**Team 4 Whitepaper:**

E-Commerce Online Shopping Intention Analysis

By:

Benjamin Hood, Olusola Adesegun, Benjamin Ciotti

Prepared for Pankaj Prakash

University of Connecticut

Table of Contents

Title Page…………………………………………………………………………...……………..1

Table of Contents…………..……..………………………………………...………………..........2

Executive Summary……………..……………………………………………….……….……….3

Business Understanding……………….………………………………………………….…….3-4

Methodology……………..…………………………………………………………………….5-11

Evaluation……………..……………………………………………………………………...12-15

Deployment……………..………………………………………………………..........................16

Appendix……………..……………………………………………………………………....17-25

References……………..……………………………………………………...………………….26

**Executive Summary:**

We have acquired Amazon’s online shoppers data. We will be analyzing the acquired data to find trends and correlations. The purpose of this analysis is to establish which attributes of an online browsing session on the e-commerce site Amazon will most heavily impact that session ending with a purchase. We will be comparing the browsing experience by using that data to analyze variables like the amount of time spent on a given web page, how the customer was directed towards the page, whether or not the purchase or visit date was close to a holiday, and how many times that customer had visited the page previously. By using this data, we are confident that we will be able to better tailor the customer experience to increase sales across Amazon’s product offerings. The analysis and recommendations that will be generated by the accurate prediction of the desired outcomes, will be invaluable to the company allowing Amazon to specifically target and leverage customers and their platform increasing the potential of a guaranteed sale.

**Business Understanding:**

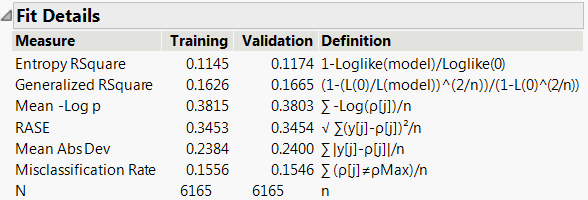
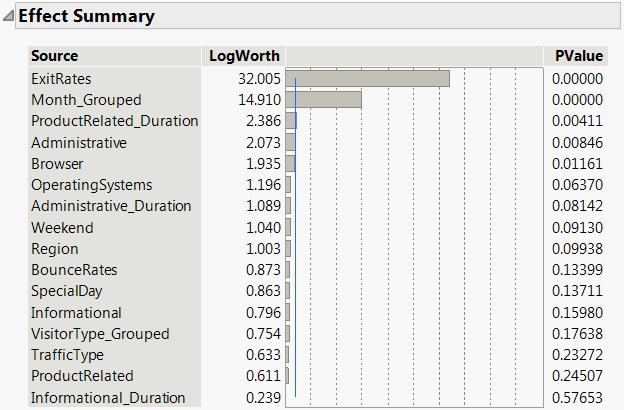
For an international conglomerate such as Amazon, turning page visits and mouse-clicks into purchases is a vital business process. From the layout of the website’s user interface, to the selection of items to be placed on sale, honing in and fine-tuning the shopping experience to nudge the customer to purchase an item in their shopping cart or wish list is an indelible aspect of the constant need to adapt and improve. While most e-commerce businesses are able to focus on building customer conversion funnels via email or other methods of filtering customers who are unlikely to purchase anything from those that will. Amazon, being the size that it is with the large customer base that it has, does not have those tools at their disposal, they must rely on website design as well as syncing discounts and advertisements with various holidays that consumers correlate to gift or other retail purchases.

Due to these facets of the business environment that Amazon is facing, we believe that the variables (located in Appendix 1) that we have targeted for data collection and tracking will reveal which area of website user interface optimization should be focused upon moving forward to provide more purchase conversions from page views. Research has shown that time spent within the browsing session is strongly correlated to purchasing, but we have broken up further the types of websites the user has been trafficked through (administrative, informational, and product details) and where the traffic was directed from - either the user specifically started the browsing session within the system or was redirected from another source. Lastly, we wanted to see whether or not the visit was near the holidays, on the weekend versus during the week, as well as determining if this was the customers first unique visit or someone returning to an unfinished shopping experience. Determining these predictive properties can assist Amazon in establishing a proper strategic goal for the future in order to increase the amount of browsing sessions that will end in a purchase.

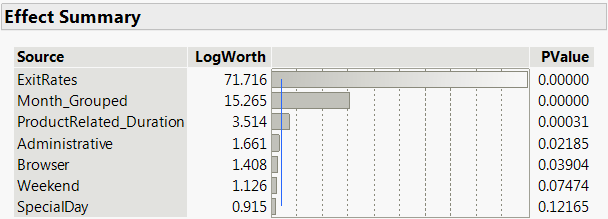
**Methodology of Analysis**

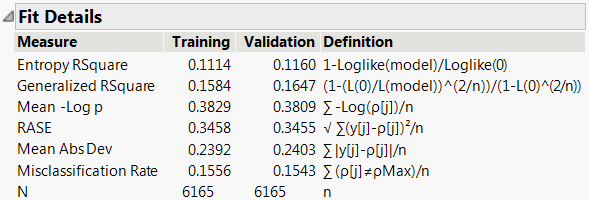
To begin the analysis, first we need to preprocess the dataset. That required removing the attribute ‘PageValues’ due to the information this attribute is tracking not being available for predicting an online browsing session purchase. It would only be available once the purchase had been completed, which would then trigger the system to assign a higher page value to the previously visited page prior to the purchase being completed. This kind of variable could introduce target leakage into the initial models. Secondly, since Amazon metrics are consistently measured by quarterly sales and projections, the months have been grouped according to their corresponding fiscal quarter for the year. Lastly, for the data preprocessing, VisitorType had three values that could be populated within the dataset: new; returning; and other. The *other* population was quite small comparatively to the other two present values, so this was grouped together with the *returning customers* nomenclature.

