**Background**

Data for 5000 customers have been collected by Universal Bank, a relatively young bank growing rapidly in terms of overall customer acquisition. The data collected includes customer demographic information like age and income, their relationship with the bank if they are mortgage holders or have securities accounts, as well as customer’s responses to the last personal loan campaign.

As per the bank, the majority of their customers are liability customers (or depositors) and the number of asset customers (or borrowers) is very small. The bank is looking for ways of converting its liability customers to personal customers, while retaining them as depositors too.

In the marketing campaign run by the bank last year for its liability customers, it witnessed a healthy conversion rate. Encouraged with this, bank wants to formulate campaigns with better target marketing.

To this end, a logistic regression model needs to be built which can classify whether a new customer will be accepting the loan offer.

**Exploratory Data Analysis**

All collected variables have been classified under 2 categories – Categorical and Numeric Variables

|  |  |
| --- | --- |
| **Categorical variables** | **Numeric Variables** |
| Education | Age |
| Family | Experience |
| Personal Loan | Income |
| Securities Account | CCAvg |
| CD Account | Mortgage |
| Online |  |
| CreditCard |  |

Exploratory Data Analysis has been performed to identify patterns with the help of Minitab. Statistical tools like Mean, Standard Deviation, Median, Skewness, distribution analysis, different types of charts and graphs have been used to try to portray picture of a typical Universal Bank customer.

**Statistics**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Mean** | **SE Mean** | **StDev** | **Minimum** | **Q1** | **Median** | **Q3** | **Maximum** | **Skewness** |
| Age | 45.338 | 0.162 | 11.463 | 23.000 | 35.000 | 45.000 | 55.000 | 67.000 | -0.03 |
| Experience | 20.105 | 0.162 | 11.468 | -3.000 | 10.000 | 20.000 | 30.000 | 43.000 | -0.03 |
| Income | 73.774 | 0.651 | 46.034 | 8.000 | 39.000 | 64.000 | 98.000 | 224.000 | 0.84 |
| CCAvg | 1.9379 | 0.0247 | 1.7477 | 0.0000 | 0.7000 | 1.5000 | 2.5000 | 10.0000 | 1.60 |
| Mortgage | 56.50 | 1.44 | 101.71 | 0.00 | 0.00 | 0.00 | 101.00 | 635.00 | 2.10 |

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart

Description automatically generatedChart, histogram

Description automatically generatedUNIVERSALBANK.MTW

**Tally for Discrete Variables: Education, Family, Personal Loan, Securities Account, CD Account, Online, CreditCard**

**Tally**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Education** | | **Count** | | | **Percent** | | | **Family** | | **Count** | | **Percent** | | **Personal Loan** | | **Count** | | **Percent** | |
| 1 | | 2096 | | | 41.92 | | | 1 | | 1472 | | 29.44 | | 0 | | 4520 | | 90.40 | |
| 2 | | 1403 | | | 28.06 | | | 2 | | 1296 | | 25.92 | | 1 | | 480 | | 9.60 | |
| 3 | | 1501 | | | 30.02 | | | 3 | | 1010 | | 20.20 | | N= | | 5000 | |  | |
| N= | | 5000 | | |  | | | 4 | | 1222 | | 24.44 | |  | |  | |  | |
|  | |  | | |  | | | N= | | 5000 | |  | |  | |  | |  | |
| **Securities Account** | **Count** | | | **Percent** | | | **CD Account** | | | | **Count** | | **Percent** | | **Online** | | **Count** | | **Percent** | |
| 0 | 4478 | | | 89.56 | | | 0 | | | | 4698 | | 93.96 | | 0 | | 2016 | | 40.32 | |
| 1 | 522 | | | 10.44 | | | 1 | | | | 302 | | 6.04 | | 1 | | 2984 | | 59.68 | |
| N= | 5000 | | |  | | | N= | | | | 5000 | |  | | N= | | 5000 | |  | |
| **CreditCard** | | | **Count** | | | **Percent** | | |
| 0 | | | 3530 | | | 70.60 | | |
| 1 | | | 1470 | | | 29.40 | | |
| N= | | | 5000 | | |  | | |

Shape

Description automatically generated

As per the analysis, it can be said that an average customer is of around 45 years of age having a professional experience of about 20 years. Almost 42% of the customers are undergrads. Also, a typical customer earns around $64,000 in annual income and has an average monthly credit card spending of approx. $1,500. Since income and credit card monthly spending data is right-skewed, we found median to be a better representative. Further, around 70% of the customers are without any mortgage.

