000519

38 BILL\_TILE41\_H12 -0.078216 -0.000410

39 BILL\_TILE42\_H12 -0.172920 -0.000107

40 CLASSGold\_H12 -0.021922 0.000354

41 CLASSPlatinum\_H12 -0.548625 0.000348

42 CLASSSilver\_H12 0.317490 0.000322

43 LTIME\_TILE41\_H12 0.054726 -0.000253

44 LTIME\_TILE42\_H12 0.217985 -0.000082068

45 LTIME\_TILE43\_H12 -0.189995 -0.000185

46 NGROUPA\_H12 -0.108016 -0.000016396

47 NGROUPB\_H12 0.293189 -0.000110

48 NGROUPC\_H12 0.393859 0.000050697

49 NGROUPD\_H12 0.070902 -0.000195

50 NGROUPE\_H12 0.119485 0.000160

51 NGROUPF\_H12 0.029800 -0.000022524

52 TV\_REGBorder\_H12 1.064056 0.000291

53 TV\_REGCScotland\_H12 -0.271569 0.000365

54 TV\_REGEast\_H12 0.011994 0.000360

55 TV\_REGLondon\_H12 0.282237 0.000561

56 TV\_REGMidlands\_H12 0.135925 0.000305

57 TV\_REGNEast\_H12 -0.379676 0.000313

58 TV\_REGNScot\_H12 -0.544399 0.000277

59 TV\_REGNWest\_H12 0.069465 0.000411

60 TV\_REGS\_SEast\_H12 0.053536 0.000294

61 TV\_REGSWest\_H12 -0.354308 0.000355

62 TV\_REGWales\_West\_H12 -0.054269 0.000302

63 AFFL\_TILE41\_H13 0.562192 0.001937

64 AFFL\_TILE42\_H13 -0.285002 0.001819

65 AFFL\_TILE43\_H13 0.270194 0.003659

66 BILL\_TILE41\_H13 0.370487 -0.002684

67 BILL\_TILE42\_H13 -0.229541 0.000046750

68 CLASSGold\_H13 0.274331 0.002386

69 CLASSPlatinum\_H13 -0.334278 0.001753

70 CLASSSilver\_H13 -0.393871 0.002092

71 LTIME\_TILE41\_H13 0.631010 -0.000417

72 LTIME\_TILE42\_H13 -0.885621 -0.000617

73 LTIME\_TILE43\_H13 0.033280 -0.000635

74 NGROUPA\_H13 0.716257 -0.000086250

75 NGROUPB\_H13 0.421380 0.000175

76 NGROUPC\_H13 0.424997 -0.000023406

77 NGROUPD\_H13 0.518660 -0.000357

78 NGROUPE\_H13 0.383241 0.000387

79 NGROUPF\_H13 -0.164848 -0.000496

80 TV\_REGBorder\_H13 -0.088386 0.002426

81 TV\_REGCScotland\_H13 0.059916 0.002701

82 TV\_REGEast\_H13 -0.555465 0.002255

83 TV\_REGLondon\_H13 -0.624476 0.002999

84 TV\_REGMidlands\_H13 0.112844 0.002912

85 TV\_REGNEast\_H13 0.240197 0.002466

86 TV\_REGNScot\_H13 0.698431 0.002438

87 TV\_REGNWest\_H13 -0.897412 0.002536

88 TV\_REGS\_SEast\_H13 -0.245436 0.003237

89 TV\_REGSWest\_H13 0.768405 0.002458

90 TV\_REGWales\_West\_H13 -0.118899 0.002211

91 BIAS\_H11 0.131021 0.001148

92 BIAS\_H12 1.535084 -0.000309

93 BIAS\_H13 -1.176347 -0.000359

94 H11\_ORGYN1 -2.370573 -0.000657

95 H12\_ORGYN1 -1.481301 -0.000980

96 H13\_ORGYN1 -1.208266 0.000787

97 BIAS\_ORGYN1 -0.344029 0.000199

Value of Objective Function = 0.5206639912

The NEURAL Procedure

Levenberg-Marquardt Optimization

Minimum Iterations 0

Maximum Iterations 50

Maximum Function Calls 2147483647

Maximum CPU Time 14400

ABSGCONV Gradient Criterion 0.00001

GCONV Gradient Criterion 1E-8

GCONV2 Gradient Criterion 0

ABSFCONV Function Criterion 0

FCONV Function Criterion 0.0001

FCONV2 Function Criterion 0

FSIZE Parameter 0

ABSXCONV Parameter Change Criterion 0

XCONV Parameter Change Criterion 0

XSIZE Parameter 0

ABSCONV Function Criterion 0.0013863817

Trust Region Initial Radius Factor 1

Singularity Tolerance (SINGULAR) 1E-8

Levenberg-Marquardt Optimization

Scaling Update of More (1978)