With the dataset being a categorical response, the first method of modeling this data focused on is the Nominal Logistic Fit for Revenue where the value for Revenue is TRUE, indicating a purchase was completed at the end of the browsing session. Running the initial model with all reasonable predictive attributes resulted in these preliminary results:

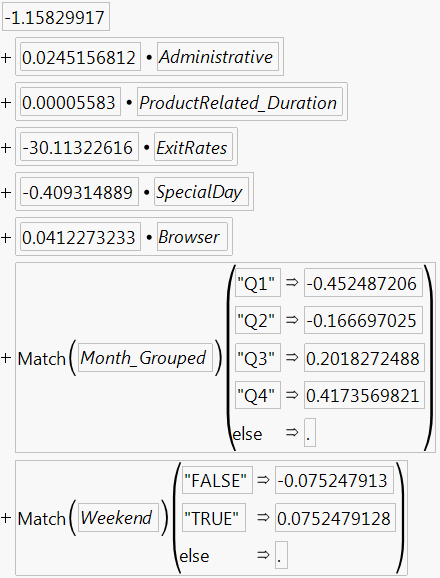
  
*Figure 1: M1 – Nominal Logistic Effect Summary & Fit Details*

Our fledgling model started with an AUC of 0.7340 (See A1 FNL\_M1) and RASE of 0.3453, so we are confident that we are capturing enough of the information with a low amount of error in our model, so we can continue by trimming statistically insignificant attributes from the model to highlight our best predictors. Refer to Appendix 1 for the order and reasoning for removing the attributes to achieve the final model, but after removing statistically insignificant variables and attributes, the remaining predictors are as follows: ExitRates, Month\_Grouped, ProductRelated\_Duration, Administrative, Browser, Weekend, and SpecialDay.

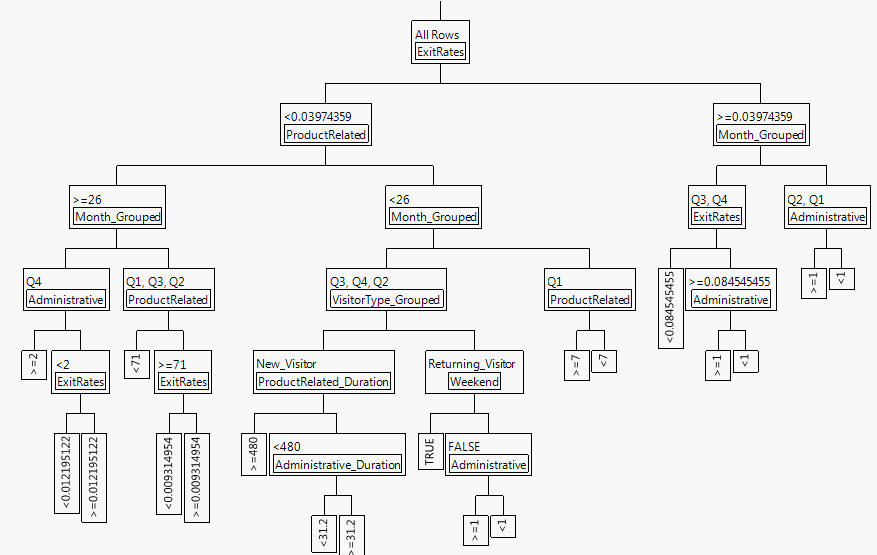


  
*Figure 2: M2 – Nominal Logistic Effect Summary & Fit Details*

As we can see, ExitRates, defined by Google Analytics as the percentage of all pageviews to the page that were the last page in the session, is currently the strongest predictor of the attributes remaining in the final model. This is intuitive initially, as ExitRates compared to BounceRates are far more valuable due to the nature of multiple pages viewed within the browsing session as BounceRates are singular page views, with no redirection within the website. It is surprising to see that SpecialDay is not as strong of a predictor as the fiscal quarter, although Q4 is the much stronger of a predictor, which includes several “SpecialDays,” so this information is most likely being captured within the Month\_Grouped attribute. The aspect of whether or not it is the weekend is also shown to be a large contributor to the probability formula, which has been calculated to be:

  
*Fig. 3: Final Nominal Logistic Probability Formula*

Next in the analysis, a decision tree, bootstrap forest, and boosted tree modeling strategy was used in order to determine which should be utilized in the final model comparison. The initial partition for Revenue using the Decision Tree yielded 18 splits, starting with ExitRates:



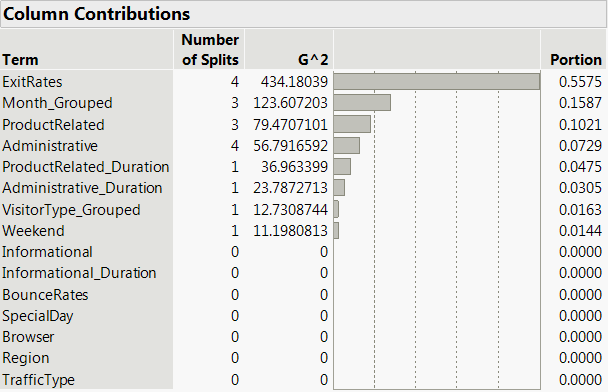


Fig 3: Small Tree View Splits & Column Contributions

Our initial RSquare value for the validation data is 0.111, indicating we have a very small amount of standard error in the model which is encouraging that this could be a very useful model for analyzing our strategic future. After running through 8 prunes and splits, respectively, for marginal gains on decreasing our RSquare (-0.01 for splitting further, -0.008 for pruning; see Appendix 2 for exact values based on splits) values we would also see a decrease in the AUC of the validation data in the model. This confirmed that the initial split of 18 branches was the correct amount to use as well as the fact that ExitRates and our grouped months attribute are still strong predictors of purchases made by customers.

Comparing the three decision tree based models, our Bootstrap Forest model returns the most favorable results with an AUC of 0.7564 and RSquare of 0.3411. However, the differences between the largest and smallest values of AUC and RSquare between the three models are 4 and 2%, respectively, so we have confidence that all of these models are providing accurate and valuable information. Also to consider is the Bootstrap Forest performing the best when calculating the precision of the model via the confusion matrix generated during the analysis:

|  |  |  |
| --- | --- | --- |
| **Bootstrap Forest** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5185 | 26 |
| **TRUE** | 931 | 23 |
|  |  |  |
| Precision | Recall | Accuracy |
| 0.847776324 | 0.995011 | 0.844768856 |

[insert app & table #]

So we know that the model is correct almost 85% of the time when it is predicting the event of a purchase to be true.

Next we ran the dataset through three penalized regression models: logistic, lasso, and ridge. The primary observation we had when running these three models was that within the lasso penalized regression, BounceRate had been zeroed out in the process, which indicated that we could re-run the models excluding this attribute for possibly better results. Removing it didn’t incur any perceptible benefits with regards to RMSE or AUC, but not including BounceRate clearly wasn’t going to negatively affect the viability of the outputs.

When comparing the models on the basis of RMSE and AUC, the lasso penalized regression is our clear ‘best model’ out of the three, but the difference between both values is lower than 1%. Similarly, the differences between the accuracy, recall, and precision are negligible indicating that all three models perform adequately at predicting a purchasing event in the dataset. Furthermore, inspecting the odds ratios of these models continues to validate that the Month\_Grouped, ExitRates, and the Weekend attributes are still the strongest predictive variables within the dataset. However, we see in this model that the probability of a purchase occurring when a new and returning customer increases by almost 13%. Even though the values are very close in range, we will be choosing the lasso penalized regression model of the three discussed for further analysis in our final models comparisons.