As far as relationship with the bank is concerned, 1/10th of the customers have Securities Account with the bank while 6% hold Bank Certificate of Deposit. Further, 30 out of 100 people use Credit Card issued by Universal Bank. And it might also be interesting to note that more than 90% of the customers rejected the personal loan offered to them in the last campaign.

**Linear Probability Model**

We can try to predict whether a customer will accept the Personal Loan or not given certain conditions using a Linear Probability Model.

Personal Loan has been kept as response variable with continuous predictors being Income, Family, CCAvg, Mortgage, Age and Experience, while categorical predictors are Education and CD Account with confidence level as 95%.

Linear Regression model determined through Minitab are as below.

**Regression Equation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Education** | **CD Account** |  |  |  |
| 1 | 0 | Personal Loan | = | -0.2072 + 0.003105 Income + 0.02996 Family + 0.01256 CCAvg + 0.000079 Mortgage - 0.00563 Age + 0.00616 Experience |
|  |  |  |  |  |
| 1 | 1 | Personal Loan | = | 0.0590 + 0.003105 Income + 0.02996 Family + 0.01256 CCAvg + 0.000079 Mortgage - 0.00563 Age + 0.00616 Experience |
|  |  |  |  |  |
| 2 | 0 | Personal Loan | = | -0.0596 + 0.003105 Income + 0.02996 Family + 0.01256 CCAvg + 0.000079 Mortgage - 0.00563 Age + 0.00616 Experience |
|  |  |  |  |  |
| 2 | 1 | Personal Loan | = | 0.2065 + 0.003105 Income + 0.02996 Family + 0.01256 CCAvg + 0.000079 Mortgage - 0.00563 Age + 0.00616 Experience |
|  |  |  |  |  |
| 3 | 0 | Personal Loan | = | -0.0494 + 0.003105 Income + 0.02996 Family + 0.01256 CCAvg + 0.000079 Mortgage - 0.00563 Age + 0.00616 Experience |
|  |  |  |  |  |
| 3 | 1 | Personal Loan | = | 0.2167 + 0.003105 Income + 0.02996 Family + 0.01256 CCAvg + 0.000079 Mortgage - 0.00563 Age + 0.00616 Experience |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | -0.2072 | 0.0700 | -2.96 | 0.003 |  |
| Income | 0.003105 | 0.000097 | 32.12 | 0.000 | 1.86 |
| Family | 0.02996 | 0.00291 | 10.30 | 0.000 | 1.05 |
| CCAvg | 0.01256 | 0.00245 | 5.12 | 0.000 | 1.73 |
| Mortgage | 0.000079 | 0.000033 | 2.41 | 0.016 | 1.05 |
| Age | -0.00563 | 0.00275 | -2.04 | 0.041 | 93.66 |
| Experience | 0.00616 | 0.00275 | 2.24 | 0.025 | 93.56 |
| Education |  |  |  |  |  |
| 2 | 0.14753 | 0.00823 | 17.93 | 0.000 | 1.28 |
| 3 | 0.15773 | 0.00821 | 19.22 | 0.000 | 1.33 |
| CD Account |  |  |  |  |  |
| 1 | 0.2662 | 0.0140 | 19.07 | 0.000 | 1.04 |

**Model Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **S** | **R-sq** | **R-sq(adj)** | **R-sq(pred)** |
| 0.230614 | 38.84% | 38.73% | 38.49% |

Adjusted R-square of the model is 38.73% which implies the goodness of the fit. Also p-value for all variables is lesser than 0.05, which is our α, denoting that they are significant for our model. However, variables Age and Experience have VIF (Variance Inflation Factor) is 93.66 and 93.56, respectively, which is very high. This shows there is multi-collinearity between these two variables. Also p-value for Age is 0.041 which is very close to α, and that of Experience is 0.025.