Parameter Estimates 97

Optimization Start

Active Constraints 0 Objective Function 0.5206639912

Max Abs Gradient Element 0.0040791288 Radius 1

Ratio

Between

Actual

Objective Max Abs and

Function Active Objective Function Gradient Predicted

Iter Restarts Calls Constraints Function Change Element Lambda Change

1 0 5 0 0.52036 0.000305 0.00403 0.122 0.130

2 0 6 0 0.52012 0.000242 0.00290 0.0802 0.130

3 0 8 0 0.51957 0.000551 0.00186 0.477 0.367

4 0 9 0 0.51940 0.000163 0.00170 0.113 0.242

5 0 11 0 0.51920 0.000201 0.00104 0.690 0.394

6 0 12 0 0.51913 0.000076 0.000816 0.230 0.339

7 0 13 0 0.51908 0.000044 0.000722 0.211 0.221

8 0 14 0 0.51904 0.000046 0.000711 0.224 0.225

9 0 15 0 0.51901 0.000025 0.000699 0.187 0.130

10 0 16 0 0.51898 0.000036 0.000729 0.211 0.164

11 0 17 0 0.51896 0.000011 0.000732 0.181 0.0527

12 0 18 0 0.51888 0.000080 0.000441 0.967 0.434

13 0 19 0 0.51886 0.000023 0.000377 0.273 0.376

14 0 20 0 0.51884 0.000018 0.000298 0.287 0.333

15 0 21 0 0.51883 0.000017 0.000360 0.257 0.341

16 0 22 0 0.51881 0.000016 0.000292 0.262 0.329

17 0 23 0 0.51879 0.000015 0.000363 0.243 0.321

18 0 24 0 0.51878 0.000015 0.000304 0.250 0.310

19 0 25 0 0.51877 0.000014 0.000373 0.238 0.290

20 0 26 0 0.51875 0.000014 0.000324 0.247 0.286

21 0 27 0 0.51874 0.000013 0.000386 0.240 0.259

22 0 28 0 0.51873 0.000014 0.000345 0.251 0.266

Optimization Results

Iterations 22 Function Calls 30

Jacobian Calls 24 Active Constraints 0

Objective Function 0.5187250013 Max Abs Gradient Element 0.0003448865

Lambda 0.2507791654 Actual Over Pred Change 0.2659571534

Radius 0.0093053093

Convergence criterion (FCONV=0.0001) satisfied.