The last model we chose to utilize for this analysis would be designing a neural model. Running three simple models of each neural node type, we reduced the amount of models to compare to the following: 3 nodes using the sigmoid activation function (Neural - Revenue), a 6 node using the radial activation function (Neural - Revenue 1), a model with 2 layers having 3 nodes each using the identity activation function (Neural - Revenue 2), and lastly a model using 2 layers with 6 nodes each in all three of the activation functions (Neural - Revenue 3). Comparing these models surprisingly yielded the results that the most simple neural model we ran turned out to be the model capturing the most data as well as minimizing the standard error within the model.

Now, analyzing the predictors that the neural model will deem stronger contributors to the prediction formula, the data shows that returning customers, shopping later in the year on a weekend tend to be the strongest indicators of whether or not a purchase will be completed during the browsing session. These findings trend with the previous models run throughout this analysis, so it continues to strengthen the theory that these prediction attributes are being correctly identified within our models. These models also spotlight that the longer the amount of time that a potential customer spends on a given page, whether it be informational, administrative, or related to the product, that there is a downward pressure on the probability that they will end up purchasing the product they may be researching. This type of behavior may indicate that multiple, but quick, page visits are a good indication of a purchase, and that this is a facet of our future planning strategy to focus on improving upon.

Reaching our final comparison we are left with the following best models to compare out of our analysis thus far: the 3 node sigmoid neural model, our best fit nominal logistic, the bootstrap forest model, and the lasso penalized regression model. These four models up until this point in the analysis of this particular dataset have yielded us our best results for AUC, RMSE, and prediction attributes. Most of the models corroborated that the strongest predictors of the dataset are ExitRates, Weekend, & Month\_Grouped, but each gave a unique spin on the ‘least’ predictive attribute or jockeyed the attributes for positioning after the three strongest variables. After all the extensive research and development of these models, in conclusion, we have determined that the bootstrap forest model will be our best predictive model moving forward as it gives us the best AUC, RMSE, and precision of all the models, our foremost data metrics that indicate performance.

**Evaluation**

With that we have established the model with which we are most confident in continuing our analysis, what does this entirely mean? First, we have a clear understanding of the window of time of when Amazon is most likely going to make a sale (Q3 or Q4, on a weekend), however, we now can focus our research and development into the following areas: Exit and bounce rates, duration of time on a products page directly or on a related page with product information, and how to convert those new customers into becoming returning customers, even if their first browsing session does not culminate with a purchase. Second, even if an attribute is a ‘weak’ predictor, we now can study how to improve the prediction rate of some of the smaller contributors in the model such as non-end-of-year holiday shopping, minimizing the amount of time that a customer may spend on a non-product related page, and creating a landing page sticky enough to convert ‘bouncing’ customers into multi-page-viewing ones, or simply by designing the browsing experience so that we can start converting new visitors into purchasing customers without the need for them to exit and return to the page. By identifying areas of weakness that Amazon can improve upon within their user experience, we can formulate strategies that involve strengthening those areas which can do a greater job of raising our average purchase rate percentage than improving our already strong predictors.

While there were no wildcard or curveball attributes making a surprising leap to the top of the pile for strongest predictor, the benefits of either confirming (potentially) previously known information is best for prediction modeling or revealing that a discounted attribute within the data organization may actually be a stronger predictor than previously thought, is intangible. But, any amount of customers who are now purchasing items instead of simply not purchasing or – possibly worse – going to a competing online marketplace for their goods and/or services will increase profitability in the business. Even with the continued collection of data, we can narrow our scope to better identify areas of improvement.

For this analysis, we have concluded that the best ways of implementing the knowledge from this data analysis would be:

1. Increase customer engagement around non-traditional retail holidays.
2. Work to increase compatibility across browser systems, as there is downward pressure on the probability of a purchase depending on which browser is being utilized (ahem: Internet Explorer…)
3. Create a singular landing page for new customers who are redirected to the website, welcoming the customer, while also requiring a click to continue to the original page of interest.

Our data confirms the commonly held notion that during the end of the year timeframe, people are purchasing more consumer goods. Be it for gift giving, food purchases, or home goods supplies, there tends to be a spike in spending towards the end, holiday time of the year. By either designing sales to occur or recoloring the product related pages to reflect the appropriate ‘holiday’ or weather season, could induce gains in purchasing customers. This type of strategy would help increase the values of strong contributors (product related pages) as well as possibly help change weak predictors into stronger ones by boosting the metrics value indirectly (new customers might enjoy non-uniform coloration; converting bouncing, new customers to multi-page viewing or returning customers).

Similarly, by ensuring that we have seamless compatibility across browsers (or by not allowing our service to be displayed in certain browser environments) should be of similar priority to Amazon moving forward. The browsing experience of every customer should be unique to them, but identical across all platforms with which the customer may be using to peruse our wares. Amazon can work to nudge customers to preferred platforms by not allowing their website to be seen on platforms deemed obsolete or out-of-style to the current, better performing platforms, or by investing in making the browsing session equally as pleasant in all browsing environments. If customers are allowed to traffic Amazon on a browser that is not optimized (or properly accounted for backwards compatibility) for browsing, this could turn new customers away and induce more bounces into our systems, but also could repel returning customers from continuing a previously ended browsing session. Both of these are the poorest outcomes we could have as a business, so we do not want to be the root cause of any of these customers being lost for good.

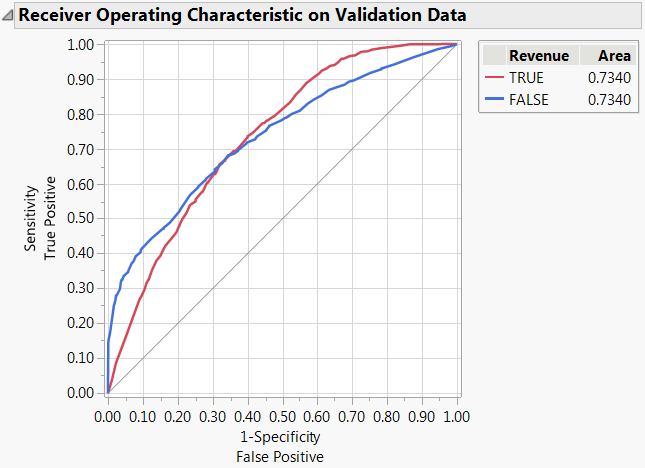
Finally, by creating a landing page for new customers that would require a click to continue their journey to the page they had been directed to could lead to a large conversion of new customers bouncing out of our marketplace to becoming returning customers, or at minimum having the potential customer view several pages (specifically product related and informational pages) before leaving our website. Through our models, we have seen time and time again that the longer the amount of time, as well as the amount of clicks intra-site, a potential customer spends on a product related or informational page in our website increases the probability of a purchase. By creating a landing page, we can optimize the customers initial impression of the website with a seasonally themed layout, discount offer, or other piece of marketing material, and then capitalize on the multi-click adventure they are hopefully about to have within our marketplace.