To alleviate the problem of multi-collinearity, we tried fitting regression line by removing first only Age, then only Experience and then both Age and Experience. After considering adjusted R-square and complexity of the line equations, we determined it’s best to remove both Age and Experience as predictor variables. Resulting regression models are as below :

**Regression Equation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Education** | **CD Account** |  |  |  |
| 1 | 0 | Personal Loan | = | -0.3350 + 0.003110 Income + 0.02940 Family + 0.01217 CCAvg + 0.000079 Mortgage |
|  |  |  |  |  |
| 1 | 1 | Personal Loan | = | -0.0676 + 0.003110 Income + 0.02940 Family + 0.01217 CCAvg + 0.000079 Mortgage |
|  |  |  |  |  |
| 2 | 0 | Personal Loan | = | -0.1896 + 0.003110 Income + 0.02940 Family + 0.01217 CCAvg + 0.000079 Mortgage |
|  |  |  |  |  |
| 2 | 1 | Personal Loan | = | 0.0778 + 0.003110 Income + 0.02940 Family + 0.01217 CCAvg + 0.000079 Mortgage |
|  |  |  |  |  |
| 3 | 0 | Personal Loan | = | -0.1815 + 0.003110 Income + 0.02940 Family + 0.01217 CCAvg + 0.000079 Mortgage |
|  |  |  |  |  |
| 3 | 1 | Personal Loan | = | 0.0860 + 0.003110 Income + 0.02940 Family + 0.01217 CCAvg + 0.000079 Mortgage |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | -0.3350 | 0.0109 | -30.71 | 0.000 |  |
| Income | 0.003110 | 0.000097 | 32.21 | 0.000 | 1.85 |
| Family | 0.02940 | 0.00290 | 10.12 | 0.000 | 1.04 |
| CCAvg | 0.01217 | 0.00245 | 4.97 | 0.000 | 1.72 |
| Mortgage | 0.000079 | 0.000033 | 2.41 | 0.016 | 1.05 |
| Education |  |  |  |  |  |
| 2 | 0.14539 | 0.00818 | 17.78 | 0.000 | 1.27 |
| 3 | 0.15357 | 0.00794 | 19.33 | 0.000 | 1.24 |
| CD Account |  |  |  |  |  |
| 1 | 0.2674 | 0.0140 | 19.15 | 0.000 | 1.04 |

**Model Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **S** | **R-sq** | **R-sq(adj)** | **R-sq(pred)** |
| 0.230756 | 38.74% | 38.65% | 38.44% |

As can be seen from the above model, the adjusted R-square is 38.65% which is not very different from our earlier adjusted R-square of 38.73%, however VIF for all variables is close to 1 implying problem of multi-collinearity is not there anymore.

On the other hand, there are certain disadvantages while applying linear probability model to predict a categorical variable. Few of these shortcomings are :

1. It ignores the limitations of dependent variable, i.e. it might produce outcomes or predictions which are outside the possible range of values of dependent variable.
2. It lacks all the restrictions on the range of predicted values of the outcome.

For example, in the above case Personal Loan is a binary dependent variable with only two possible outcomes – 0 and 1 (No and Yes), but the model can predict values other than 0 and 1. It may even produce negative values.

**Logistic Regression Model**

Considering the shortcomings of fitting a Linear Probability model to a binary dependent variable, we have tried fitting a Logistic Regression Model to predict if the customer is like to accept the offer of a Personal Loan (which is our response variable) keeping Income, Family, CCAvg, Mortgage, Age and Experience as continuous predictors and Education and CD Account as categorical predictors.