The NEURAL Procedure

Optimization Results

Parameter Estimates

Gradient

Objective

N Parameter Estimate Function

1 AGE\_H11 -0.122495 -0.000098930

2 AGE\_H12 3.544919 -0.000141

3 AGE\_H13 -0.138784 -0.000033130

4 GENDERF\_H11 -0.191767 0.000087584

5 GENDERF\_H12 0.002656 0.000137

6 GENDERF\_H13 -0.251618 -0.000146

7 AFFL\_TILE41\_H11 0.302486 0.000114

8 AFFL\_TILE42\_H11 0.315197 -0.000083555

9 AFFL\_TILE43\_H11 -0.124863 0.000135

10 BILL\_TILE41\_H11 -0.077221 -0.000128

11 BILL\_TILE42\_H11 0.107602 -0.000144

12 CLASSGold\_H11 -0.052118 0.000128

13 CLASSPlatinum\_H11 0.220666 0.000015729

14 CLASSSilver\_H11 -0.006131 0.000155

15 LTIME\_TILE41\_H11 -0.191765 -0.000044385

16 LTIME\_TILE42\_H11 0.176940 -0.000009774

17 LTIME\_TILE43\_H11 0.010649 -0.000124

18 NGROUPA\_H11 -0.115910 -0.000005846

19 NGROUPB\_H11 0.013777 -0.000053670

20 NGROUPC\_H11 -0.066722 0.000027051

21 NGROUPD\_H11 -0.078151 0.000054499

22 NGROUPE\_H11 -0.081343 0.000061533

23 NGROUPF\_H11 0.101144 -0.000054250

24 TV\_REGBorder\_H11 -0.262922 0.000044029

25 TV\_REGCScotland\_H11 0.259041 0.000023159

26 TV\_REGEast\_H11 0.213125 0.000023564

27 TV\_REGLondon\_H11 0.190509 0.000076505

28 TV\_REGMidlands\_H11 -0.002215 -0.000008680

29 TV\_REGNEast\_H11 -0.087223 0.000110

30 TV\_REGNScot\_H11 -0.230660 0.000021793

31 TV\_REGNWest\_H11 0.335750 0.000043357

32 TV\_REGS\_SEast\_H11 0.279685 0.000033771

33 TV\_REGSWest\_H11 -0.454613 0.000043723

34 TV\_REGWales\_West\_H11 0.076990 0.000019101

35 AFFL\_TILE41\_H12 -0.196410 0.000077138

36 AFFL\_TILE42\_H12 0.250895 0.000081457

37 AFFL\_TILE43\_H12 -0.015602 0.000233

38 BILL\_TILE41\_H12 -0.070921 -0.000002029

39 BILL\_TILE42\_H12 -0.092993 -0.000072277

40 CLASSGold\_H12 0.102917 0.000011648

41 CLASSPlatinum\_H12 -0.583739 -0.000035772

42 CLASSSilver\_H12 0.289769 0.000000958

43 LTIME\_TILE41\_H12 -0.023427 -0.000008757

44 LTIME\_TILE42\_H12 0.275956 -0.000051447

45 LTIME\_TILE43\_H12 -0.191469 0.000002320

46 NGROUPA\_H12 -0.139741 0.000021737

47 NGROUPB\_H12 0.346329 -0.000008470

48 NGROUPC\_H12 0.423951 0.000063983

49 NGROUPD\_H12 0.036767 0.000019544

50 NGROUPE\_H12 0.110658 0.000096630

51 NGROUPF\_H12 0.069554 -0.000023393

52 TV\_REGBorder\_H12 1.310480 -0.000176

53 TV\_REGCScotland\_H12 -0.354418 -0.000173

54 TV\_REGEast\_H12 -0.039396 -0.000174

55 TV\_REGLondon\_H12 0.307368 -0.000146

56 TV\_REGMidlands\_H12 0.122362 -0.000177

57 TV\_REGNEast\_H12 -0.383647 -0.000172

58 TV\_REGNScot\_H12 -0.461184 -0.000176

59 TV\_REGNWest\_H12 0.056540 -0.000168

60 TV\_REGS\_SEast\_H12 0.102648 -0.000178

61 TV\_REGSWest\_H12 -0.467412 -0.000175

62 TV\_REGWales\_West\_H12 -0.079308 -0.000180

63 AFFL\_TILE41\_H13 1.149171 0.000044575

64 AFFL\_TILE42\_H13 -0.530443 0.000017847

65 AFFL\_TILE43\_H13 0.255706 0.000044982

66 BILL\_TILE41\_H13 0.559160 -0.000207

67 BILL\_TILE42\_H13 -0.534262 0.000001852

68 CLASSGold\_H13 0.351394 0.000058120

69 CLASSPlatinum\_H13 -0.656906 0.000161

70 CLASSSilver\_H13 -0.198618 -0.000044831

71 LTIME\_TILE41\_H13 0.722723 -0.000184

72 LTIME\_TILE42\_H13 -0.811295 -0.000051417

73 LTIME\_TILE43\_H13 -0.040922 -0.000135

74 NGROUPA\_H13 1.077415 -0.000011318

75 NGROUPB\_H13 0.337945 -0.000056202

76 NGROUPC\_H13 0.549191 -0.000032893

77 NGROUPD\_H13 0.689956 -0.000067583

78 NGROUPE\_H13 0.348891 -0.000005422

79 NGROUPF\_H13 -0.220047 -0.000109

80 TV\_REGBorder\_H13 0.287310 0.000286

81 TV\_REGCScotland\_H13 -0.637484 0.000306

82 TV\_REGEast\_H13 -0.819510 0.000320

83 TV\_REGLondon\_H13 -0.952657 0.000328

84 TV\_REGMidlands\_H13 0.058448 0.000316

85 TV\_REGNEast\_H13 0.372543 0.000205

86 TV\_REGNScot\_H13 0.776058 0.000282

87 TV\_REGNWest\_H13 -1.191919 0.000283

88 TV\_REGS\_SEast\_H13 -0.659888 0.000345

89 TV\_REGSWest\_H13 1.380335 0.000276

90 TV\_REGWales\_West\_H13 -0.175146 0.000297

91 BIAS\_H11 0.103068 0.000188

92 BIAS\_H12 1.830059 0.000082458

93 BIAS\_H13 -1.479786 -0.000077737

94 H11\_ORGYN1 -2.276749 -0.000025532

95 H12\_ORGYN1 -1.413423 0.000003337

96 H13\_ORGYN1 -1.134640 -0.000153

97 BIAS\_ORGYN1 -0.355584 -0.000002766

Value of Objective Function = 0.5187250013

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_DFT\_ Total Degrees of Freedom 5994.00 .

\_DFE\_ Degrees of Freedom for Error 5897.00 .

\_DFM\_ Model Degrees of Freedom 97.00 .

\_NW\_ Number of Estimated Weights 97.00 .

\_AIC\_ Akaike's Information Criterion 6435.72 .

\_SBC\_ Schwarz's Bayesian Criterion 7085.48 .

\_ASE\_ Average Squared Error 0.17 0.19

\_MAX\_ Maximum Absolute Error 0.96 0.97

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_RASE\_ Root Average Squared Error 0.42 0.43

\_SSE\_ Sum of Squared Errors 2084.83 961.53

\_SUMW\_ Sum of Case Weights Times Freq 11988.00 5144.00

\_FPE\_ Final Prediction Error 0.18 .

\_MSE\_ Mean Squared Error 0.18 0.19

\_RFPE\_ Root Final Prediction Error 0.42 .

\_RMSE\_ Root Mean Squared Error 0.42 0.43

\_AVERR\_ Average Error Function 0.52 0.56

\_ERR\_ Error Function 6241.72 2855.24

\_MISC\_ Misclassification Rate 0.25 0.28

\_WRONG\_ Number of Wrong Classifications 1520.00 729.00

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 73.9421 76.8264 2324 38.7721

1 0 26.0579 27.5850 819 13.6637

0 1 24.5879 23.1736 701 11.6950

1 1 75.4121 72.4150 2150 35.8692

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 71.4717 72.9584 947 36.8196

1 0 28.5283 29.6703 378 14.6967

0 1 28.1476 27.0416 351 13.6470

1 1 71.8524 70.3297 896 34.8367

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

819 2324 701 2150

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

378 947 351 896

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 93.8107 1.93811 1.93811 96.0000 96.0000 300 0.95592

10 92.1283 1.90446 1.92128 94.3333 95.1667 300 0.92665

15 85.2866 1.71603 1.85287 85.0000 91.7778 300 0.87741

20 80.6704 1.66776 1.80670 82.6087 89.4912 299 0.83037

25 75.3541 1.54106 1.75354 76.3333 86.8579 300 0.75710

30 67.9953 1.31226 1.67995 65.0000 83.2129 300 0.68990

35 62.8176 1.31665 1.62818 65.2174 80.6482 299 0.65375

40 58.4444 1.27861 1.58444 63.3333 78.4821 300 0.61568

45 55.1934 1.29207 1.55193 64.0000 76.8718 300 0.55989

50 49.4106 0.97229 1.49411 48.1605 74.0073 299 0.49957

55 43.4696 0.84119 1.43470 41.6667 71.0646 300 0.45023

60 37.5093 0.72006 1.37509 35.6667 68.1123 300 0.40977

65 32.1041 0.67295 1.32104 33.3333 65.4349 300 0.36430

70 27.9350 0.73597 1.27935 36.4548 63.3699 299 0.31696

75 22.7662 0.50472 1.22766 25.0000 60.8096 300 0.26716

80 18.3703 0.52490 1.18370 26.0000 58.6322 300 0.22122

85 13.6823 0.38487 1.13682 19.0635 56.3101 299 0.18166

90 9.2318 0.33648 1.09232 16.6667 54.1057 300 0.14770

95 4.6121 0.21535 1.04612 10.6667 51.8174 300 0.11657

100 0.0000 0.12154 1.00000 6.0201 49.5329 299 0.06734

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 83.1039 1.83104 1.83104 90.6977 90.6977 129 0.95327