**Deployment**

Obviously, the more data we are able to track and establish metrics for, the better we will become at predicting purchases from our customers. However, we must be careful not to dive too deep into what metrics we choose to track and how detailed the recording of the data tracked will be. Keylogging and audio/visual are obviously ethically, and potentially legally, off the table for gathering data. Current methods seem to have a good balance between too detailed and vagueness of the variables that we could get a good predictive model, while not also divining the personal secrets of the customer that they might not want Amazon to know. Similarly, we should focus first and foremost on converting new customers to becoming returning customers. Once a customer has had a positive purchasing experience with us, they are likely to continue shopping utilizing our marketplace for their purchases. We also do not want to become overly burdensome on a customer who is already paying for our goods with more and more nudges to purchase goods, because if we become too much of a financial albatross on their finances, we may lose a customer that will never return. New customers are much more difficult to convert to returning customers, not even considering a purchasing customer. However, we must still be delicate in our handling of new customers, if we are too aggressive in our improvements to push page visitors to purchasing customers, we run the risk of losing a customer for good instead of piquing their curiosity enough to entice them with a return visit.

**Appendix:**

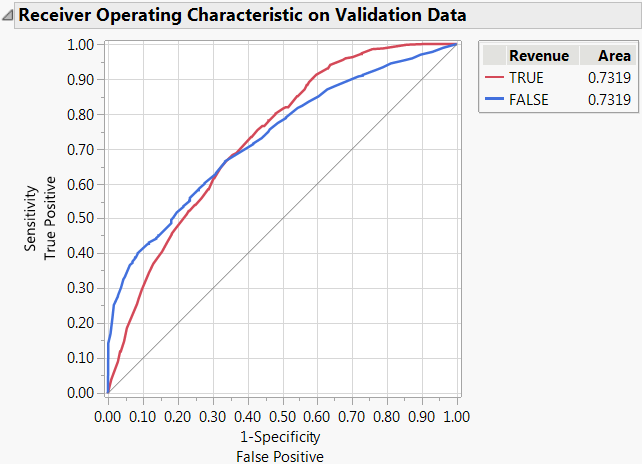
*Appendix 1 – Fit Nominal Logistic Model*



*A1 Model 1 – Initial FNL ROC Curve – Validation*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **attribute removed** | **RASE** | **Delta RASE** | **AUC** | **Delta AUC** |
| - | 0.3454 | 0.0000 | 0.7340 | 0 |
| Information\_Duration | 0.3453 | 0.0001 | 0.7342 | -0.0002 |
| Product\_Related | 0.3452 | 0.0001 | 0.7344 | -0.0002 |
| TrafficType | 0.3453 | -0.0001 | 0.7341 | 0.0003 |
| Informational | 0.3453 | 0.0000 | 0.7333 | 0.0008 |
| VisitorType\_Grouped | 0.3458 | -0.0005 | 0.7303 | 0.0030 |
| Admin\_Duration | 0.3456 | 0.0002 | 0.7308 | -0.0005 |
| BounceRates | 0.3455 | 0.0001 | 0.7321 | -0.0013 |
| Region | 0.3454 | 0.0001 | 0.7324 | -0.0003 |
| OpeartingSystem | 0.3455 | -0.0001 | 0.7319 | 0.0005 |

*A1 Table 1: Order of Removals and Associated Deltas*



*A1 Figure 2: Final ‘Best’ Model ROC Curve - Validation*

Appendix 2 Decision Tree, Boosted Tree, Bootstrap Forest

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Splits** | **Rsquare** | **Delta RSquare** | **AUC** | **Delta AUC** |
| 18 | 0.111 | 0 | 0.7265 | 0.0000 |
| 19 | 0.108 | -0.003 | 0.7228 | -0.0037 |
| 20 | 0.108 | 0 | 0.7223 | -0.0005 |
| 21 | 0.103 | -0.005 | 0.7208 | -0.0015 |
| 22 | 0.101 | -0.002 | 0.7194 | -0.0014 |
| 23 | 0.102 | 0.001 | 0.7208 | 0.0014 |
| 24 | 0.101 | -0.001 | 0.7207 | -0.0001 |
| 25 | 0.103 | 0.002 | 0.7218 | 0.0011 |
| 26 | 0.101 | -0.002 | 0.7209 | -0.0009 |
| Totals: | -0.010 |  | -0.0056 |  |

*A2 Table 1: Order of Tree Splitting & Associated Deltas*

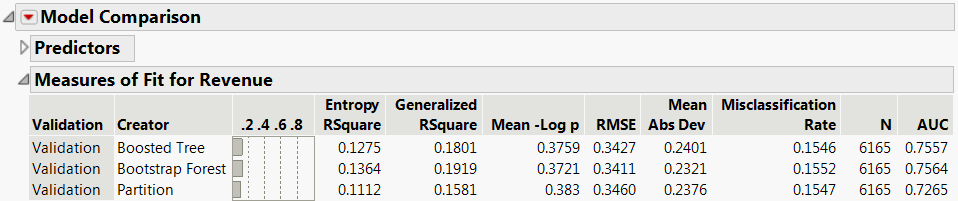
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Prunes** | **Rsquare** | **Delta RSquare** | **AUC** | **Delta AUC** |
| 18 | 0.111 | 0.000 | 0.7265 | 0.0000 |
| 17 | 0.109 | -0.002 | 0.7257 | -0.0008 |
| 16 | 0.107 | -0.002 | 0.7242 | -0.0015 |
| 15 | 0.106 | -0.001 | 0.7208 | -0.0034 |
| 14 | 0.108 | 0.002 | 0.7213 | 0.0005 |
| 13 | 0.106 | -0.002 | 0.7192 | -0.0021 |
| 12 | 0.104 | -0.002 | 0.7160 | -0.0032 |
| 11 | 0.106 | 0.002 | 0.7153 | -0.0007 |
| 10 | 0.103 | -0.003 | 0.7144 | -0.0009 |
| Totals: | -0.008 |  | -0.0121 |  |