**Regression Equation**

|  |  |  |  |
| --- | --- | --- | --- |
| P(1) | = | exp(Y')/(1 + exp(Y')) | |
| **Education** | | | **CD Account** | |  |  |  |
| 1 | | | 0 | | Y' | = | -13.16 + 0.06005 Income + 0.6124 Family + 0.1561 CCAvg + 0.000793 Mortgage - 0.02479 Age + 0.03307 Experience |
|  | | |  | |  |  |  |
| 1 | | | 1 | | Y' | = | -10.64 + 0.06005 Income + 0.6124 Family + 0.1561 CCAvg + 0.000793 Mortgage - 0.02479 Age + 0.03307 Experience |
|  | | |  | |  |  |  |
| 2 | | | 0 | | Y' | = | -9.210 + 0.06005 Income + 0.6124 Family + 0.1561 CCAvg + 0.000793 Mortgage - 0.02479 Age + 0.03307 Experience |
|  | | |  | |  |  |  |
| 2 | | | 1 | | Y' | = | -6.692 + 0.06005 Income + 0.6124 Family + 0.1561 CCAvg + 0.000793 Mortgage - 0.02479 Age + 0.03307 Experience |
|  | | |  | |  |  |  |
| 3 | | | 0 | | Y' | = | -9.143 + 0.06005 Income + 0.6124 Family + 0.1561 CCAvg + 0.000793 Mortgage - 0.02479 Age + 0.03307 Experience |
|  | | |  | |  |  |  |
| 3 | | | 1 | | Y' | = | -6.626 + 0.06005 Income + 0.6124 Family + 0.1561 CCAvg + 0.000793 Mortgage - 0.02479 Age + 0.03307 Experience |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **Z-Value** | **P-Value** | **VIF** |
| Constant | -13.16 | 1.79 | -7.35 | 0.000 |  |
| Income | 0.06005 | 0.00289 | 20.75 | 0.000 | 2.62 |
| Family | 0.6124 | 0.0751 | 8.15 | 0.000 | 1.29 |
| CCAvg | 0.1561 | 0.0433 | 3.61 | 0.000 | 1.50 |
| Mortgage | 0.000793 | 0.000587 | 1.35 | 0.177 | 1.04 |
| Age | -0.0248 | 0.0662 | -0.37 | 0.708 | 102.11 |
| Experience | 0.0331 | 0.0658 | 0.50 | 0.615 | 101.93 |
| Education |  |  |  |  |  |
| 2 | 3.946 | 0.264 | 14.92 | 0.000 | 2.69 |
| 3 | 4.012 | 0.261 | 15.38 | 0.000 | 2.73 |
| CD Account |  |  |  |  |  |
| 1 | 2.518 | 0.243 | 10.38 | 0.000 | 1.07 |

**Odds Ratios for Continuous Predictors**

|  |  |  |
| --- | --- | --- |
|  | **Odds Ratio** | **95% CI** |
| Income | 1.0619 | (1.0559, 1.0679) |
| Family | 1.8449 | (1.5923, 2.1375) |
| CCAvg | 1.1689 | (1.0739, 1.2724) |
| Mortgage | 1.0008 | (0.9996, 1.0019) |
| Age | 0.9755 | (0.8568, 1.1107) |
| Experience | 1.0336 | (0.9086, 1.1758) |

**Odds Ratios for Categorical Predictors**

|  |  |  |  |
| --- | --- | --- | --- |
| **Level A** | **Level B** | **Odds Ratio** | **95% CI** |
| Education |  |  |  |
| 2 | 1 | 50.4659 | (30.1319, 84.5220) |
| 3 | 1 | 52.6993 | (31.8727, 87.1345) |
| 3 | 2 | 1.0443 | (0.7309, 1.4920) |
| CD Account |  |  |  |
| 1 | 0 | 12.7615 | (7.9391, 20.5130) |