10 83.1039 1.83104 1.83104 90.6977 90.6977 129 0.91890

15 82.5323 1.81380 1.82532 89.8438 90.4145 128 0.87115

20 76.4033 1.58064 1.76403 78.2946 87.3786 129 0.83114

25 70.4867 1.46681 1.70487 72.6563 84.4479 128 0.76720

30 65.0113 1.37719 1.65011 68.2171 81.7358 129 0.70027

35 57.7427 1.14244 1.57743 56.5891 78.1354 129 0.66519

40 53.2277 1.21446 1.53228 60.1563 75.8989 128 0.63079

45 48.0133 1.06419 1.48013 52.7132 73.3161 129 0.57089

50 43.4851 1.02519 1.43485 50.7813 71.0731 128 0.51054

55 38.8219 0.92334 1.38822 45.7364 68.7633 129 0.45946

60 34.4149 0.86075 1.34415 42.6357 66.5803 129 0.41849

65 29.5582 0.70975 1.29558 35.1563 64.1746 128 0.37528

70 25.0985 0.67295 1.25098 33.3333 61.9656 129 0.32892

75 21.8210 0.75706 1.21821 37.5000 60.3421 128 0.27908

80 16.7355 0.40690 1.16736 20.1550 57.8231 129 0.22843

85 12.8039 0.50080 1.12804 24.8062 55.8756 129 0.18815

90 8.0493 0.26813 1.08049 13.2813 53.5205 128 0.15196

95 3.9157 0.29735 1.03916 14.7287 51.4730 129 0.12149

100 0.0000 0.25235 1.00000 12.5000 49.5334 128 0.07015

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.95-1.00 242 11 0.95733 4.22089

0.90-0.95 321 17 0.93040 5.63897

0.85-0.90 300 55 0.87500 5.92259

0.80-0.85 209 42 0.82618 4.18752

0.75-0.80 141 40 0.77523 3.01969

0.70-0.75 135 61 0.72144 3.26994

0.65-0.70 270 145 0.67310 6.92359

0.60-0.65 234 126 0.62681 6.00601

0.55-0.60 159 79 0.57691 3.97064

0.50-0.55 139 125 0.52400 4.40440

0.45-0.50 130 164 0.47290 4.90490

0.40-0.45 142 232 0.42539 6.23957

0.35-0.40 109 208 0.37528 5.28862

0.30-0.35 113 202 0.32544 5.25526

0.25-0.30 82 222 0.27411 5.07174

0.20-0.25 86 254 0.22443 5.67234

0.15-0.20 81 352 0.17441 7.22389

0.10-0.15 58 411 0.12469 7.82449

0.05-0.10 15 212 0.07548 3.78712

0.00-0.05 3 67 0.03998 1.16783

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.95-1.00 78 6 0.95740 3.26594

0.90-0.95 142 14 0.92915 6.06532

0.85-0.90 137 17 0.87311 5.98756

0.80-0.85 93 30 0.82934 4.78227

0.75-0.80 68 20 0.77775 3.42146

0.70-0.75 69 29 0.72369 3.81026

0.65-0.70 115 81 0.67389 7.62053

0.60-0.65 86 57 0.62871 5.55988

0.55-0.60 51 38 0.57402 3.46034

0.50-0.55 57 59 0.52504 4.51011

0.45-0.50 59 69 0.47410 4.97667

0.40-0.45 71 92 0.42523 6.33748

0.35-0.40 47 87 0.37636 5.20995

0.30-0.35 49 92 0.32677 5.48212

0.25-0.30 44 77 0.27553 4.70451

0.20-0.25 27 127 0.22286 5.98756

0.15-0.20 40 135 0.17264 6.80404

0.10-0.15 26 169 0.12573 7.58165

0.05-0.10 11 71 0.07674 3.18818

0.00-0.05 4 28 0.03865 1.24417

# APPENDIX O ENSEMBLE MODEL SUMMARY REPORT

Variable Summary

Measurement Frequency

Role Level Count

INPUT BINARY 1

INPUT INTERVAL 1

INPUT NOMINAL 1

REJECTED INTERVAL 5

REJECTED NOMINAL 11

REJECTED UNARY 1

TARGET BINARY 1

Model Events

Number

Measurement of

Target Event Level Levels Order Label

ORGYN 1 BINARY 2 Descending Organics Purchased?

Predicted and decision variables

Type Variable Label

TARGET ORGYN Organics Purchased?

PREDICTED P\_ORGYN1 Predicted: ORGYN=1

RESIDUAL R\_ORGYN1 Residual: ORGYN=1

PREDICTED P\_ORGYN0 Predicted: ORGYN=0

RESIDUAL R\_ORGYN0 Residual: ORGYN=0

FROM F\_ORGYN From: ORGYN

INTO I\_ORGYN Into: ORGYN

Imported Models to be Combined

Probability Function: Average

Node Id rpt\_modeltype

REG4 Regression

TREE5 Extension Class

NEURAL2 Neural

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

Fit Statistics

Target=ORGYN Target Label=Organics Purchased?