*A2 Table 2: Order of Tree Pruning & Associated Deltas*

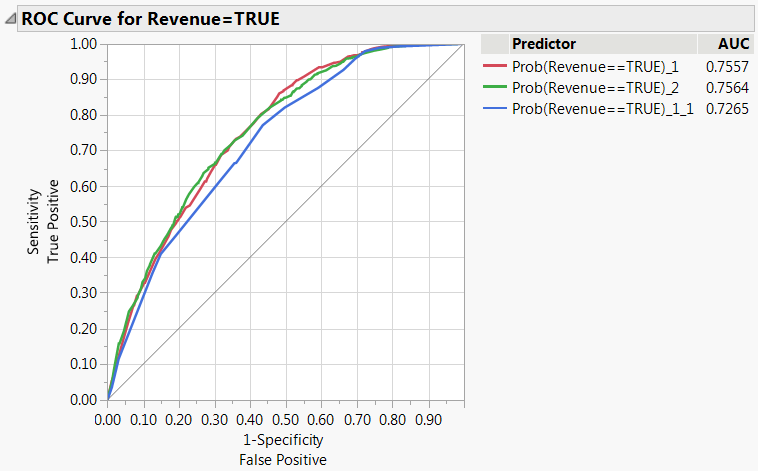
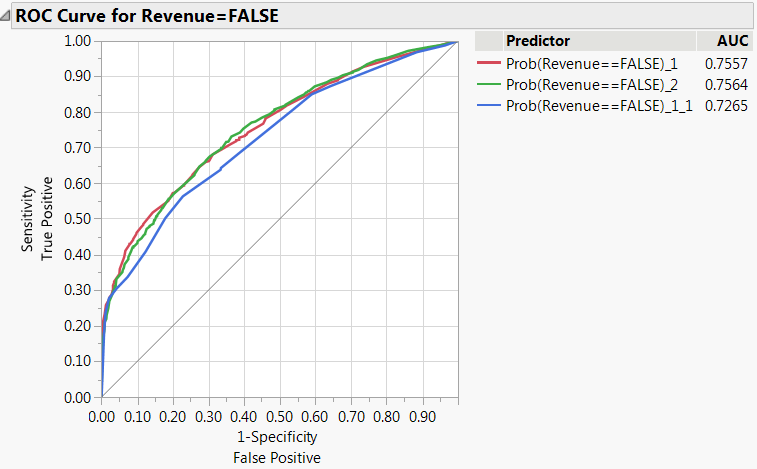
|  |  |  |
| --- | --- | --- |
| RSquare Values |  |  |
| Boosted Tree | 0.3427 |  |
| **Bootstrap Forest** | **0.3411** | % Diff |
| Partition | 0.346 | 1.42% |

|  |  |  |
| --- | --- | --- |
| AUC Values |  |  |
| Boosted Tree | 0.7557 |  |
| **Bootstrap Forest** | **0.7564** | % Diff |
| Partition | 0.7265 | 3.95% |

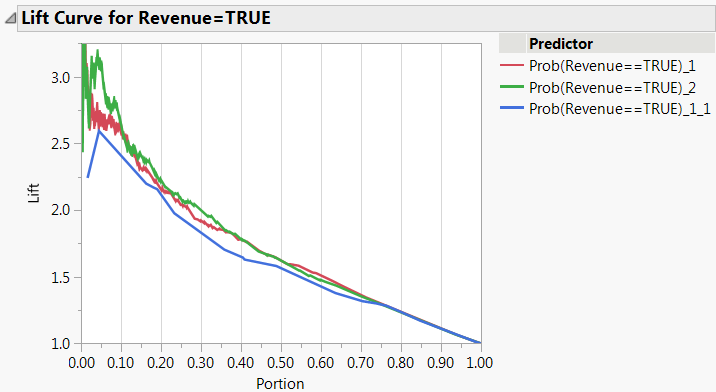
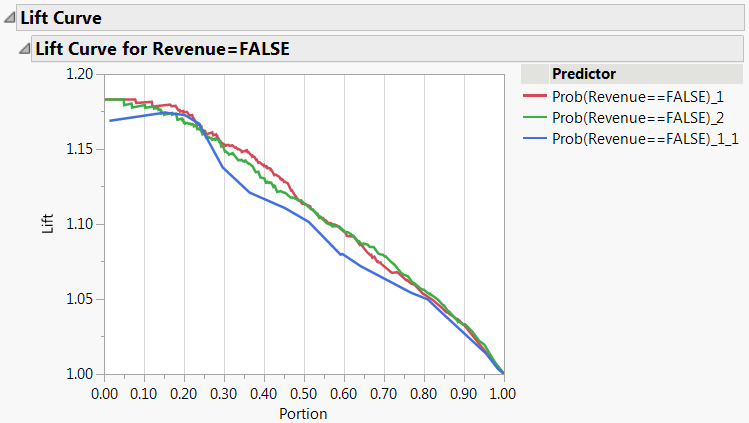
*A2 Table 3: RSquare and AUC Values for Ensemble Models*



*A2 Figure 1: Ensemble Model’s Measures of Fit Compairson*



*A2 Figure 2: Ensemble Model’s ROC Curve Comparison - Revenue = True & False*



*A2 Figure 3: Ensemble Model’s Lift Curve Comparison - Revenue = True & False*

|  |  |  |
| --- | --- | --- |
| **Boosted Tree** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5210 | 1 |
| **TRUE** | 952 | 2 |
|  |  |  |
| Precision | Recall | Accuracy |
| 0.845504706 | 0.999808 | 0.84541768 |

*A2 Table 1: Boosted Tree Confusion Matrix & Values*

|  |  |  |
| --- | --- | --- |
| **Bootstrap Forest** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5185 | 26 |
| **TRUE** | 931 | 23 |
|  |  |  |
| Precision | Recall | Accuracy |
| 0.847776324 | 0.995011 | 0.844768856 |

*A2 Table 2: Bootstrap Forest Confusion Matrix & Values*

|  |  |  |
| --- | --- | --- |
| **Decision Tree** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5211 | 0 |
| **TRUE** | 954 | 0 |
|  |  |  |
| Precision | Recall | Accuracy |
| 0.845255474 | 1 | 0.845255474 |