*Odds ratio for level A relative to level B*

**Model Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Deviance R-Sq** | **Deviance R-Sq(adj)** | **AIC** | **AICc** | **BIC** | **Area Under ROC Curve** |
| 61.47% | 61.19% | 1238.33 | 1238.38 | 1303.50 | 0.9585 |

**Goodness-of-Fit Tests**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **DF** | **Chi-Square** | **P-Value** |
| Deviance | 4990 | 1218.33 | 1.000 |
| Pearson | 4990 | 18577.70 | 0.000 |
| Hosmer-Lemeshow | 8 | 67.96 | 0.000 |

**Analysis of Variance**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Wald Test** | |
| **Source** | **DF** | **Chi-Square** | **P-Value** |
| Regression | 9 | 563.65 | 0.000 |
| Income | 1 | 430.67 | 0.000 |
| Family | 1 | 66.47 | 0.000 |
| CCAvg | 1 | 13.01 | 0.000 |
| Mortgage | 1 | 1.82 | 0.177 |
| Age | 1 | 0.14 | 0.708 |
| Experience | 1 | 0.25 | 0.615 |
| Education | 2 | 262.34 | 0.000 |
| CD Account | 1 | 107.73 | 0.000 |

Adjusted deviance R-square is 61.2% which indicates that the model explains approximately 61.2% of deviance in the response. But coming to p-values, it is also seen that predictors Age and Experience have high p-values of 0.708 and 0.615 respectively, which are more than 0.05 level of significance. This implies that these variables are not significant while predicting the value of dependent variable. Also the VIF for both Age and Experience is more than 100 suggesting multi-collinearity between the two.

Hence, we would go ahead with fitting another model with the variables Age and Experience.

The revised model is as below :

**Regression Equation**

|  |  |  |  |
| --- | --- | --- | --- |
| P(1) | = | exp(Y')/(1 + exp(Y')) | |
| **Education** | | | **CD Account** | |  |  |  |
| 1 | | | 0 | | Y' | = | -13.58 + 0.06003 Income + 0.6102 Family + 0.1505 CCAvg + 0.000766 Mortgage |
|  | | |  | |  |  |  |
| 1 | | | 1 | | Y' | = | -11.05 + 0.06003 Income + 0.6102 Family + 0.1505 CCAvg + 0.000766 Mortgage |
|  | | |  | |  |  |  |
| 2 | | | 0 | | Y' | = | -9.644 + 0.06003 Income + 0.6102 Family + 0.1505 CCAvg + 0.000766 Mortgage |
|  | | |  | |  |  |  |
| 2 | | | 1 | | Y' | = | -7.110 + 0.06003 Income + 0.6102 Family + 0.1505 CCAvg + 0.000766 Mortgage |
|  | | |  | |  |  |  |
| 3 | | | 0 | | Y' | = | -9.599 + 0.06003 Income + 0.6102 Family + 0.1505 CCAvg + 0.000766 Mortgage |
|  | | |  | |  |  |  |
| 3 | | | 1 | | Y' | = | -7.065 + 0.06003 Income + 0.6102 Family + 0.1505 CCAvg + 0.000766 Mortgage |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **Z-Value** | **P-Value** | **VIF** |
| Constant | -13.585 | 0.557 | -24.39 | 0.000 |  |
| Income | 0.06003 | 0.00288 | 20.85 | 0.000 | 2.60 |
| Family | 0.6102 | 0.0751 | 8.12 | 0.000 | 1.29 |
| CCAvg | 0.1505 | 0.0431 | 3.49 | 0.000 | 1.49 |
| Mortgage | 0.000766 | 0.000587 | 1.31 | 0.192 | 1.04 |
| Education |  |  |  |  |  |
| 2 | 3.941 | 0.264 | 14.91 | 0.000 | 2.69 |
| 3 | 3.986 | 0.258 | 15.46 | 0.000 | 2.67 |
| CD Account |  |  |  |  |  |
| 1 | 2.533 | 0.242 | 10.47 | 0.000 | 1.07 |