Fit

Statistics Statistics Label Train Validation

\_ASE\_ Average Squared Error 0.18 0.18

\_DIV\_ Divisor for ASE 11988.00 5144.00

\_MAX\_ Maximum Absolute Error 0.93 0.94

\_NOBS\_ Sum of Frequencies 5994.00 2572.00

\_RASE\_ Root Average Squared Error 0.42 0.43

\_SSE\_ Sum of Squared Errors 2107.57 944.13

\_DISF\_ Frequency of Classified Cases 5994.00 2572.00

\_MISC\_ Misclassification Rate 0.26 0.27

\_WRONG\_ Number of Wrong Classifications 1532.00 699.00

Classification Table

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 72.8218 78.7438 2382 39.7397

1 0 27.1782 29.9427 889 14.8315

0 1 23.6137 21.2562 643 10.7274

1 1 76.3863 70.0573 2080 34.7014

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Target Outcome Frequency Total

Target Outcome Percentage Percentage Count Percentage

0 0 71.5934 76.5023 993 38.6081

1 0 28.4066 30.9262 394 15.3188

0 1 25.7384 23.4977 305 11.8585

1 1 74.2616 69.0738 880 34.2146

Event Classification Table

Data Role=TRAIN Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

889 2382 643 2080

Data Role=VALIDATE Target=ORGYN Target Label=Organics Purchased?

False True False True

Negative Negative Positive Positive

394 993 305 880

Assessment Score Rankings

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 94.4837 1.94484 1.94484 96.3333 96.3333 300 0.92161

10 91.1189 1.87754 1.91119 93.0000 94.6667 300 0.90355

15 85.5109 1.74295 1.85511 86.3333 91.8889 300 0.84620

20 80.6704 1.66100 1.80670 82.2742 89.4912 299 0.79369

25 73.7379 1.46031 1.73738 72.3333 86.0574 300 0.73458

30 69.5664 1.48723 1.69566 73.6667 83.9911 300 0.68367

35 64.6460 1.35041 1.64646 66.8896 81.5539 299 0.65880

40 59.9598 1.27188 1.59960 63.0000 79.2327 300 0.62650

45 54.8193 1.13729 1.54819 56.3333 76.6864 300 0.55750

50 47.7265 0.83725 1.47727 41.4716 73.1732 299 0.46348

55 41.2040 0.76044 1.41204 37.6667 69.9424 300 0.41011

60 35.7694 0.76044 1.35769 37.6667 67.2505 300 0.37641

65 32.2595 0.90176 1.32260 44.6667 65.5119 300 0.34971

70 27.4058 0.64144 1.27406 31.7726 63.1077 299 0.31767

75 22.5417 0.54509 1.22542 27.0000 60.6984 300 0.27816

80 17.4442 0.41050 1.17444 20.3333 58.1735 300 0.25177

85 12.9691 0.41187 1.12969 20.4013 55.9568 299 0.22354

90 8.5208 0.32975 1.08521 16.3333 53.7535 300 0.19832

95 4.5767 0.33648 1.04577 16.6667 51.7998 300 0.17364

100 0.0000 0.12829 1.00000 6.3545 49.5329 299 0.13772

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Mean

Cumulative % Cumulative Number of Posterior

Depth Gain Lift Lift Response % Response Observations Probability

5 87.0164 1.87016 1.87016 92.6357 92.6357 129 0.91989

10 87.7989 1.88581 1.87799 93.4109 93.0233 129 0.89762

15 80.9632 1.67185 1.80963 82.8125 89.6373 128 0.83850

20 76.7954 1.64324 1.76795 81.3953 87.5728 129 0.79275

25 70.1727 1.43527 1.70173 71.0938 84.2924 128 0.74345

30 67.6263 1.54934 1.67626 76.7442 83.0311 129 0.68838

35 59.0871 1.07984 1.59087 53.4884 78.8013 129 0.65705

40 55.9744 1.34063 1.55974 66.4063 77.2595 128 0.62632

45 51.3257 1.14244 1.51326 56.5891 74.9568 129 0.56797

50 45.3689 0.91479 1.45369 45.3125 72.0062 128 0.47767

55 39.8206 0.84510 1.39821 41.8605 69.2580 129 0.41741

60 34.9379 0.81380 1.34938 40.3101 66.8394 129 0.38299

65 29.9205 0.69398 1.29920 34.3750 64.3541 128 0.35728

70 25.4348 0.67295 1.25435 33.3333 62.1321 129 0.32812

75 20.7745 0.55203 1.20774 27.3438 59.8237 128 0.29141

80 16.6375 0.54775 1.16637 27.1318 57.7745 129 0.26009

85 12.8039 0.51645 1.12804 25.5814 55.8756 129 0.22893

90 8.4853 0.34699 1.08485 17.1875 53.7365 128 0.20002

95 3.9983 0.23475 1.03998 11.6279 51.5139 129 0.17503

100 0.0000 0.23658 1.00000 11.7188 49.5334 128 0.13894

Assessment Score Distribution

Data Role=TRAIN Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 488 25 0.91604 8.55856

0.85-0.90 204 25 0.87577 3.82049

0.80-0.85 219 32 0.81992 4.18752

0.75-0.80 246 62 0.78046 5.13847

0.70-0.75 133 65 0.71938 3.30330

0.65-0.70 392 155 0.67353 9.12579

0.60-0.65 224 141 0.62855 6.08942

0.55-0.60 91 69 0.57563 2.66934

0.50-0.55 83 69 0.52587 2.53587

0.45-0.50 62 103 0.47314 2.75275

0.40-0.45 140 192 0.42321 5.53887

0.35-0.40 207 315 0.37370 8.70871

0.30-0.35 150 278 0.32815 7.14047

0.25-0.30 132 377 0.27137 8.49183

0.20-0.25 103 446 0.22368 9.15916

0.15-0.20 84 464 0.17674 9.14248

0.10-0.15 11 207 0.13094 3.63697

Data Role=VALIDATE Target Variable=ORGYN Target Label=Organics Purchased?