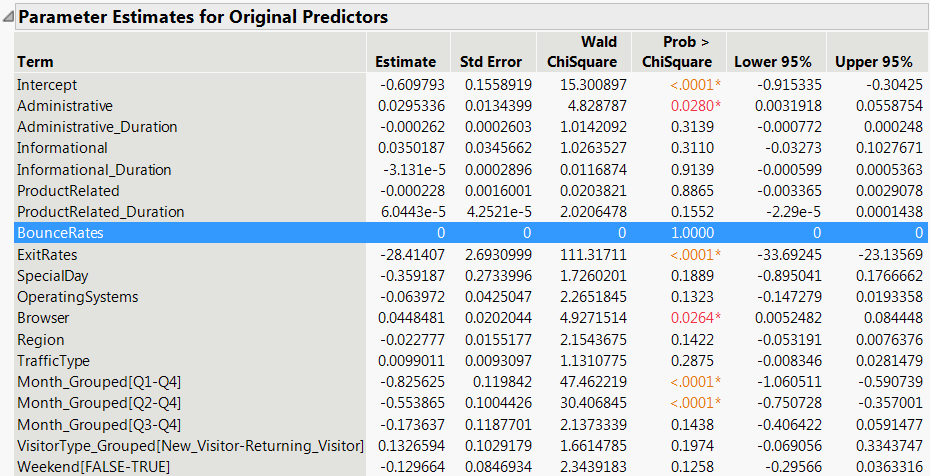
*A2 Table 2: Decision Tree Confusion Matrix & Values*

Appendix 3 Penalized Regression Analysis

|  |  |  |
| --- | --- | --- |
| RMSE Values |  |  |
| Logistic Regression | 0.3454 |  |
| Ridge | 0.3453 | % Diff |
| **Lasso** | **0.345** | 0.12% |

|  |  |  |
| --- | --- | --- |
| AUC Values |  |  |
| Logistic Regression | 0.734 |  |
| Ridge | 0.7347 | % Diff |
| **Lasso** | **0.7358** | 0.24% |

*A3 Table 1: RSquare and AUC Values for Penalized Regression Models*



*A3 Figure 1: Parameter Estimates for Penalized Regression Models*

|  |  |  |
| --- | --- | --- |
| **Generalized Logistic** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5208 | 3 |
| **TRUE** | 950 | 4 |
|  |  |  |
| Precision | Recall | Accuracy |
| **0.84573** | 0.99942 | 0.845418 |

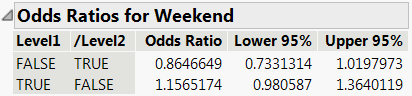
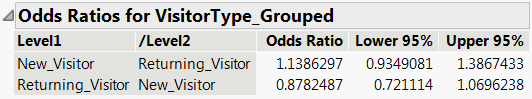
*A3 Table 2: Penalized Regression Confusion Matrix & Associated Values*

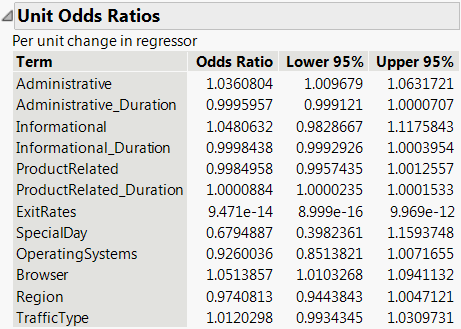
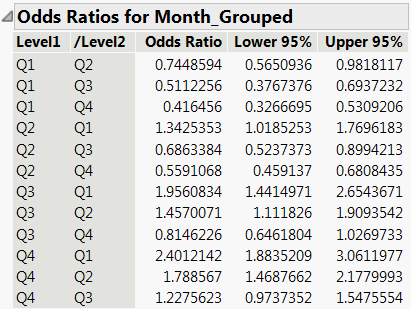
|  |  |  |
| --- | --- | --- |
| **Ridge** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5208 | 3 |
| **TRUE** | 950 | 4 |
|  |  |  |
| Precision | Recall | Accuracy |
| **0.84573** | 0.99942 | 0.845418 |

*A3 Table 3: Penalized Regression Confusion Matrix & Associated Values*

|  |  |  |
| --- | --- | --- |
| **Lasso** | Validation Data | |
| Actual | Predicted | |
| Revenue | **FALSE** | **TRUE** |
| **FALSE** | 5209 | 2 |
| **TRUE** | 951 | 3 |
|  |  |  |
| Precision | Recall | Accuracy |
| 0.84562 | **0.99962** | 0.845418 |

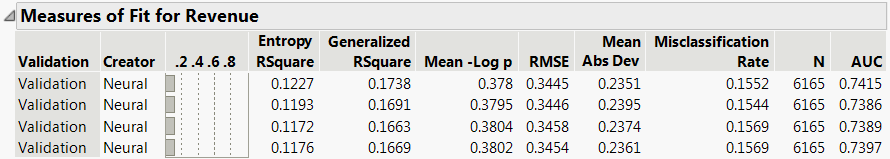
*A3 Table 4: Penalized Regression Confusion Matrix & Associated Values*

