BFS CAPSTONE PROJECT (CRED X CASE STUDY) SUBMISSION

SUBMITTED BY:

AMIT SOOD

RAJAT GUPTA

PRAKHAR GUPTA

ABHINAV KUMAR

PROBLEM STATEMENT

Identifying the right customers for Acquisition by creating predictive models using past data of the bank's applicants. Also, determine the factors affecting credit risk and assess the financial benefit of the project.

Case Study Analysis Process

Data Extraction and Preparation

- Import Demographic and Credit data into Python notebook.
- Data Cleanup i.e. check for duplicates, null values etc.
- Separate the rows where Target variable is missing. They will be used as reject population for Model Evaluation.

Exploratory Data Analysis

- Segment the Numerical variables in to buckets and analyze the default rates.
- Check default rates for various categorical variables.
- Perform outlier analysis and wherever required impute the outliers with closest percentile value.
- Calculate WOE and IV values for the variables and identify the important variables.

Model Building

- Create Logistic Regression model using the Demographic variables having WOE value.
- Create the Logistic model using the original Demographic data and compare the ROC value with the one created before.
- Create Logistic model using Principal components derived from demographic and credit data.
- Create Random forest model using Principal components derived from demographic and credit data. Tune the hyperparameters of the model.

Application Scorecard & Model Evaluation

- Check the evaluation metrics of final model such as Accuracy, Sensitivity, Specificity, auc_roc
- Identify cut off score for Accept/Reject of application.
- Evaluate the final model using the Rejected population.
- Assess the financial benefit of the model.

WOE & IV ANALYSIS

After calculating the Weight of Evidence, Information values for variables, we identified below key variables:

1. Average CC Utilization in last 12 months
2. No of trades opened in last 12 months
3. No of inquiries in last 12 months
4. Total no of Trades
5. No of times 30 DPD in last 12 months

MODEL BUILDING

We created Logistic Regression model using only Demographic data and using both Demographic and Credit Bureau data but the auc_roc value was not good. Hence, we employed PCA and created the Random Forest model with tuned hyperparameters leading to auc_roc value of 0.67 on train data and 0.68 on test data.

APPLICATION SCORECARD

We created Application scorecard considering base score of 400 doubling at every 20 points with Good to Bad odd ratio of 10:1. We identified cut-off score of 332 below which credit cards application will be rejected.

MODEL EVALUATION

We used the Reject Population(i.e. rows with Application ID having missing Performance Tag) to evaluate our final model. We used the Application scorecard to calculate the score for the reject population and it came out to be less than 332 i.e. all of these applications will get reject when we use our model.

FINANCIAL BENEFIT OF PROJECT

Assumptions:

- 100 people applied for credit card. Based on our model, 60 of them will be granted credit card and 40 will be rejected.
- Average credit loss per default is 50,000 Rs and average revenue generated from a customer is 1500 Rs.
- Default rate for Approved/Rejected applications is as per calculations done on test data.

1. Credit loss, if model wasn't used = ((296+579)/(12587+8269)) * 100 * 50,000 = 2,10,000 Rs
2. Net credit loss avoided will be difference of (Credit loss from Rejected applicants - Credit loss from Approved applicants who may default) = ((7/100) * 40) * 50,000 - ((2.4/100) * 60) * 50,000 = 68,000 Rs
3. Avoided credit loss % = (68,000/2,10,000) * 100 = 32.4%
4. Revenue loss due to rejection of good customers = (93/100) * 40 * 1500 = 55,800 Rs