Posterior Number Mean

Probability of Number of Posterior

Range Events Nonevents Probability Percentage

0.90-0.95 183 14 0.91505 7.65941

0.85-0.90 95 9 0.87792 4.04355

0.80-0.85 101 23 0.81931 4.82115

0.75-0.80 116 33 0.77801 5.79316

0.70-0.75 58 28 0.72199 3.34370

0.65-0.70 133 65 0.67584 7.69829

0.60-0.65 120 70 0.62860 7.38725

0.55-0.60 40 26 0.57789 2.56610

0.50-0.55 34 37 0.52721 2.76050

0.45-0.50 40 43 0.47421 3.22706

0.40-0.45 57 83 0.42178 5.44323

0.35-0.40 90 149 0.37338 9.29238

0.30-0.35 64 122 0.32679 7.23173

0.25-0.30 53 155 0.27344 8.08709

0.20-0.25 50 160 0.22380 8.16485

0.15-0.20 29 208 0.17645 9.21462

0.10-0.15 11 73 0.12966 3.26594

# APPENDIX P MODEL COMPARISON (STEPWISE-DT GINI-NEURAL NETWORK-ENSEMBLE)

Variable Summary

Measurement Frequency

Role Level Count

TARGET BINARY 1

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.787 0.574 0.436 0.542 0.431 0.651

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.779 0.558 0.423 0.476 0.412 0.527

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.818 0.635 0.496 0.512 0.492 0.559

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.791 0.581 0.44 0.564 0.431 0.511

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.785 0.57 0.487 0.477 0.486 0.562

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.771 0.543 0.458 0.477 0.455 0.562

BINNED\_KS\_

\_KS\_PROB\_ PROB\_

Obs TARGET TARGETLABEL \_AUR\_ \_GINI\_ KS CUTOFF \_KS\_BIN\_ CUTOFF

1 ORGYN Organics Purchased? 0.814 0.627 0.489 0.515 0.489 0.556

\_VKS\_ \_VBINNED\_

PROB\_ \_VKS\_ KS\_PROB\_

Obs TARGET TARGETLABEL \_VAUR\_ \_VGINI\_ VKS CUTOFF\_ BIN\_ CUTOFF\_

1 ORGYN Organics Purchased? 0.796 0.593 0.462 0.527 0.458 0.566

Fit Statistics

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Train: Valid:

Valid: Average Train: Average

Selected Model Misclassification Squared Misclassification Squared

Model Node Model Description Rate Error Rate Error

Y Tree5 DT-GINI 0.27061 0.18116 0.25576 0.18837

Ensmbl2 Ensemble 0.27177 0.17581 0.25559 0.18354

Neural2 Neural Network 0.28344 0.17391 0.25359 0.18692

Reg4 Regression(STEPWISE) 0.29316 0.18818 0.29096 0.19184

Fit Statistics Table

Target: ORGYN

Data Role=Train

Statistics Tree5 Ensmbl2 Neural2 Reg4

Train: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff 0.56 0.56 0.56 0.65

Train: Kolmogorov-Smirnov Statistic 0.49 0.49 0.50 0.44

Train: Akaike's Information Criterion . . 6435.72 6689.62

Train: Average Squared Error 0.18 0.18 0.17 0.19

Train: Roc Index 0.79 0.81 0.82 0.79

Train: Average Error Function . . 0.52 0.56

Train: Cumulative Percent Captured Response 18.49 19.13 19.23 18.98

Train: Percent Captured Response 9.25 9.40 9.53 9.39

Selection Criterion: Valid: Misclassification Rate 0.27 0.27 0.28 0.29

Train: Degrees of Freedom for Error . . 5897.00 5988.00

Train: Model Degrees of Freedom . . 97.00 6.00

Train: Total Degrees of Freedom 5994.00 . 5994.00 5994.00

Train: Frequency of Classified Cases . 5994.00 . .

Train: Divisor for ASE 11988.00 11988.00 11988.00 11988.00

Train: Error Function . . 6241.72 6677.62

Train: Final Prediction Error . . 0.18 0.19

Train: Gain 84.76 91.12 92.13 89.57

Train: Gini Coefficient 0.57 0.63 0.64 0.57

Train: Bin-Based Two-Way Kolmogorov-Smirnov Statistic 0.49 0.49 0.49 0.43

Train: Kolmogorov-Smirnov Probability Cutoff 0.48 0.52 0.51 0.54

Train: Cumulative Lift 1.85 1.91 1.92 1.90

Train: Lift 1.85 1.88 1.90 1.88

Train: Maximum Absolute Error 0.92 0.93 0.96 0.96

Train: Misclassification Rate 0.26 0.26 0.25 0.29

Train: Mean Square Error . . 0.18 0.19

Train: Sum of Frequencies 5994.00 5994.00 5994.00 5994.00

Train: Number of Estimate Weights . . 97.00 6.00

Train: Root Average Sum of Squares 0.43 0.42 0.42 0.43

Train: Cumulative Percent Response 91.52 94.67 95.17 93.90

Train: Percent Response 91.52 93.00 94.33 92.92

Train: Root Final Prediction Error . . 0.42 0.43

Train: Root Mean Squared Error . . 0.42 0.43

Train: Schwarz's Bayesian Criterion . . 7085.48 6729.81

Train: Sum of Squared Errors 2171.69 2107.57 2084.83 2255.87

Train: Sum of Case Weights Times Freq . . 11988.00 11988.00

Train: Number of Wrong Classifications . 1532.00 1520.00 .

Data Role=Valid

Statistics Tree5 Ensmbl2 Neural2 Reg4

Valid: Kolmogorov-Smirnov Statistic 0.46 0.46 0.44 0.42

Valid: Average Squared Error 0.19 0.18 0.19 0.19

Valid: Roc Index 0.77 0.80 0.79 0.78

Valid: Average Error Function . . 0.56 0.57

Valid: Bin-Based Two-Way Kolmogorov-Smirnov Probability Cutoff 0.56 0.57 0.51 0.53

Valid: Cumulative Percent Captured Response 18.26 18.84 18.37 18.60

Valid: Percent Captured Response 9.13 9.46 9.18 8.96

Valid: Frequency of Classified Cases . 2572.00 . .

