Case Study: BFS Capstone Project – Mid Submission

Group Details :

1. Sharath Athrey
2. Sumit Kaushik
3. Vandhana Shri

Business Understanding

*Problem Statement :*

CredX is a leading credit card provider that gets thousands of credit card applicants every year. But in the past few years, it has experienced an increase in credit loss. The CEO believes that the best strategy to mitigate credit risk is to ‘acquire the right customers.

*Objective :*

Help CredX identify the right customers using predictive models. Using past data of the bank’s applicants, determine the factors affecting credit risk, create strategies to mitigate the acquisition risk and assess the financial benefit.

Data Understanding

*Datasets :*

* **Demographics Data** : Applicants information collected during the credit card application. It contains 12 variables listed below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| App ID | Age | Gender | Marital Status | Dependents | Income | Education | Profession | Type of Residence |

|  |  |  |
| --- | --- | --- |
| No of Months in Current Residence | No of Months in Current Company | Performance tag |

* **Credit Bureau Data** : The information is obtained directly from credit bureau. It contains 19 variables listed below.

|  |  |  |  |
| --- | --- | --- | --- |
| App ID | No of times 90 DPD or worse in last 6 months | No of times 60 DPD or worse in last 6 months | No of times 30 DPD or worse in last 6 months |
| No of times 90 DPD or worse in last 12 months | No of times 60 DPD or worse in last 12 months | No of times 30 DPD or worse in last 12 months | Avgas CC Utilization in last 12 months |
| No of trades opened in last 6 months | No of trades opened in last 12 months | No of PL trades opened in last 6 months | No of PL trades opened in last 12 months |
| No of Inquiries in last 6 months (excluding home & auto loans) | No of Inquiries in last 12 months (excluding home & auto loans) | Presence of open home loan | Outstanding Balance |
| Total No of Trades | Presence of open auto loan | Performance Tag |

**Data Observations :**

* 71295 observations (rows) in both Demographics and Credit Bureau Data
* Both datasets can be merged using the common variable “Application ID”
* “Performance tag” variable is the target variable

Data Preparation / Cleaning :

**Data Quality (Both Demographics and Credit Bureau):**

1. **Unwanted Data**
   * **3** Duplicate Application ID observations found – excluded
   * **65** records found with age < 18 (-3, 0, 15,16, 17) – excluded
   * **107** records found with income <=0 – excluded
2. **Missing Data**
   * **1425** observations found to be missing “Performance tag”
   * **2** observations found to be missing “Gender”
   * **6** observations found to be missing “Marital Status”
   * **3** observations found to be missing “No of Dependents”
   * **119** observations found to be missing “Education”
   * **14** observations found to be missing “Profession”
   * **8** observations found to be missing “Type of Residence”
   * **1058** observations found to be missing “Avgas CC Utilization in last 12 months”
   * **1** observations found to be missing “No of trades opened in last 6 months”
   * **272** observations found to be missing “Presence of Open Home Loan”
   * **272** observations found to be missing “Outstanding Balance”

**Weight Of Evidence (WOE), IV Analysis :**

* All varibales are analysed using the for WOE using the woe.binning and information packages.
* WOE values are used to replace the missing values in the variables
* From the analysis, it is evident that the variables in the Credit Bureau are significant as compated to the demographic data.
* There are no variables found with predictive power which is strong
* We have identified a total of 12 variables that has relatively less strong (medium) predictive power based on their IV values that are listed below. (Marked yellow)

|  |  |
| --- | --- |
| **Significant Variables (Medium Pred Power)** | **IV Values** |
| No.of.Inquiries.in.last.12.months | 0.2715 |
| Avgas.CC.Utilization.in.last.12.months | 0.2607 |
| No.of.times.30.DPD.or.worse.in.last.6.months | 0.2415 |
| No.of.times.90.DPD.or.worse.in.last.12.months | 0.2138 |
| No.of.times.60.DPD.or.worse.in.last.6.months | 0.2058 |
| No.of.times.30.DPD.or.worse.in.last.12.months | 0.1982 |
| No.of.trades.opened.in.last.12.months | 0.1943 |
| No.of.times.60.DPD.or.worse.in.last.12.months | 0.1854 |
| Total.No.of.Trades | 0.1822 |
| No.of.PL.trades.opened.in.last.12.months | 0.1766 |
| No.of.trades.opened.in.last.6.months | 0.1697 |
| No.of.times.90.DPD.or.worse.in.last.6.months | 0.1601 |
| **InSignificant Variables** | **IV Values** |
| No.of.PL.trades.opened.in.last.6.months | 0.12474369 |
| No.of.Inquiries.in.last.6.months | 0.09293914 |
| No.of.months.in.current.residence | 0.07894352 |
| Income | 0.07894352 |
| No.of.months.in.current.company | 0.02175441 |
| Presence.of.open.home.loan | 0.01762652 |
| Outstanding.Balance | 0.0142395 |
| Age | 0.0033491 |
| woe.Profession.binned | 0.0021820 |
| Presence.of.open.auto.loan | 0.001654 |
| Application.ID | 0.00150419 |
| woe.Gender.binned | 0.00032497 |
| woe.Type.of.residence.binned | 0.00028927 |
| woe.Education.binned | 0.0002694 |
| woe.Marital.Status..at.the.time.of.application..binned | 9.52E-05 |
| No.of.dependents | 5.56E-05 |
| Performance.Tag | 0 |