INSY5339 001 Principles of Business Data Mining Spring 2019

GROUP 5

|  |  |
| --- | --- |
| S. No | Name |
| 1. | Monisha Mallarapu |
| 2. | Reddy Sushmitha Yethirajula |
| 3. | Vishal Reddy Duvvuru |
| 4. | Ashwin Kumar SathishKumar |

Analysis of Data in JSS Groups, an Auto finance company (Tier-3) to acquire knowledge on Business insights and help the business grow successfully using various Data mining Techniques.

**Objective:**

In this project, To perform the data mining techniques and predict various insights of the business using the knowledge discovered through the available data, We are using first hand data. We have acquired the data from authorized auto finance company named JSS Groups in India and have received clearance to use the data for exploring the insights from the employer.

**Brief Description of Data:**

The data we are using for the project consists of 65 variables which describes various components of the business like loan amount, status of the loan, various details about the vehicle and information regarding the customer STATUS, SUB\_STATUS, SZUREDT, RELEASE DT, BRANCH NAME, PRODUCT SEGMENT, FIELD EXECUTIVE, LOAN NUMBER, SCHEMECODE, CP\_LOANNO, CP\_STATUS, ARREARS MONTHS, BUCKET, AGRMNT-DAY ,DUE-DAY ,INSTALLMENT DUE DATE ,LAST COLL TYPE ,LAST RECEIPT DT ,LAST RECEIPT RI AMT ,CURRENT EMI ,CURRENT DUE AMOUNT, MONTH TBC etc.,

**Description of Project:**

TheAuto Finance company data has numerous details about the loan, period of plan, EMI, Arrears of payment, Personal details of the customer and guarantee, Vehicle details etc, .This is a company where there are customers who haven’t paid the EMI’s (1 or more months i.e. arrears) along with ones with 0 arrears. We would like to perform analysis by sampling the data with the help of bootstrapping and by selecting best model among decision trees, radiant boost , random forests, KNN based on accuracy, precision and recall for the available data based on arrears. Perform predictive analysis so that the company would be able to recover the loans from the faulty customers in either settlement with a plan or seizure of the vehicle and use this available model to predict the scenario for new customers.

Analysis of the data available for the auto finance company

|  |  |
| --- | --- |
| How many observations in the dataset? | 2300 |
| How many binary/nominal variables? | 4 |
| How many continuous variables? | 6 |
| What is the outcome / target variable? | SETTLEMENT STATUS |
| What is the level of the target variable (binary/nominal or continuous) | NOMINAL |
| If binary or nominal: What percentage of the variables belong to each class. | SETTLEMENT STATUS:  **No action required** :60%  **Settlement with a Plan**: 30 %  **Seizure of asset**:10% |
| If continuous: What is the mean value of the target variable? | NA |
| Before doing any further processing, what would your prediction of the target variable be? That is what is the initial default value you will predict for the target variable? | SETTLEMENT STATUS by default is NO ACTION REQUIRED |