Valid: Divisor for VASE 5144.00 5144.00 5144.00 5144.00

Valid: Error Function . . 2855.24 2910.99

Valid: Gain 82.08 87.80 83.10 85.45

Valid: Gini Coefficient 0.54 0.59 0.58 0.56

Valid: Bin-Based Two-Way Kolmogorov-Smirnov Statistic 0.46 0.46 0.43 0.41

Valid: Kolmogorov-Smirnov Probability Cutoff 0.48 0.53 0.56 0.48

Valid: Cumulative Lift 1.82 1.88 1.83 1.85

Valid: Lift 1.82 1.89 1.83 1.79

Valid: Maximum Absolute Error 0.92 0.94 0.97 0.96

Valid: Misclassification Rate 0.27 0.27 0.28 0.29

Valid: Mean Square Error . . 0.19 0.19

Valid: Sum of Frequencies 2572.00 2572.00 2572.00 2572.00

Valid: Root Average Squared Error 0.43 0.43 0.43 0.44

Valid: Cumulative Percent Response 90.19 93.02 90.70 91.86

Valid: Percent Response 90.19 93.41 90.70 88.53

Valid: Root Mean Square Error . . 0.43 0.44

Valid: Sum of Square Errors 969.00 944.13 961.53 986.81

Valid: Sum of Case Weights Times Freq . . 5144.00 5144.00

Valid: Number of Wrong Classifications . 699.00 729.00 .

Event Classification Table

Model Selection based on Valid: Misclassification Rate (\_VMISC\_)

Model Data False True False True

Node Model Description Role Target Target Label Negative Negative Positive Positive