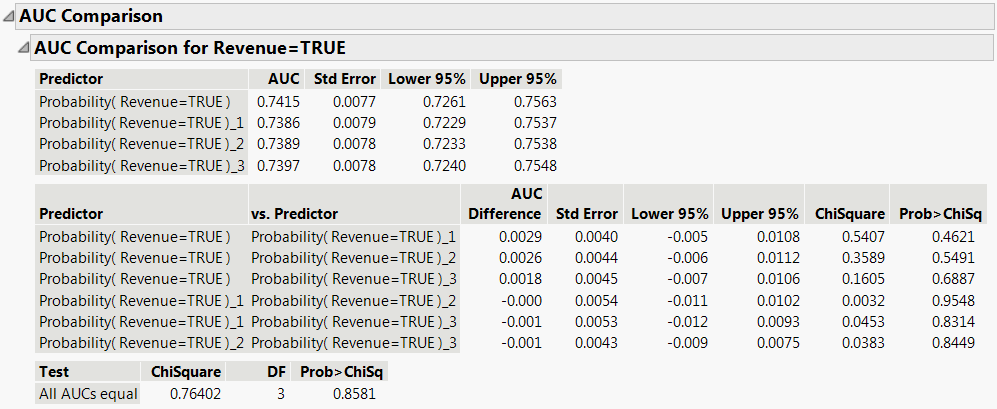
 



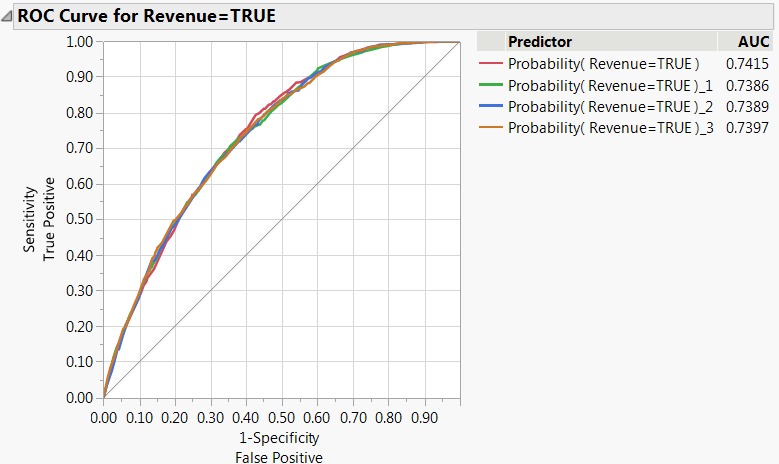
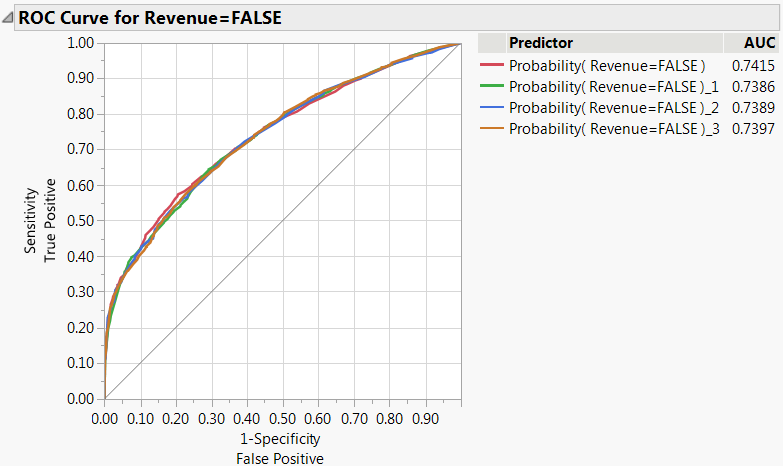
*A3 Figure 3: Penalized Regression Odds Ratios*

Appendix 4 Neural Analysis

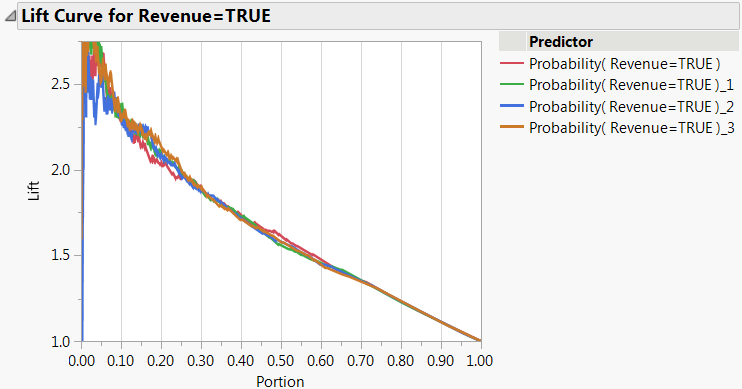
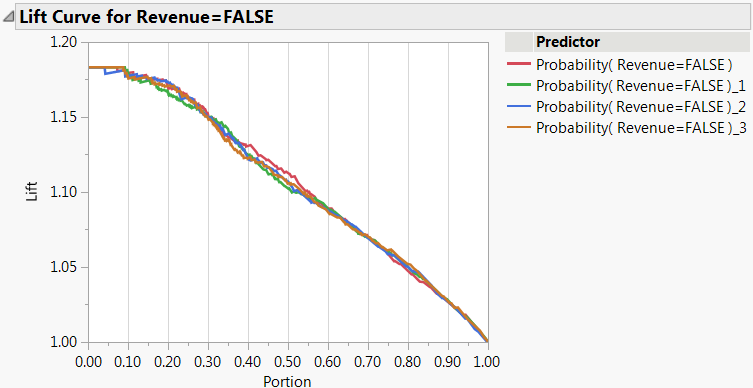




*A4 Figure 1 & 2: Neural Measures of Fit & AUC Comparison*

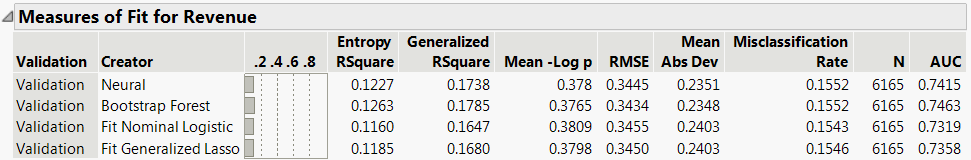


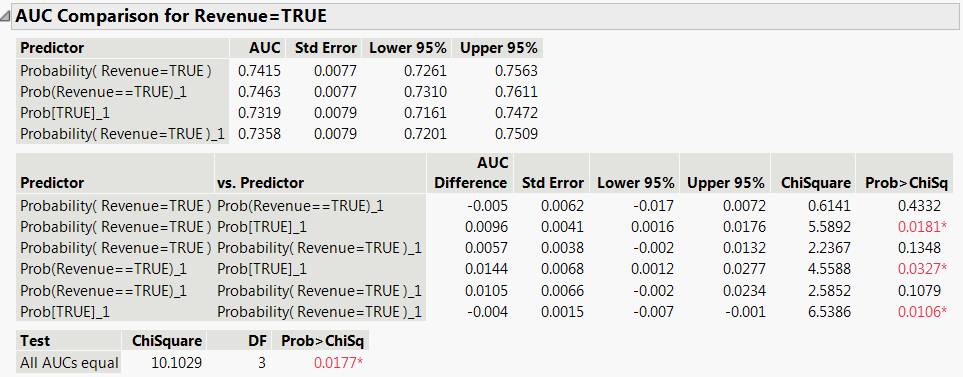
*A4 Figure 3: Neural ROC Curve Comparison*