**Odds Ratios for Continuous Predictors**

|  |  |  |
| --- | --- | --- |
|  | **Odds Ratio** | **95% CI** |
| Income | 1.0619 | (1.0559, 1.0679) |
| Family | 1.8408 | (1.5888, 2.1327) |
| CCAvg | 1.1625 | (1.0682, 1.2650) |
| Mortgage | 1.0008 | (0.9996, 1.0019) |

**Odds Ratios for Categorical Predictors**

|  |  |  |  |
| --- | --- | --- | --- |
| **Level A** | **Level B** | **Odds Ratio** | **95% CI** |
| Education |  |  |  |
| 2 | 1 | 51.4596 | (30.6541, 86.3863) |
| 3 | 1 | 53.8331 | (32.4738, 89.2410) |
| 3 | 2 | 1.0461 | (0.7319, 1.4953) |
| CD Account |  |  |  |
| 1 | 0 | 12.5961 | (7.8393, 20.2392) |

*Odds ratio for level A relative to level B*

**Model Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Deviance R-Sq** | **Deviance R-Sq(adj)** | **AIC** | **AICc** | **BIC** | **Area Under ROC Curve** |
| 61.41% | 61.19% | 1236.17 | 1236.20 | 1288.31 | 0.9584 |

**Goodness-of-Fit Tests**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **DF** | **Chi-Square** | **P-Value** |
| Deviance | 4992 | 1220.17 | 1.000 |
| Pearson | 4992 | 17577.49 | 0.000 |
| Hosmer-Lemeshow | 8 | 69.22 | 0.000 |

**Analysis of Variance**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Wald Test** | |
| **Source** | **DF** | **Chi-Square** | **P-Value** |
| Regression | 7 | 564.87 | 0.000 |
| Income | 1 | 434.70 | 0.000 |
| Family | 1 | 66.00 | 0.000 |
| CCAvg | 1 | 12.18 | 0.000 |
| Mortgage | 1 | 1.71 | 0.192 |
| Education | 2 | 263.07 | 0.000 |
| CD Account | 1 | 109.63 | 0.000 |

In the newer Logistic regression model, adjusted deviance R-square is 61.19% which is same as the one we had in the previous model. VIF for all the variables are low and close to 1 indicating non-collinearity among variables, however p-value for one of the variables Mortgage is 0.192 which is more than the level of significance of 0.05. This suggests that Mortgage as a predictor is not significant enough to cause variation in the dependent variable.

We will try to build another model by removing the variable Mortgage and see if it is a better model than the previous one.

Logistic Regression model to predict Personal Loan with continuous predictors Income, Family and CCAvg and categorical variables Education and CD Account is as below :

**Regression Equation**

|  |  |  |  |
| --- | --- | --- | --- |
| P(1) | = | exp(Y')/(1 + exp(Y')) | |
| **Education** | | | **CD Account** | |  |  |  |
| 1 | | | 0 | | Y' | = | -13.55 + 0.06041 Income + 0.6138 Family + 0.1453 CCAvg |
|  | | |  | |  |  |  |
| 1 | | | 1 | | Y' | = | -11.01 + 0.06041 Income + 0.6138 Family + 0.1453 CCAvg |
|  | | |  | |  |  |  |
| 2 | | | 0 | | Y' | = | -9.632 + 0.06041 Income + 0.6138 Family + 0.1453 CCAvg |
|  | | |  | |  |  |  |
| 2 | | | 1 | | Y' | = | -7.086 + 0.06041 Income + 0.6138 Family + 0.1453 CCAvg |
|  | | |  | |  |  |  |
| 3 | | | 0 | | Y' | = | -9.589 + 0.06041 Income + 0.6138 Family + 0.1453 CCAvg |
|  | | |  | |  |  |  |
| 3 | | | 1 | | Y' | = | -7.043 + 0.06041 Income + 0.6138 Family + 0.1453 CCAvg |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **Z-Value** | **P-Value** | **VIF** |
| Constant | -13.554 | 0.555 | -24.41 | 0.000 |  |
| Income | 0.06041 | 0.00287 | 21.06 | 0.000 | 2.59 |
| Family | 0.6138 | 0.0751 | 8.17 | 0.000 | 1.29 |
| CCAvg | 0.1453 | 0.0429 | 3.39 | 0.001 | 1.48 |
| Education |  |  |  |  |  |
| 2 | 3.921 | 0.263 | 14.90 | 0.000 | 2.67 |
| 3 | 3.965 | 0.257 | 15.45 | 0.000 | 2.65 |
| CD Account |  |  |  |  |  |
| 1 | 2.546 | 0.242 | 10.52 | 0.000 | 1.07 |