Reg4 Regression(STEPWISE) TRAIN ORGYN Organics Purchased? 876 2157 868 2093

Reg4 Regression(STEPWISE) VALIDATE ORGYN Organics Purchased? 356 900 398 918

Neural2 Neural Network TRAIN ORGYN Organics Purchased? 819 2324 701 2150

Neural2 Neural Network VALIDATE ORGYN Organics Purchased? 378 947 351 896

Tree5 DT-GINI TRAIN ORGYN Organics Purchased? 942 2434 591 2027

Tree5 DT-GINI VALIDATE ORGYN Organics Purchased? 417 1019 279 857

Ensmbl2 Ensemble TRAIN ORGYN Organics Purchased? 889 2382 643 2080

Ensmbl2 Ensemble VALIDATE ORGYN Organics Purchased? 394 993 305 880

\*------------------------------------------------------------\*

\* Score Output

\*------------------------------------------------------------\*

\*------------------------------------------------------------\*

\* Report Output

\*------------------------------------------------------------\*

# APPENDIX Q PREDICTION SCORE RESULT

| **GENDER** | **Age** | **TV Region** | **Neighborhood Group** | **Customer Loyalty Status** | **BILL\_TILE4** | **AFFL\_TILE4** | **LTIME\_TILE4** | **Into: ORGYN** | **Predicted: ORGYN=0** | **Predicted: ORGYN=1** | **Validated: ORGYN=0** | **Validated: ORGYN=1** | **Prediction for ORGYN** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| F | 32 | Wales & West | F | Silver | 1 | 4 | 2 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| M | 39 | S West | D | Silver | 2 | 4 | 1 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| F | 35 | Midlands | D | Tin | 1 | 4 | 4 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| F | 0 | London | A | Gold | 3 | 4 | 4 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| 0 | 36 | East | F | Tin | 1 | 4 | 2 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| F | 42 | London | C | Silver | 2 | 4 | 1 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| M | 44 | Midlands | F | Tin | 1 | 4 | 2 | 1 | 0.2224371 | 0.7775629 | 0.2339956 | 0.7660044 | 1 |
| F | 34 | S West | E | Tin | 1 | 3 | 2 | 1 | 0.3262712 | 0.6737288 | 0.2751323 | 0.7248677 | 1 |
| F | 0 | S & S East | F | Tin | 1 | 3 | 4 | 1 | 0.3262712 | 0.6737288 | 0.2751323 | 0.7248677 | 1 |
| F | 0 | Midlands | D | Tin | 1 | 3 | 4 | 1 | 0.3262712 | 0.6737288 | 0.2751323 | 0.7248677 | 1 |
| 0 | 0 | London | C | Silver | 2 | 1 | 1 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 31 | London | C | Tin | 1 | 1 | 4 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 30 | London | F | Silver | 1 | 2 | 1 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 36 | London | F | Tin | 1 | 1 | 2 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| 0 | 0 | Wales & West | C | Gold | 3 | 1 | 4 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 35 | S & S East | B | Silver | 2 | 2 | 4 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 0 | N East | B | Tin | 1 | 1 | 4 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 39 | Ulster |  | Tin | 1 | 2 | 3 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 0 | Midlands | E | Silver | 1 | 2 | 1 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 0 | C Scotland | F | Silver | 2 | 2 | 4 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| 0 | 0 | S & S East | A | Silver | 1 | 2 | 2 | 0 | 0.5098039 | 0.4901961 | 0.5477707 | 0.4522293 | 0 |
| F | 55 | London | D | Silver | 1 | 4 | 4 | 0 | 0.5538222 | 0.4461778 | 0.5525424 | 0.4474576 | 0 |
| 0 | 51 | Yorkshire | C | Silver | 2 | 4 | 1 | 0 | 0.5538222 | 0.4461778 | 0.5525424 | 0.4474576 | 0 |
| 0 | 50 | S & S East | B | Gold | 3 | 4 | 3 | 0 | 0.5538222 | 0.4461778 | 0.5525424 | 0.4474576 | 0 |
| F | 53 | N West | B | Gold | 3 | 4 | 4 | 0 | 0.5538222 | 0.4461778 | 0.5525424 | 0.4474576 | 0 |
| 0 | 47 | N West | E | Silver | 1 | 4 | 4 | 0 | 0.5538222 | 0.4461778 | 0.5525424 | 0.4474576 | 0 |
| F | 51 | N West | F | Gold | 3 | 4 | 1 | 0 | 0.5538222 | 0.4461778 | 0.5525424 | 0.4474576 | 0 |
| F | 40 | Yorkshire | F | Tin | 1 | 3 | 2 | 0 | 0.6051081 | 0.3948919 | 0.6009174 | 0.3990826 | 0 |
| 0 | 43 | London | F | Gold | 3 | 2 | 1 | 0 | 0.6051081 | 0.3948919 | 0.6009174 | 0.3990826 | 0 |
| F | 41 | Midlands | F | Silver | 1 | 2 | 1 | 0 | 0.6051081 | 0.3948919 | 0.6009174 | 0.3990826 | 0 |
| F | 44 | S West | F | Silver | 1 | 2 | 4 | 0 | 0.6051081 | 0.3948919 | 0.6009174 | 0.3990826 | 0 |
| F | 43 | London | F | Platinum | 3 | 2 | 4 | 0 | 0.6051081 | 0.3948919 | 0.6009174 | 0.3990826 | 0 |
| F | 65 | S West | E | Silver | 1 | 4 | 1 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| F | 65 | Midlands | C | Tin | 1 | 4 | 4 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| F | 63 | London | F | Silver | 2 | 4 | 3 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| 0 | 60 | Wales & West | C | Silver | 1 | 4 | 3 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| F | 72 | London | B | Gold | 3 | 4 | 3 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| F | 76 | East | C | Gold | 3 | 4 | 1 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| F | 69 | East | B | Gold | 3 | 4 | 4 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| 0 | 70 | Midlands | A | Gold | 3 | 4 | 2 | 0 | 0.6426056 | 0.3573944 | 0.6022727 | 0.3977273 | 0 |
| M | 0 | Midlands | C | Gold | 3 | 2 | 1 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 27 | London | D | Silver | 2 | 2 | 2 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 43 | London | A | Gold | 3 | 1 | 3 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 35 | Wales & West | E | Tin | 1 | 1 | 2 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 27 | London | C | Tin | 1 | 3 | 2 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 40 | S & S East | E | Tin | 1 | 3 | 2 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 36 |  | A | Tin | 1 | 1 | 2 | 0 | 0.7110694 | 0.2889306 | 0.6982759 | 0.3017241 | 0 |
| M | 53 | London | D | Silver | 1 | 4 | 4 | 0 | 0.7820738 | 0.2179262 | 0.8464912 | 0.1535088 | 0 |
| M | 55 | N West | B | Silver | 2 | 4 | 1 | 0 | 0.7820738 | 0.2179262 | 0.8464912 | 0.1535088 | 0 |
| F | 73 | N West | C | Gold | 3 | 3 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 51 | East | D | Tin | 1 | 2 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 76 |  | B | Silver | 2 | 3 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 65 | London | F | Gold | 3 | 3 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 50 | London | B | Silver | 1 | 2 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 79 | East | C | Gold | 3 | 3 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 69 | Wales & West | C | Silver | 2 | 3 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 54 | S West | C | Tin | 1 | 1 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 77 | London | D | Gold | 3 | . | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 77 | Wales & West | F | Gold | 3 | 3 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 65 | S & S East | C | Gold | 3 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 73 | S & S East | C | Silver | 1 | 1 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 74 | London | D | Silver | 2 | 3 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 76 | Wales & West | C | Tin | 1 | 3 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 61 | London | D | Tin | 1 | 3 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 47 | London | F | Silver | 1 | 3 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 57 | S West | D | Tin | 1 | 1 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 46 | Midlands | D | Tin | 1 | . | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 57 | Wales & West |  | Silver | 2 | 1 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 46 | Border | B | Gold | 3 | 2 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 55 | C Scotland | F | Silver | 2 | 3 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 60 | London | B | Gold | 3 | 1 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 76 | S West | B | Silver | 1 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 51 | London | C | Silver | 1 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 60 | Midlands | D | Gold | 3 | 3 | . | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 65 | London | B | Silver | 2 | 1 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 50 | Midlands | C | Platinum | 3 | 3 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 66 | London | D | Silver | 2 | 3 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 62 | S & S East | A | Silver | 1 | 2 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 68 | Yorkshire | D | Gold | 3 | 1 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 77 | S & S East | B | Gold | 3 | 2 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 46 | London | C | Silver | 1 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 71 | London | E | Gold | 2 | 1 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 64 | London | C | Tin | 1 | 2 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 79 | S & S East | C | Silver | 2 | 2 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 47 | C Scotland | F | Tin | 1 | 3 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 54 | London | E | Silver | 1 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 70 | C Scotland | D | Tin | 1 | 2 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 54 | Midlands | F | Silver | 2 | 2 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 51 | London | E | Silver | 1 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 73 | Midlands | A | Silver | 2 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 70 | Wales & West | F | Silver | 1 | 1 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 77 | London | B | Gold | 3 | 1 | 1 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 46 | N East | D | Tin | 1 | 2 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 73 | N West | D | Platinum | 3 | 3 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 47 | London | A | Gold | 3 | 1 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 53 | London | B | Gold | 3 | 3 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 74 | London | B | Silver | 2 | 2 | 2 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| 0 | 70 | Midlands | C | Silver | 1 | 3 | 3 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| M | 57 | East | F | Tin | 1 | 3 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |
| F | 65 | S & S East | D | Tin | 1 | 3 | 4 | 0 | 0.8802291 | 0.1197709 | 0.8836066 | 0.1163934 | 0 |