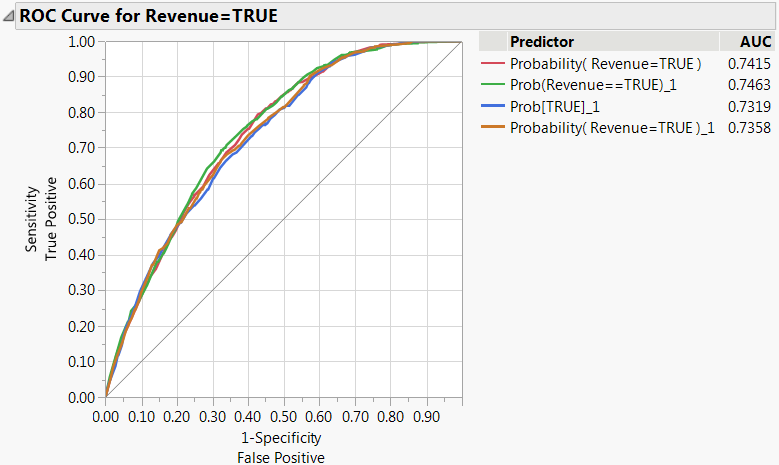
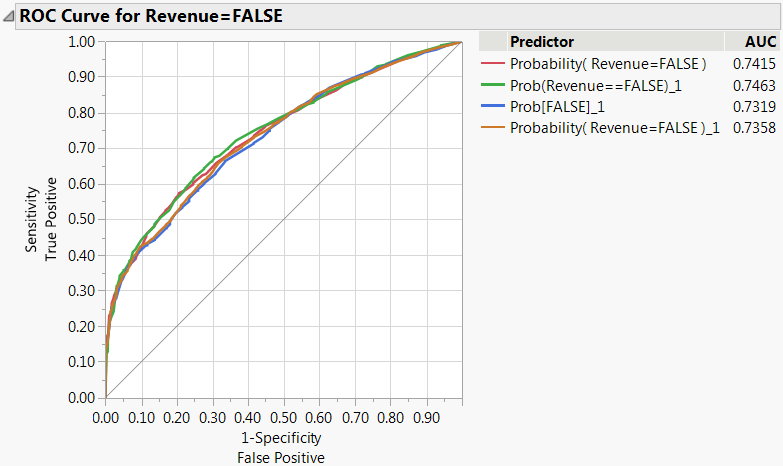
*A4 Figure 4: Neural Lift Curve Comparison*

Appendix 5: Final Models Comparison

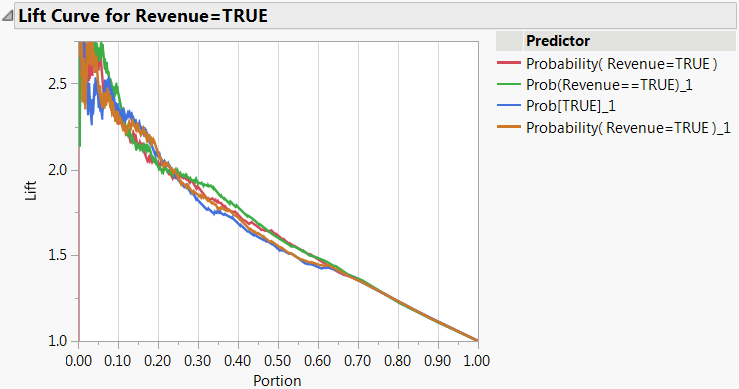
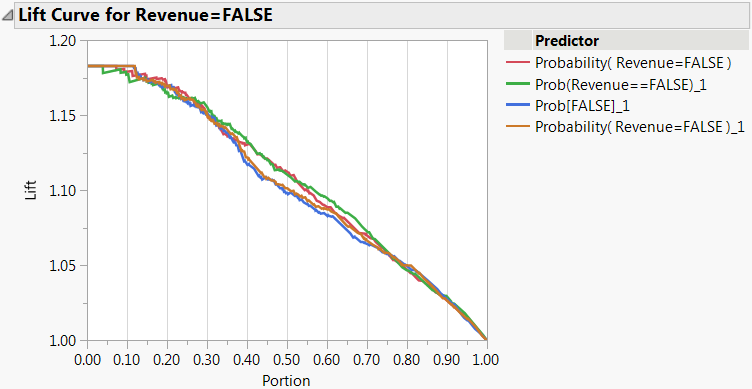




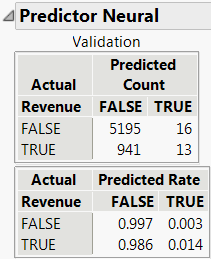
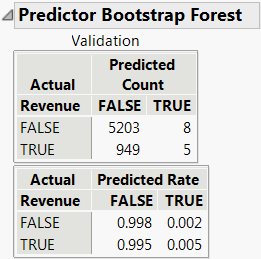
*A5 Figure 1 & 2: Best Models Measures of Fit & AUC Comparison*

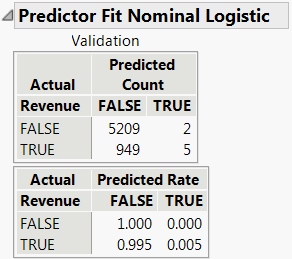
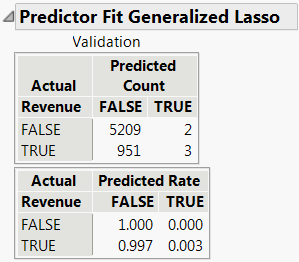


*A5 Figure 3: Best Model ROC Curve Comparison*



*A5 Figure 4: Best Model Lift Curve Comparison*

*A5 Figure 5: Neural & Bootstrap Forest Confusion Matrix Comparison*

**References:**

“Bounce Rate vs. Exit Rate: What's the Difference?” *CXL*, 18 Sept. 2020, cxl.com/guides/bounce-rate/bounce-rate-vs-exit-rate/.

“Exit Rate vs. Bounce Rate - Analytics Help.” *Google*, Google, 2020, support.google.com/analytics/answer/2525491?hl=en.

G, Yiannis. “Exit Rate % Metric Explained Like Never Before.” *Digital Insights*, 8 June 2016, digitalinsightsworld.com/digital-insights/exit-rate-metric-explained/.

Morris, Robert. “How to Convert Your Website Visitors into Buyers.” *Return On Now*, 8 Sept. 2020, returnonnow.com/2017/04/convert-website-visitors-buyers/.

Patel, Neil. “How to Convert Visitors into Customers Using a Landing Page.” *Neil Patel*, 23 Jan. 2020, neilpatel.com/blog/how-to-convert-visitors-into-customers-using-a-landing-page/.

Patel, Neil. “People Pay, Not Pageviews. Here's Why People Are Leaving Your Website Without Converting.” *Neil Patel*, 23 Jan. 2020, neilpatel.com/blog/why-people-leave-without-converting/.

C. Okan Sakar, and Yomi Kastro. “Online Shoppers Purchasing Intention Dataset Data Set .” *UCI Machine Learning Repository*, Center for Machine Learning and Intelligent Systems, 31 Aug. 2018, archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset#.