|  |
| --- |
| SCM HW4 Group 1 |
| Loan Analysis |
| Syracuse University |

|  |
| --- |
| **Nicholas Starr, Jared M Mosley & Sathish Kumar Rajendiran**  6-7-2020 |

**The Problem**

We’re give Universal Bank data to determine the factors influence whether a customer takes out a loan. Based on the underlying data, following analysis were performed.

Universal Bank Data Fields:

A screenshot of a cell phone

Description automatically generated

**Logit and Probit Analysis**

Using Rcmdr & R Studio – Logit and Probit analysis were performed. We ran the analysis with all the variables to find out their significance.

**Logit Analysis**

|  |  |
| --- | --- |
| Based on Logit Analysis, we can safely say that the following independent variables may have higher influence on dependent variable “PersonalLoan” (i.e. likelihood of taking out a loan).   * Income * Family * Education * Credit Card * Online and CDAccount   In addition, CCAvg and SecuritiesAccount also have p-value ~0.001 |  |
| Among these independent variables, CDAccount, Education, Family, Income increases the chances of taking out a loan; where as Creditcard, SecurityAccount and Online variables with negative coefficient may decrease the chances of taking out a loan.  In addition, the coefficient value of Intercept is also in negative. This negative value of coefficient simply means that the expected value of our dependent variable will be less that 0 when all the independent/predictor variables are set to 0. |  |
| **Probit Analysis**  Probit analysis also confirms the more or less the same influence of the independent variables on the dependent variable “PersonalLoan”.  Age, Mortgage and Experience continue to have higher p-value suggesting these values prone to introduce lot of errors in the prediction.  Intercept value has reduced to 50% than Logit; however, still being -ve confirms that the chances of taking out loan is still No when all the predictor values are set to 0. |  |

**Impact of moderating effects (Interactions of variables)**

|  |  |
| --- | --- |
| We have simply introduced a moderating variable by multiplying Income and Age variables. In our example, we took Logit Analysis to analyze the impact of this newly introduced variable. As you can see this variable had made the following changes,   * Income: Age variable is statistically significant having p-value <0.05 * p-value of “Age” is significantly improved * Intercept coefficient has reduced   We don’t see a huge shift in the outcome, yet the moderating variable proved that the interactions of Age & Income variables can result in significant changes to the regression model. |  |
| A screenshot of text  Description automatically generated |  |

**Regression Model & 2-way Sensitive Analysis**

|  |  |  |
| --- | --- | --- |
| Let’s use the same variables (including the moderating variable) to run a regression analysis to find out % of customers taking out loan.  First, run a Linear Regression model using Logit Analysis and based on the coefficient values build a spreadsheet prediction model.  As you can see it on the right, Steps followed are   * Copy the Coefficient values from the regression model output (Copied from R-Studio) * Set Input values to those variables – we just took random assumptions. Example Income to be in 80k, Family Size of 4 and Age as 40 and so on. For, Moderating variable (multiply Age & Income input variables) * Multiply coefficient values with their respect input values (note that for Intercept it’s just multiplied by 1) * Sum of all these values as U * Take Exponential value of U 🡪 EXP(U) * Add 1 to Exp (U) * Probability = EXP(U)/(1+EXP(U))   It is shown as for our input values combination the probability is 76%. It implies that chances of customer taking loan is ~76%. | |  |
| A screenshot of text  Description automatically generated | | A screenshot of a cell phone  Description automatically generated |
| Now, let’s plot a 2-way analysis by taking and Family and Income. You may consider other combinations as well.  This is to prove that outcome of Prediction model is impacted by variations in the input variables. As you can see higher the value of Income and Increase in family size – increases the chances of customer taking loan. |  | |

**Neural Network Analysis & Prediction Model**

|  |  |
| --- | --- |
| Next, let’s build a neural network by taking Education, CDAccount, Family and Income as independent variables. We also introduced 3 hidden layers to reduce the errors improve the accuracy of the prediction.  R-Studio script as shows the script used in plotting neural network. | A screenshot of a social media post  Description automatically generated |

|  |  |  |
| --- | --- | --- |
| A close up of a map  Description automatically generated | | A screenshot of a cell phone  Description automatically generated |
| Build a Prediction model using spreadsheet    A screenshot of a cell phone  Description automatically generated | | Input variables:  A screenshot of a cell phone  Description automatically generated  Output values:  A screenshot of a cell phone  Description automatically generated |
| Now, let’s plot a 2-way analysis by taking and Family and Income. You may consider other combinations as well.  This is to prove that outcome of Neural network Prediction model is impacted by variations in the input variables. As you can see higher the value of Income and Increase in family size – increases the chances of customer taking loan. | A picture containing computer  Description automatically generated | |