**Odds Ratios for Continuous Predictors**

|  |  |  |
| --- | --- | --- |
|  | **Odds Ratio** | **95% CI** |
| Income | 1.0623 | (1.0563, 1.0683) |
| Family | 1.8475 | (1.5946, 2.1406) |
| CCAvg | 1.1564 | (1.0631, 1.2579) |

**Odds Ratios for Categorical Predictors**

|  |  |  |  |
| --- | --- | --- | --- |
| **Level A** | **Level B** | **Odds Ratio** | **95% CI** |
| Education |  |  |  |
| 2 | 1 | 50.4659 | (30.1319, 84.5220) |
| 3 | 1 | 52.6993 | (31.8727, 87.1345) |
| 3 | 2 | 1.0443 | (0.7309, 1.4920) |
| CD Account |  |  |  |
| 1 | 0 | 12.7615 | (7.9391, 20.5130) |

*Odds ratio for level A relative to level B*

**Model Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Deviance R-Sq** | **Deviance R-Sq(adj)** | **AIC** | **AICc** | **BIC** | **Area Under ROC Curve** |
| 61.36% | 61.17% | 1235.87 | 1235.89 | 1281.49 | 0.9584 |

**Goodness-of-Fit Tests**

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **DF** | **Chi-Square** | **P-Value** |
| Deviance | 4993 | 1221.87 | 1.000 |
| Pearson | 4993 | 18208.73 | 0.000 |
| Hosmer-Lemeshow | 8 | 52.06 | 0.000 |

**Analysis of Variance**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Wald Test** | |
| **Source** | **DF** | **Chi-Square** | **P-Value** |
| Regression | 6 | 566.01 | 0.000 |
| Income | 1 | 443.67 | 0.000 |
| Family | 1 | 66.77 | 0.000 |
| CCAvg | 1 | 11.46 | 0.001 |
| Education | 2 | 263.17 | 0.000 |
| CD Account | 1 | 110.57 | 0.000 |

This seems to be an optimal Binary Logistic Regression model. P-value of Regression is 0 which is lesser than α of 0.05 suggesting this model is statistically significant in predicting whether a customer will be accepting Personal Loan. P-values for all variables are also lesser than 0.05 indicating they are significant for the model. Adjusted deviance R-square suggests that 61.17% deviance in response variable can be explained with the model. VIF values for all variables are extremely low which means there is no issue of multi-collinearity among various predictors.

Odds ratio for all three continuous predictors is more than 1 implying better the chances of customer accepting Personal Loans with increase in income, family size or average monthly spending on credit cards. Coming to categorical predictors, customer with Advanced/Professional qualification is approximately 53 times and customer having a Graduate qualification is almost 50 times more likely to accept the loan as compared to customer with Undergrad qualification. Similarly, a customer already having a CD account with the bank is 12 times more like to accept the loan than customers who don’t.

As far as goodness of fit is concerned, P-value for Deviance is 1 which is greater than 0.05 level of significance. This means there is not enough statistical evidence that to conclude that the model does not fit the data.

**Case study performed by**

1. Amandeep Singh
2. Dependra Nath Yogi
3. Vishal Singh Thakur
4. Madhukanth Surgi